

**Laserinduzierte Breakdownspektroskopie
zur qualitativen und quantitativen Bestimmung
von Elementgehalten in geologischen Proben
mittels multivariater Analysemethoden
am Beispiel von Kupfer und ausgewählten Seltenen Erden**



kumulative Dissertation

zur Erlangung des akademischen Grades

„doctor rerum naturalium“

(Dr. rer. nat.)

in der Wissenschaftsdisziplin „Physikalische Chemie“

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von
Pia Brinkmann

Potsdam, den 10.05.2022

Soweit nicht anders gekennzeichnet, ist dieses Werk unter einem Creative-Commons-Lizenzvertrag Namensnennung 4.0 lizenziert.

Dies gilt nicht für Zitate und Werke, die aufgrund einer anderen Erlaubnis genutzt werden. Um die Bedingungen der Lizenz einzusehen, folgen Sie bitte dem Hyperlink:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Online veröffentlicht auf dem

Publikationsserver der Universität Potsdam:

<https://doi.org/10.25932/publishup-57212>

<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-572128>

1. Gutachter: Herr Prof. Dr. Hans-Gerd Löhmannsröben[†]
2. Gutachter: Herr Prof. Dr. Uwe Altenberger^{*}
3. Gutachter: Frau Dr. Jeannette Meima[‡]

[†] Universität Potsdam, Institut für Chemie, Physikalische Chemie

^{*} Universität Potsdam, Institut für Geowissenschaften, Strukturpetrologie

[‡] Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover

Inhaltsverzeichnis

Zusammenfassung	7
Abstract	8
Publikationsliste	9
1. Einleitung	10
2. Grundlagen	13
2.1 Minerale, Gesteine, Lagerstätten und Böden.....	13
2.2 Kupfer.....	14
2.3 Seltene Erden.....	15
2.4 Laserinduzierte Breakdownspektroskopie.....	16
2.4.1 Prinzip.....	17
2.4.2 Plasma und Breakdown.....	18
2.4.3 Emissionsspektrum.....	19
2.4.4 LIBS-Instrumente.....	23
2.5 Chemometrie.....	24
2.5.1 Univariate Analyse.....	24
2.5.2 Hauptkomponentenanalyse, eine multivariate Analysemethode.....	25
Robuste Hauptkomponentenanalyse.....	26
2.5.3 PLS Regression, eine multivariate Analysemethode.....	27
Interval partial least squares-Regression.....	28
3. Publikationen	29
3.1 Darstellung des eigenen Anteils an den einzelnen Publikationen.....	29
Manuskript 1.....	29
Manuskript 2.....	29
Manuskript 3.....	29
Manuskript 4.....	30
3.2 Vergleich von Kalibrierungsansätzen in der laserinduzierten Breakdownspektroskopie für die proximale Bodenerfassung in der Präzisionslandwirtschaft (<i>Manuskript 1</i>).....	31
3.3 Detektion von Seltenen Erdelementen in Mineralen und Böden mittels laserinduzierter Breakdownspektroskopie und Interval-PLSR (<i>Manuskript 2</i>).....	48
3.4 Vergleich von Handheld- und Echelle-Spektrometer zur Bestimmung von Kupfergehalten in Erzen mit Hilfe der laserinduzierten Breakdownspektroskopie (<i>Manuskript 3</i>).....	66

3.5 Klassifizierung kupferhaltiger Minerale mittels laserinduzierter Breakdownspektroskopie und multivariater Datenanalyse (<i>Manuskript 4</i>).....	88
4. Diskussion	106
4.1 Kalibrierung.....	106
4.2 Quantifizierung.....	108
4.3 Klassifizierung.....	112
5. Zusammenfassung und Ausblick	115
Literaturverzeichnis	117
Abkürzungsverzeichnis	124
Abbildungsverzeichnis	125
Tabellenverzeichnis	126
Anhang	127
S.1 Ergänzende Informationen zu den Publikationen.....	127
S1.1 Detektion von Seltenen Erdelementen in Mineralen und Böden mittels laserinduzierter Breakdownspektroskopie und Interval-PLSR (<i>Manuskript 2</i>).....	127
S1.2 Klassifizierung kupferhaltiger Minerale mittels laserinduzierter Breakdownspektroskopie und multivariater Datenanalyse (<i>Manuskript 4</i>).....	132
Danksagung	145
Eidesstattliche Erklärung	146

Zusammenfassung

Ein schonender Umgang mit den Ressourcen und der Umwelt ist wesentlicher Bestandteil des modernen Bergbaus sowie der zukünftigen Versorgung unserer Gesellschaft mit essentiellen Rohstoffen. Die vorliegende Arbeit beschäftigt sich mit der Entwicklung analytischer Strategien, die durch eine exakte und schnelle Vor-Ort-Analyse den technisch-praktischen Anforderungen des Bergbauprozesses gerecht werden und somit zu einer gezielten und nachhaltigen Nutzung von Rohstofflagerstätten beitragen. Die Analysen basieren auf den spektroskopischen Daten, die mittels der laserinduzierten Breakdownspektroskopie (LIBS) erhalten und mittels multivariater Datenanalyse ausgewertet werden. Die LIB-Spektroskopie ist eine vielversprechende Technik für diese Aufgabe. Ihre Attraktivität machen insbesondere die Möglichkeiten aus, Feldproben vor Ort ohne Probennahme oder -vorbereitung messen zu können, aber auch die Detektierbarkeit sämtlicher Elemente des Periodensystems und die Unabhängigkeit vom Aggregatzustand. In Kombination mit multivariater Datenanalyse kann eine schnelle Datenverarbeitung erfolgen, die Aussagen zur qualitativen Elementzusammensetzung der untersuchten Proben erlaubt. Mit dem Ziel die Verteilung der Elementgehalte in einer Lagerstätte zu ermitteln, werden in dieser Arbeit Kalibrierungs- und Quantifizierungsstrategien evaluiert. Für die Charakterisierung von Matrixeffekten und zur Klassifizierung von Mineralen werden explorative Datenanalysemethoden angewendet. Die spektroskopischen Untersuchungen erfolgen an Böden und Gesteinen sowie an Mineralen, die Kupfer oder Seltene Erdelemente beinhalten und aus verschiedenen Lagerstätten bzw. von unterschiedlichen Agrarflächen stammen.

Für die Entwicklung einer Kalibrierungsstrategie wurden sowohl synthetische als auch Feldproben von zwei verschiedenen Agrarflächen mittels LIBS analysiert. Anhand der Beispielanalyten Calcium, Eisen und Magnesium erfolgte die auf uni- und multivariaten Methoden beruhende Evaluierung verschiedener Kalibrierungsmethoden. Grundlagen der Quantifizierungsstrategien sind die multivariaten Analysemethoden der partiellen Regression der kleinsten Quadrate (PLSR, von engl.: partial least squares regression) und der Intervall PLSR (iPLSR, von engl.: interval PLSR), die das gesamte detektierte Spektrum oder Teilspektren in der Analyse berücksichtigen. Der Untersuchung liegen synthetische sowie Feldproben von Kupfermineralen zugrunde als auch solche die Seltene Erdelemente beinhalten. Die Proben stammen aus verschiedenen Lagerstätten und weisen unterschiedliche Begleitmatrices auf. Mittels der explorativen Datenanalyse erfolgte die Charakterisierung dieser Begleitmatrices. Die dafür angewendete Hauptkomponentenanalyse gruppiert Daten anhand von Unterschieden und Regelmäßigkeiten. Dies erlaubt Aussagen über Gemeinsamkeiten und Unterschiede der untersuchten Proben im Bezug auf ihre Herkunft, chemische Zusammensetzung oder lokal bedingte Ausprägungen. Abschließend erfolgte die Klassifizierung kupferhaltiger Minerale auf Basis der nicht-negativen Tensorfaktorisierung. Diese Methode wurde mit dem Ziel verwendet, unbekannte Proben aufgrund ihrer Eigenschaften in Klassen einzuteilen.

Die Verknüpfung von LIBS und multivariater Datenanalyse bietet die Möglichkeit durch eine Analyse vor Ort auf eine Probennahme und die entsprechende Laboranalytik weitestgehend zu verzichten und kann somit zum Umweltschutz sowie einer Schonung der natürlichen Ressourcen bei der Prospektion und Exploration von neuen Erzgängen und Lagerstätten beitragen. Die Verteilung von Elementgehalten der untersuchten Gebiete ermöglicht zudem einen gezielten Abbau und damit eine effiziente Nutzung der mineralischen Rohstoffe.

Abstract

The sustainable use of resources and the environment is an important part of modern mining and the supply of our society with essential raw materials in the future. The present work focuses on the development of analytical strategies that address the technical-practical requirements of the mining process through accurate and rapid on-site analysis, thus contributing to the targeted and sustainable use of raw material deposits. The analyses are based on spectroscopic data obtained by laser-induced breakdown spectroscopy (LIBS) and evaluated by multivariate data analysis. LIB spectroscopy is a promising technique for this task. Its advantages are in particular the possibility to measure field samples on site without sample collection or preparation, but also the detectability of all elements of the periodic table and the independence of the state of matter. In combination with multivariate data analysis, rapid data processing can be performed, allowing statements to be made on the qualitative elemental composition of the samples investigated. With the goal of determining the distribution of elemental contents in a deposit, calibration and quantification strategies are evaluated in this work. Exploratory data analysis methods are used to characterize matrix effects and to classify minerals. Spectroscopic studies are performed on soils and rocks as well as on minerals containing copper or rare earth elements originating from different deposits or from different agricultural sites.

To develop a calibration strategy, both synthetic and field samples from two different agricultural sites were analyzed using LIBS. Using calcium, iron and magnesium as example analytes, the evaluation of different calibration methods based on univariate and multivariate methods was performed.

Basics of the quantification strategies are the multivariate analysis methods of partial least squares regression (PLSR) and interval PLSR (*i*PLSR), which consider the whole detected spectrum or partial spectra in the analysis. The investigation is based on synthetic and field samples of copper minerals as well as those containing rare earth elements. The samples are from different deposits and have varying accompanying matrices.

Exploratory data analysis was used to characterize these accompanying matrices. The principal component analysis used for this purpose groups data on the basis of differences and regularities. This allows conclusions to be drawn about similarities and differences between the samples examined in terms of their origin, chemical composition or locally determined characteristics. Finally, the classification of copper-bearing minerals was based on non-negative tensor factorization. This method was used with the aim of classifying unknown samples based on their properties.

The combination method of LIBS and multivariate data analysis offers the possibility to avoid sampling and the corresponding laboratory analysis as far as possible by an on-site analysis and can thus contribute to environmental protection as well as to a conservation of natural resources during the prospection and exploration of new ore veins and deposits. The distribution of element contents of the investigated areas also enables a precise mining and thus an efficient utilization of the mineral raw materials.

Publikationsliste

P. Brinkmann, N. Köllner, S. Merk, T. Beitz, U. Altenberger, H.-G. Löhmannsröben, „Comparison of Handheld and Echelle Spectrometer to assess Copper in Ores by means of Laser-Induced Breakdown Spectroscopy (LIBS)“, eingereicht bei *Minerals* (2022).

N. Rethfeldt, P. Brinkmann, D. Riebe, T. Beitz, N. Köllner, U. Altenberger, H.-G. Löhmannsröben, "Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS", *Minerals* **11** (12), 1379 (2021).

M. Wójcik, P. Brinkmann, R. Zdunek, D. Riebe, T. Beitz, S. Merk, K. Cieślik, D. Mory, A. Antończak, "Classification of Copper Minerals by Handheld Laser-Induced Breakdown Spectroscopy and Nonnegative Tensor Factorisation", *Sensors* **20** (18), 5152 (2020).

D. Riebe, A. Erler, P. Brinkmann, T. Beitz, H.-G. Löhmannsröben, R. Gebbers, „Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture“, *Sensors* **19** (23), 5244 (2019).

1. Einleitung

„Natürliche Ressourcen [...] sind die Grundlage unseres täglichen Lebensbedarfs und unseres Wohlstands.“¹ Es sind die Bestandteile der Natur, die dem Menschen Nutzen bringen und auf den vielfältigsten Wegen eingesetzt werden. Durch ihren Gebrauch sollen einerseits die globalen Lebensstandards der Weltbevölkerung verbessert werden. Andererseits führt jede Entnahme von natürlichen Ressourcen aus der Natur zu zum Teil schweren Einschnitten in die Umwelt, gefolgt von weltweiten negativen sozialen Auswirkungen und Umweltproblemen. Voraussetzung für die Nutzung der natürlichen Ressourcen ist daher ein schonender und nachhaltiger Umgang mit den zur Verfügung stehenden Bestandteilen der Natur. Neben den strömenden Ressourcen wie Wind oder Sonnenenergie und den Umweltmedien, Luft, Wasser und Boden, umfassen die natürlichen Ressourcen unter anderem die erneuerbaren und nicht-erneuerbaren Rohstoffe. Teil der nicht-erneuerbaren Rohstoffe sind fossile Energieträger, nicht-metallische Minerale und Metallerze. Da insbesondere die Vorkommen der nicht-nachwachsenden Rohstoffe in der Natur begrenzt sind, ist deren Nutzung endlich. Verwendet werden die nicht-erneuerbaren Rohstoffe unter anderem als Basis für Baumaterialien, Energiequellen oder Grundstoffe in der chemischen Industrie sowie im Maschinenbau. Daraus resultiert eine stetig steigende Nachfrage an diesen Rohstoffen getrieben von Megatrends wie Digitalisierung, Energie- und Verkehrswende, sowie einer wachsenden Weltbevölkerung und Weltwirtschaft. Der permanente Bedarf, aber auch die übermäßige Ausbeutung und die ineffiziente Nutzung der Rohstoffe führen zu einer Verschlechterung der Ökosysteme, und zu einer zunehmenden Umweltverschmutzung sowie Verknappung der Ressourcen. Die möglichen Auswege aus der Verknappung der nicht-erneuerbaren Rohstoffe, vorwiegend der Metallerze, sind vielfältig, bedingen aber grundsätzlich einen nachhaltigen Einsatz und Umgang mit den Ressourcen. Neben einer detaillierten Abbauplanung, die auch eine Nachnutzung der veränderten Umwelt berücksichtigt, können Stoffkreisläufe und Substitution von Materialien mögliche Wege aus der Verknappung sein. Zudem bietet sich die Möglichkeit bis dato unentdeckte oder unrentable Lagerstätten von Metallerzen nachhaltig zu prospektieren und explorieren. Für eine optimale Erkundung und Ausbeutung von Lagerstätten werden effiziente und zuverlässige Analysetechniken benötigt, die Minerale, Gesteine und Böden sicher detektieren und identifizieren können. Gleichzeitig müssen diese Techniken die chemischen Komponenten der mineralische Rohstoffe am besten vor Ort, genau und schnell analysieren können, um den Bergbauprozess durch Kenntnisse über die lokalen Gegebenheiten unterstützen und damit eine sinnvolle Ausbeutung und Nutzung der Ressourcen fördern zu können.^{1, 2, 3, 4, 5}

In der vorliegenden Arbeit werden zwei Beispiele nicht-erneuerbarer, primärer Rohstoffe betrachtet: zum einen Kupfer - eines der weltweit wichtigsten Funktionsmetalle - und zum anderen Seltene Erden - essentielle Rohstoffe der Energiewende.^{6, 7}

Kupfer ist ein Buntmetall der Superlative. Es weist nahezu die höchste elektrische Leitfähigkeit, Wärmeleitfähigkeit und Duktilität aller Metalle auf. In diesen Eigenschaften wird es nur vom Edelmetall Silber übertroffen.^{8, 9} Ergänzend zu diesen Eigenschaften machen auch die gute Verformbarkeit, die Korrosionsbeständigkeit, die antibakterielle Wirkung sowie die Möglichkeit den Rohstoff ohne Qualitätsverlust zu recyceln Kupfer zu einem der wichtigsten Metalle der Menschheit.^{6, 10} Aufgrund seiner vielfältigen, hervorragenden Eigenschaften, wird Kupfer in zahlreichen Bereichen des täglichen Lebens eingesetzt. Hauptsächlich verwendet wird Kupfer in der Elektronik und Elektrotechnik sowie im Bau- und Transportsektor. Weitere Verbraucher sind der

Maschinenbau und diverse andere Verbrauchsgüter.^{10, 11, 12} Die verschiedenen Anwendungsfelder für das Metall Kupfer werden zukünftig bestehen bleiben. Durch die globalen Megatrends, die auf die Verfügbarkeit des Metalls angewiesen sind, wird die Nachfrage für Kupfer zusätzlich steigen. Der zunehmende Bedarf an Kupfer macht den Abbau bis dato unattraktiver Lagerstätten mit geringen Kupfergehalten lukrativ. Da Kupfer vorrangig im Tagebau gefördert wird, wird dadurch gleichzeitig die Umwelt stärker belastet. Der Abbau der Kupfererze mit geringen Gehalten geht mit beträchtliche Mengen ungenutzter Materialien einher, die als Abraum weitere Flächen bedecken. Verbunden mit den niedrigen Erzgehalten ist auch der Energieverbrauch durch Kraftstoffe für den Abbau und elektrische Energie für die Aufbereitung erhöht und wird weiter steigen. Problematisch beim Kupferbergbau ist zudem der relativ hohe Wasserverbrauch und die Bildung saurer Grubenwässer.¹² Eine Alternative zum primären Rohstoff ist das Kupfer aus Sekundärrohstoffen, wie Kupferschrotten. Da Kupfer ohne Einbußen in der Funktionalität beliebig häufig recyclebar ist, bietet sich eine nachhaltige Kreislaufwirtschaft an. Diese ermöglicht Einsparungen an den für den Bergbau benötigten Ressourcen, wie Wasser, Land und Energie. Problematisch ist dabei zum einen die lange Lebensdauer kupferhaltiger Produkte und zum anderen, dass die Schrotte zunehmend wertmetallärmer und komplexer werden. Eine Deckung des Bedarfs an Kupfer durch das Recycling ist damit derzeit nicht möglich. Auch die Substitution von Kupfer in Produkten ist aufgrund der charakteristischen Eigenschaften des Metalls nicht in allen Bereichen möglich.^{11, 12, 13}

Das zweite Beispiel für mineralische Rohstoffe in dieser Arbeit sind die Metalle der Seltenen Erdelemente (SEE). Sie zeigen vor allem ungewöhnliche optische und magnetische Eigenschaften. Bedingt durch ihre Charakteristika kommen sie in den verschiedensten Anwendungen zum Einsatz: Gläser, Keramiken, Polituren, Phosphore, Legierungen, Katalysatoren und Magneten, um nur einige Beispiele zu nennen. Eines der wichtigsten Einsatzgebiete ist die Verwendung der SEE in Dauermagneten, die wiederum hauptsächlich in Generatoren zur Energiegewinnung und in Elektromotoren verbaut werden. Dadurch sind sie unerlässliche Rohstoffe insbesondere für die Energie- und Verkehrswende.^{7, 10, 14} Gleichzeitig wird damit auch der stetig steigende Bedarf an SEE deutlich.¹⁵ Der Abbau der SEE-Mineraler erfolgt typischerweise im Tagebau. Durch die niedrigen SEE-Gehalte, kommt es auch hier, wie bei der Kupfergewinnung, zu einem intensiven Flächenverbrauch und zu großen Abbaumengen. Problematisch ist vor allem die Vergesellschaftung der SEE mit Elementen, wie Thorium und Uran, die zu radioaktiv kontaminierten Abwässern und Stäuben in den Abbauregionen führen.⁷ Wirtschaftlich nutzbares Recycling ist aufgrund des hohen technischen Aufwands, geringer Rücklaufquoten und der starken Verteilung der SEE in den Endprodukten sehr schwierig und trägt aktuell nicht zur globalen Versorgung mit SEE bei. Auch die Substitution mit anderen Elementen ist wegen der speziellen Eigenschaften der SEE kompliziert.⁷

Eine Lösung für eine nachhaltige Gewinnung von Kupfer und SEE bietet die Prospektion und Exploration neuer und die gezielte Ausbeutung bestehender Lagerstätten. Die Prospektion und Exploration einer potentiellen Lagerstätte erfolgt üblicherweise über viele Jahre und ist kleinschrittig. Dabei kommen etliche Techniken zum Auffinden und Erkunden der neuen Lagerstätten zum Einsatz. Beispiele sind geophysikalische Methoden, Kartierung, Fernerkundung, Bohrungen und geochemische Analysemethoden. Überdies beruht die herkömmliche Verfahrensweise auf den fachlichen Einschätzungen von Geologen. In seiner Gesamtheit ist dieses Vorgehen personal-, zeit- und kostenintensiv. Vor allem die systematischen Bohrungen und Probennahmen sowie die Laboranalytik der Feldproben sind teuer.^{16, 17, 18} Zudem wird der lange Nachweiszyklus, bedingt durch die komplexe Probennahme und -vorbereitung sowie den hohen Personalaufwand, den

Anforderungen einer schnellen Analyse in der Praxis nicht gerecht. Insbesondere die Probennahme und -analyse erfordern beim bisher üblichen Vorgehen einen zeitlichen Aufwand von mehreren Stunden bis Tagen, bis die laboranalytischen Ergebnisse vorliegen, auf deren Grundlage Entscheidungen zur weiteren Prospektion und Exploration getroffen werden können. Untersuchungsmethoden, die mittels tragbarer Geräte eine Analyse vor Ort und ohne aufwändige Probenvorbereitung erlauben, können eine Alternative sein, da sie vor allem die Probennahme und damit unnötige Eingriffe in die Natur durch ihre minimalinvasive und oberflächennahe Anwendung reduzieren oder überflüssig machen und zusätzlich die Detektion von Rohstoffen und Ressourcen in Kombination mit Beprobung durch Bohrungen erheblich beschleunigen.¹⁶ Ein Beispiel für eine geeignete mobile Technik ist die Röntgenfluoreszenzanalyse (RFA). Sie ist eine etablierte Methode für die Analyse der chemischen Zusammensetzung geologischer Materialien, wie Gesteine oder Sedimente.¹⁹ Basierend auf der durch Röntgenstrahlung ausgelösten Fluoreszenzstrahlung wird bei der RFA die Elementzusammensetzung von vorrangig festen Proben analysiert. Robuste Kalibrierungsverfahren ermöglichen hierbei eine Quantifizierung der Elementgehalte, wobei eine Probenvorbereitung nicht zwingend erforderlich ist. Die Analyse der leichten Elemente, wie Lithium und Magnesium ist für die RFA problematisch.²⁰ Die jüngste Ergänzung der mobilen Techniken für die Analyse geologischer Materialien vor Ort ist die laserinduzierte Breakdownspektroskopie (LIBS), die eine schnelle Analyse aller Elemente des Periodensystems in Probenmaterial jeden Aggregatzustands erlaubt.²¹ Hierbei wechselwirkt ein Laserpuls mit der zu untersuchenden Probe unter Ausbildung eines Plasmas. Die vom Plasma emittierte charakteristische Strahlung erlaubt in situ eine Analyse der elementaren, chemischen Zusammensetzung einer Feldprobe. Neben der Möglichkeit die LIBS-Instrumente so weit zu miniaturisieren, dass sie als portables Gerät im Feld eingesetzt werden können, machen die simultane Multielementanalytik und die geringfügige bis nicht nötige Probenvorbereitung die Methode für Feldanalysen attraktiv. Vor allem qualitativ hat sich die Methode in Forschung und Industrie etabliert. Eine quantitative Analyse mittels LIBS ist bis dato Gegenstand der Forschung.²² Die Forschungsschwerpunkte dieser Arbeit umfassen sowohl die qualitative LIB-Spektroskopie, beispielsweise zur Klassifizierung kupferhaltiger Minerale, als auch die quantitative LIBS, zur Gehaltsbestimmung von Kupfer in Erzen sowie von SEE in Erzen und Böden. Dies dient der Evaluierung verschiedener Kalibrierungsmethoden und Quantifizierungsansätze. Für eine verbesserte Genauigkeit der Analysen wird die LIBS mit multivariaten Methoden kombiniert. Diese Methoden können die Komplexität der Spektren erfassen sowie wichtige Informationen aus den Spektren extrahieren. Damit können sie zur Differenzierung und zur Bestimmung der chemischen Zusammensetzung und Herkunft der zu analysierenden Materialien eingesetzt werden.¹⁶ Die Untersuchungen erfolgen mit dem Ziel LIBS als Methode zur ressourcenschonenden und damit umweltfreundlichen Prospektion und Exploration neuer Kupfer- und SEE-Lagerstätten sowie zur Identifikation neuer Erzgänge in bestehenden Lagerstätten zu erforschen. Ein zusätzliches Augenmerk liegt auf Analyse landwirtschaftlicher Böden mittels LIBS.

2. Grundlagen

2.1 Minerale, Gesteine, Lagerstätten und Böden

Minerale sind definiert als natürlich gebildete, anorganische, stofflich homogene, meist kristalline Festkörper, die eine charakteristische chemische Zusammensetzung aufweisen. Basierend auf der chemischen Zusammensetzung der Anionengruppen können die Minerale, je nach Autor, in bis zu neun Mineralklassen eingeteilt werden (Tabelle 1). Zusätzlich werden gediegene Minerale der Klasse „Elemente“ zugeordnet.^{23, 24, 25, 26} Die der Klassifizierung vorausgehende Identifikation der Minerale beruht auf den chemischen und physikalischen Mineraleigenschaften, wie beispielsweise Farbe, Glanz, Bruch, Härte und Kristallform.

Tabelle 1: Einteilung der Minerale in Mineralklassen nach der Mineralsystematik nach Strunz.^{26, 27}

	Mineralklasse	Beispiel	
		Mineral	Chemische Formel
I	Elemente	Kupfer	Cu
II	Sulfide , Sulfosalze	Chalkopyrit	CuFeS ₂
III	Halogenide	Boleit	KPb ₂₆ Ag ₉ Cu ₂₄ (OH) ₄₈ Cl ₆₂
IV	Oxide , Hydroxide	Cuprit	Cu ₂ O
V	Karbonate , Nitrate	Malachit	Cu ₂ (CO ₃)(OH) ₂
VI	Borate	Bandyolith	CuB(OH) ₄ Cl
VII	Sulfate , Chromate, Molybdate, Wolframate	Brochantit	Cu ₄ (SO ₄)(OH) ₆
VIII	Phosphate , Arsenate, Vanadate	Konichalcit	CaCu(AsO ₄)(OH)
IX	Silikate	Dioptas	CuSiO ₃ · H ₂ O
X	Organische Verbindungen	Weddellit	Ca[C ₂ O ₄] · 2 H ₂ O

Neben der Identifikation bieten die Charakteristika der Minerale die Möglichkeit die Entstehung der Gesteine zu rekonstruieren, da die Gesteine aus den Mineralen aufgebaut sind. Ein Gestein ist per Definition ein natürliches festes Aggregat aus Mineralen, Mineral- oder Gesteinsbruchstücken und Resten von Organismen.^{23, 24, 25} Gesteine werden in drei Gruppen gegliedert: magmatische Gesteine, Sedimentgesteine und metamorphe Gesteine. Die Gesteine dieser drei Gruppen können unter dem Einfluss von Plattentektonik und Klima permanent ineinander überführt werden. Dieser Gesteinskreislauf führt mitunter zur Anreicherung von Mineralen in der Erdkruste, den sogenannten Lagerstätten, die eine dreidimensionale Ausdehnung von zum Teil mehreren hundert Metern aufweisen. Diese Lagerstätten sind nicht ausschließlich für die Gewinnung von Metallen interessant, sondern auch als Quelle für beispielsweise Baustoffe und Keramiken sowie als Rohmaterialien für Düngemittel und Halbleiterprodukte.^{23, 24} Im Hinblick auf Lagerstätten mit Mineralen, die für die Metallgewinnung genutzt werden können, bezeichnet man die Vorkommen als Erze und die Minerale als Erzminerale. Die größten Anteile an den Erzmineralen haben die Klassen der Sulfide, Oxide und Silikate, deren Vorkommen weltweit in Tage- und Untertagebauen abgebaut werden.^{23, 24, 25} Eine weitere Folge des Gesteinskreislaufs ist die Ablagerung von unverfestigtem Material. Diese als

Lockergestein beschriebenen Sedimente sind durch Sedimentation und durch die Verwitterung von Gesteinen entstanden. Lockergesteine sind sehr heterogen und können organische Bestandteile beinhalten. Sie werden auch als Boden bezeichnet.^{23,24,25} In der Land- und Forstwirtschaft steht der Begriff Boden nicht synonym für Lockergestein. Vielmehr werden hier komplexe Auf- und Abbauprozesse zwischen Atmo-, Bio-, Hydro- und Lithosphäre betrachtet.²⁴

2.2 Kupfer

Kupfer hat eine lange Geschichte. Bereits im Neolithikum wurden farbige Kupferminerale wie Malachit sowie gediegenes Kupfer verwendet. Mit der Kupferverhüttung beginnt das Chalkolithikum, die Kupferzeit. Später, mit zunehmender Entwicklung und Verwendung des Metalls, werden auch die Kupfersteinzeit und die Bronzezeit nach dem Buntmetall benannt.²⁸ Kupfer kann in der Natur gediegen vorkommen. Häufiger tritt es jedoch in Verbindung mit anderen Elementen als Minerale auf.^{29,30} Bis dato sind etwa 630 kupferhaltige Minerale bekannt.³¹ Einige Beispiele sind in Tabelle 1 aufgeführt. Die sulfidischen Minerale Chalkosin (Cu_2S) und Chalkopyrit (CuFeS_2) sowie die Minerale Cuprit (Cu_2O) und Malachit $\text{Cu}_2(\text{CO}_3)(\text{OH})_2$ gehören zu den wirtschaftlich relevanten Kupfererzen. Für die Produktion von Kupfermetall ist das häufig auftretende Chalkopyrit das wichtigste Erz.¹⁷ Kupfererze treten in Lagerstätten verschiedenster Typen auf, die sich vor allem im Erzgehalt, der Tonnage, den Begleitmetallen und dem geologischen Umfeld stark voneinander unterscheiden können. Daneben haben die mechanischen, mineralogischen und physikalischen Eigenschaften der Kupfererze und Begleitgesteine Auswirkungen auf die Bauwürdigkeit einer Lagerstätte.¹² Der Begriff der Bauwürdigkeit beschreibt dabei die Rentabilität des Erzabbaus.³² Kupferlagerstätten können magmatischen oder hydrothermalen Ursprungs sein oder sie sind ein Resultat aus Sedimentation und Verwitterung. Die meisten Kupferlagerstätten gehören zu den hydrothermalen Depots, wie beispielsweise die vulkanogenen massiven Sulfidlagerstätten (VMS, von engl.: volcanic massive sulfide deposit).¹⁷ Diese Lagerstätten entstehen durch die Exhalation von Sulfiden aus Schwarzen Rauchern am Meeresboden der Tiefsee. Die mit zum Beispiel Kissenlaven bedeckten Erzhügel unterhalb der Schwarzen Raucher bilden die Lagerstätten. Dieser massive Sulfidkörper ist der obere Abschluss der sogenannten Stockwerkzone. Das sind Kluftbereiche, die durch den Aufstieg von hydrothermale Wasser mit Erzen gefüllt wurden.¹⁷ Weltweit gibt es zahlreiche VMS-Lagerstätten. Sie unterscheiden sich in ihren Charakteristika und können beispielsweise aufgrund der Typlokalität klassifiziert werden. Eine solche Typlokalität ist der Zypern-Typ VMS. Hier sind die Schwarzen Raucher im Basalt der Kissenlaven im Troodos-Ophiolith, einem Teil ozeanischer Lithosphäre, das nach seiner Entstehung in der Kreidezeit auf einen Kontinentrand geschoben wurde, lokalisiert.^{17,33} Auf Zypern gibt es zahlreiche VMS-Lagerstätten, deren Abbau bereits zur Bronzezeit begann. In der Neuzeit existierten 24 Kupferminen. Heute wird noch in einer Mine aktiv das Kupfererz Chalkopyrit abgebaut.^{15,28,34}

Sedimentäre Kupferlagerstätten, die durch Verwitterung und Sedimentation entstehen, stellen die zweithäufigste Quelle für Kupfererze dar.¹² Zu diesen Depots gehören die stratiformen sedimentgebundenen Kupferlagerstätten (SSC, von engl.: stratiform sediment-hosted copper deposits), wie zum Beispiel der in Mitteleuropa auftretende Kupferschiefer. Als Kupferschiefer wird hier eine Ablagerungsschicht des Zechsteinmeeres bezeichnet. Der Schwarzschiefer ist ein mit organischem Material reicher Tonstein. Die Anreicherung des Schwarzschiefers mit Kupfer und seinem Begleitmetallen erfolgte während der Diagenese, also der Verfestigung der lockeren Sedimente. In Mitteleuropa ist das polnische Revier Lubin das bedeutendste Vorkommen. Mit seinen

drei aktiven Minen gehört es weltweit zu den wichtigsten Lagerstätten für Kupfer. Neben dem Kupfererz Chalkopyrit, finden sich in der Kupferzone des polnischen Kupferschiefers auch die kupferhaltigen Minerale Chalkosin und Bornit.^{17,35}

Weltweit werden die Kupfererze im industriellen Bergbau gefördert. Der Abbau erfolgt dabei, aufgrund der Erzgehalte, fast ausschließlich im Tagebau. Ein Abbau Untertage findet häufig erst dann statt, wenn der Tagebau weit fortgeschritten ist. Vor Ort wird das Erz in der Regel durch Sprengungen gelöst und einer Vorzerkleinerung unterzogen. Die weitere Erzaufbereitung richtet sich nach der Mineralogie der Lagerstätte und Erze. Sie erfolgt entweder pyrometallurgisch für sulfidische Erze oder hydrometallurgisch in Kombination mit Laugung für oxidische Erze.^{12,17}

2.3 Seltene Erden

Die Geschichte der Seltenen Erdelemente begann 1794 mit der Entdeckung des Elements Yttrium. 1945 wurde mit Promethium das letzte der 17 SEE beschrieben. Zur Gruppe der Seltenen Erden zählen die 15 Elemente der Lanthanide, sowie Scandium und Yttrium.^{7,15,28,36} Üblicherweise werden die SEE aufgrund ihrer Atomgewichte in schwere und leichte SEE eingeteilt (Tabelle 2). Die leichten Seltenen Erden umfassen alle Elemente von Lanthan bis Europium, während die Elemente Gadolinium bis Lutetium sowie Yttrium die Gruppe der schweren SEE bilden.¹⁴

Tabelle 2: Einteilung der Lanthanide und Yttrium in leichte und schwere SEE.

Leichte Seltene Erdelemente								
La	Ce	Pr	Nd	Pm	Sm	Eu		
Lanthan	Cer	Praesodym	Neodym	Promethium	Samarium	Europium		
Schwere Seltene Erdelemente								
Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Y
Gadolinium	Terbium	Dysprosium	Holmium	Erbium	Thulium	Ytterbium	Lutetium	Yttrium

Rar sind die meisten SEE nicht. Entgegen ihres Namens kommen sie durchschnittlich genauso häufig vor wie Kupfer. Die SEE liegen in der Natur jedoch nicht gediegen vor. Sie sind vielmehr, aufgrund ihres ähnlichen chemischen Verhaltens, in komplexen Mineralen vergesellschaftet.^{7,14} Heute sind etwa 245 SEE-haltige Minerale bekannt. Darunter vor allem Karbonate, Oxide, Phosphate und Silikate. Zu den wirtschaftlich wichtigsten Mineralen zählen vorrangig Bastnäsit $(\text{Ce, La, Y, Nd})\text{CO}_3\text{F}$, ein Karbonatmineral sowie die Phosphate Monazit $(\text{La, Ce, Nd, Sm, Th})\text{PO}_4$ und Xenotim $(\text{Y, Yb})\text{PO}_4$. Die Komplexität der Minerale wird durch das Auftreten vieler verschiedener SEE in unterschiedlicher Zusammensetzung und die Assoziation mit radioaktiven Elemente wie Thorium oder Uran in einem Mineral ausgemacht. Obwohl die Minerale mehrere SEE enthalten, tendieren sie dazu, entweder nur leichte oder nur schwere SEE in ihr Kristallgitter einzubauen. Die Minerale Bastnäsit und Monazit sind eine Quelle für hauptsächlich leichte SEE, wie Cer, Lanthan oder Neodym. Schwere SEE, wie Yttrium und Ytterbium, sind stärker in Xenotim angereichert.^{7,14,17,28,37} Trotz ihres vergleichsweise häufigen Vorkommens in der Erdkruste, sind die SEE selten in abbauwürdigen Lagerstätten vorzufinden.

Neben Lagerstätten, die durch hydrothermale, magmatische oder metamorphe Prozesse gebildet wurden, gibt es auch Depots, die durch Verwitterung und Erosion entstanden sind. SEE-Lagerstätten treten in den unterschiedlichsten geologischen Umgebungen auf, sodass sie nur schwer klassifiziert

werden können. Zu den bedeutendsten Lagerstätten zählen die Karbonatit-Intrusionen. Sie sind magmatischen Ursprungs und enthalten vorrangig die Minerale Bastnäsit und Monazit. Ähnlich wie in den Mineralen variiert auch die Zusammensetzung der Selten Erden-Lagerstätten stark: Die leichten SEE treten deutlich häufiger auf, als die schweren SEE. Eine Quelle für schwere SEE sind die sogenannten Ionenadsorptionstone, an deren Oberfläche die SEE adsorbiert sind. Peralkaline Lagerstätten, wie sie beispielsweise in Kanada und Schweden vorkommen, sind derzeit nicht bauwürdig, aufgrund der komplexen Aufbereitung der SEE. Sie weisen vor allem ein großes Potential an schweren SEE auf. Der Abbau von SEE-Lagerstätten erfolgt hauptsächlich in industriellen Tagebauen. Die Aufarbeitung der abgebauten Minerale ist, aufgrund der Ähnlichkeit der SEE, aufwändig und erfordert zahlreiche Schritte, die von den jeweiligen enthaltenen SEE abhängen. Sie umfasst physikalische Anreicherungsverfahren, gefolgt von einer Konzentrierung mittels chemischer Methoden. Wenn möglich, wird auf die vollständige Aufarbeitung zu reinen SEE-Metallen oder -oxiden zugunsten von Mischmetallen verzichtet.^{7, 14, 36, 38} Neben dem Vorkommen der SEE in Lagerstätten, finden sie sich auch als Spurenelemente in Böden, in tierischen Zellen und in Pflanzen. Die Verfügbarkeit der Seltenen Erden in diesen Teilen der Umwelt ist, in Abhängigkeit vom Ausgangsmaterial, durch geochemische und biologische Prozesse bedingt. Von besonderer Relevanz ist dabei die Verwitterung der SEE-haltigen Minerale. Dadurch gelangen die SEE primär in die Böden, können aber auch in neue Minerale eingebaut werden. Aus den Böden nehmen die Pflanzen die SEE auf. Durch die pflanzliche Nahrungsaufnahme von Tieren gelangen sie in tierische Zellen.^{39, 40}

2.4 Laserinduzierte Breakdownspektroskopie

Die laserinduzierte Breakdownspektroskopie ist Teil der optischen Elementanalytik. Sie beruht auf der Emission angeregter Spezies und erlaubt dadurch die qualitative und quantitative Bestimmung der Elementzusammensetzung zahlreicher Probenmaterialien.^{41, 42, 43} Die Entstehung dieser spezifischen Laserablationmethode ist eng an die Entwicklung des Lasers gebunden. Aufgrund der Wechselwirkung eines Laserpulses mit der zu analysierenden Probe wird ein Plasma induziert, das die probenspezifischen Spektrallinien emittiert.^{41, 42} 1960 konnte erstmals ein Laser vorgestellt werden. Dessen Weiterentwicklung führt wenige Jahre später zu einem Laser, der erstmals ein analytisch nutzbares Laserplasma an einer Oberfläche erzeugen konnte.⁴¹ Mit den Jahren wuchs das Interesse an der LIBS. Anfänglich lagen das Augenmerk vorrangig auf der Grundlagenphysik der Plasmabildung. Mit der Zeit eröffneten jedoch die technologischen Weiterentwicklungen der einzelnen LIBS-Komponenten, wie Laser, Spektrometer und Detektoren neue Messmöglichkeiten und Anwendungsfelder.⁴¹ Innerhalb kurzer Zeit entwickelte sich die qualitative Methode zu einer quantitativen Technik weiter.⁴² Heute ist die LIBS eine etablierte Analysemethode sowohl in wissenschaftlichen, als auch in technischen Bereichen. Neben Grundlagenforschung zu Plasmaphysik und -spektroskopie, findet sie in zahlreichen analytischen Fragestellungen Anwendung. Zum Einsatz kommt die LIBS beispielsweise in den Bereichen der Industrie und Forschung, Forensik, Geochemie, Biomedizin, Pflanzen- und Bodenkunde. Aber auch in der Fernerkundung sowie in der Weltraumerforschung wird die Technik der laserinduzierten Breakdownspektroskopie genutzt. Beispielsweise bei der Erforschung des Mars, bei der die LIBS Teil der Marsrover „Curiosity“ und „Perseverance“ ist, die unter anderem Boden- und Gesteinsanalytik vornehmen.^{44, 45, 46} Der Einsatz der LIBS-Spektroskopie in vielfältigen Anwendungen ist durch die zahlreichen Vorteile der Methode begründet. Insbesondere die Möglichkeit Gase, Flüssigkeiten und Feststoffe ohne Einbußen in der Qualität messen und dabei simultan alle enthaltenen Elemente detektieren zu können, macht die

Technik für ein breites Anwendungsspektrum interessant. Die schnelle oder gar Echtzeitanalyse kann on-line oder in situ erfolgen. Dafür sind keine speziellen Messbedingungen und keine oder lediglich eine minimale Probenvorbereitung erforderlich. Die Messung kann orts- und tiefenaufgelöst erfolgen. Das analysierte Material wird dabei minimal verbraucht. Problematisch ist die quantitative Analytik von Elementen, bedingt durch Matrixeffekte und die Notwendigkeit von Referenzmaterialien für eine Kalibrierung. Ein weiteres Manko ist, dass keine molekularen Informationen erhalten werden können, da durch die hohen Plasmatemperaturen alle chemischen Bindungen aufgebrochen werden. Allerdings ist es möglich die LIBS mit anderen Techniken, z.B. der Raman-Spektroskopie zu kombinieren und somit Zugang zu elementaren als auch zu molekularen Informationen zu erhalten.⁴⁷ Weiterhin ist es möglich das LIBS-Messsystem zu miniaturisieren und die Technik in portablen Geräten einzusetzen.^{41, 42, 43, 48}

2.4.1 Prinzip

Das Messsystem der LIB-Spektroskopie ist im Wesentlichen, wie in Abbildung 1 dargestellt, aus drei Komponenten aufgebaut: einem Lasersystem (1), Lichtleitungs- und Sammeloptiken (2) und einer spektral aufgelösten, optischen Detektions-einheit (3). Die spektroskopische Analyse der Elementzusammensetzung der Probe erfolgt mittels eines fokussierten Laserstrahls. Dieser wird auf die Probe gerichtet. Ist die Laserenergie pro Fläche hoch genug, wird die Probenoberfläche so stark erhitzt, dass Probenmaterial verdampft und ein Plasma gezündet wird. Das Plasma erscheint als ein sehr heller, weißer Blitz und wird von einer akustischen Emission begleitet. Die entstandene Plasmawolke emittiert Licht, das zunächst gesammelt und spektral zerlegt wird. Anschließend wird es detektiert und als Spektrum der Intensität gegenüber der Wellenlänge von einer Computereinheit ausgegeben.^{41, 42, 44, 48}

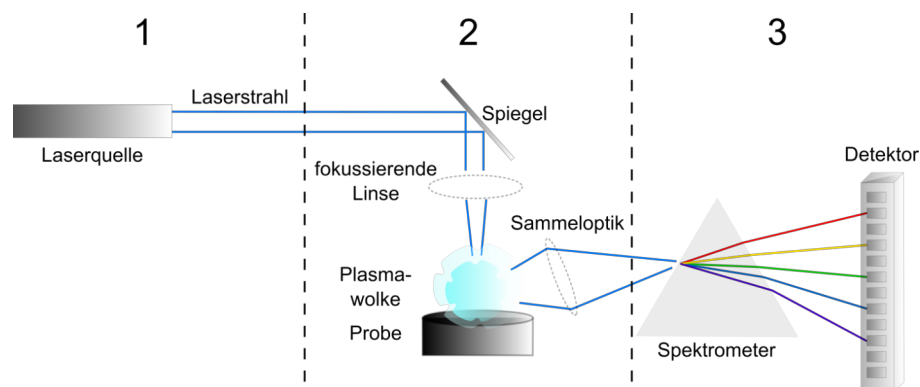


Abbildung 1: Schematischer Aufbau des Messsystems der LIB-Spektroskopie mit 1 Lasersystem, 2 Lichtleitungs- und Sammeloptiken und 3 optischer Detektoreinheit, nach SciAps⁴⁶

In der LIB-Spektroskopie werden unterschiedliche Typen von Lasern genutzt, von UV-Excimer-Lasern bis hin zu Infrarot-Festkörper-Lasern. Dabei beeinflussen die verschiedenen Absorptionscharakteristika der Laserstrahlung während der Plasmabildung das Verhalten des resultierenden Plasmas. Am häufigsten werden als Laserquellen in der LIBS gepulste Nd:YAG-Laser mit einer Wellenlänge von $\lambda = 1064 \text{ nm}$, einer Pulsdauer von 1 bis 20 ns und Pulsenergien von 5 bis 150 mJ verwendet.^{42, 49}

Als Detektoren werden in der LIB-Spektroskopie hauptsächlich Arraydetektoren genutzt, die auf der CCD- oder ICCD-Technologie (von engl.: (intensified) charge-coupled device) basieren. Das

ladungsgekoppelte Halbleiter-Bauelement (CCD) detektiert Photonen, wandelt sie in Photoelektronen um und detektiert sie als elektrische Ladungen. In der ICCD-Technologie werden verstärkte CCD-Sensoren genutzt. Der Verstärker vervielfacht die Photoelektronen und besitzt die Fähigkeit die Blende sehr schnell zu schließen, während bereits detektierte Photonen weiter umgewandelt werden. Dadurch sind ICCD-Sensoren für Anwendungen mit Einzelphotonendetektion und Belichtungsdauern im Subnanosekundenbereich, wie der LIBS, interessant.^{41, 50}

Die folgenden Betrachtungen der Plasmabildung und -charakteristika werden am Beispiel einer festen Probe vollzogen, da in der vorliegenden Arbeit ausschließlich mit Feststoffen als Probenmaterial gearbeitet wurde.

2.4.2 Plasma und Breakdown

Die Wechselwirkung des Laserstrahls mit der Probe sind zentraler Bestandteil der LIB-Spektroskopie und Grundlage für die Bildung eines analytisch nutzbaren Plasmas. Aus der Wechselwirkung resultiert die Zündung des Plasmas und damit verbunden die Atomisierung und Anregung des untersuchten Probenmaterials sowie die Emission charakteristischen Lichts. Betrachtet man diesen Schritt der LIBS im Detail, kann folgender zeitlicher Ablauf der Plasmaentwicklung (Abbildung 2) beobachtet werden. Wenn ein Laserstrahl mit ausreichend hoher Energie mit einer Probenoberfläche interagiert, absorbiert das Probenmaterial die zugeführte Energie.⁵¹ Aufgrund der hohen Energiedichte kommt es zu einer sehr schnellen und lokalen Erwärmung und infolgedessen zu einer Ablation des Probenmaterials. Durch einen weiteren Energieeintrag mittels des Laserstrahls wird das ablatierte Material verdampft, weiter erwärmt und atomisiert. Das Plasma erreicht Temperaturen typischer-

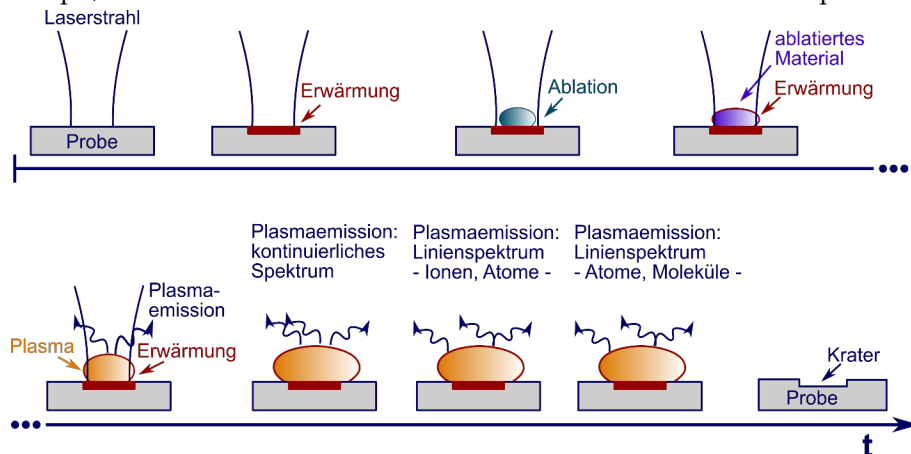


Abbildung 2: Zeitlicher Verlauf der Plasmabildung in der LIB-Spektroskopie unter Einwirkung eines Laserstrahls auf eine Oberfläche eines Feststoffs, nach Bohling.⁴⁹

weise in der Größenordnung von 10^3 K bis 10^5 K. Die gebildeten Atome erfahren durch die zugeführte Laserenergie eine Anregung und werden teilweise ionisiert. Durch diese Initialionisation entstehen freie Elektronen. Die Absorption weiterer Laserenergie führt zu einer Beschleunigung dieser Elektronen.⁵⁰ Diese Kaskadenionisation resultiert in einem Plasma. Die Gesamtheit der Bildungsprozesse des Plasmas wird als Breakdown bezeichnet. Das Plasma selbst ist ein gasförmiges Gemisch, in dem sich neutrale Atome und Moleküle in Grund- und angeregten Zuständen, positiv und negativ geladene Ionen in Grund- und angeregten Zuständen sowie freie Elektronen befinden. Es tritt als weißer Blitz auf, der durch einen hörbaren Knall begleitet wird. Dieser resultiert aus einer Druckwelle, die durch die Plasmabildung entsteht. Erreicht die Druckwelle die Schallgeschwindigkeit wird ein lauter Knall erzeugt. Das entstandene Plasma dehnt sich in alle Richtungen aus. Die Zufuhr weiterer

Energie wird mit der Beendigung des Laserpulses eingestellt. Die Abkühlphase des Plasmas setzt während der Expansionsphase ein und es kommt zur Emission charakteristischer, elektromagnetischer Strahlung durch die angeregten Plasmateilchen. Die Abkühlphase endet mit dem Zerfall des Plasmas. Nach Beendigung der LIBS-Messung resultiert aus der Ablation des Probenmaterials eine kleine kraterförmige Vertiefung. In Abhängigkeit von der Laserenergie und dem untersuchten Material weist sie einen Durchmesser im Mikrometerbereich auf.^{35, 36, 43, 46, 47}

Das Emissionsspektrum wird in den ersten Nanosekunden nach dem Breakdown durch die kontinuierliche Untergrundstrahlung dominiert (Abbildung 3). Diese Kontinuumsstrahlung tritt

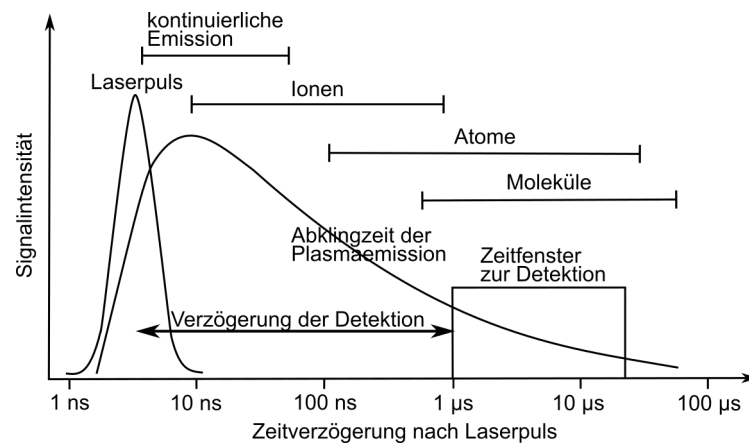


Abbildung 3: Zeitliche Entwicklung des Plasmas in Abhängigkeit von der Signalintensität in der LIB-Spektroskopie, nach Telle³⁹

durch die Brems- und Rekombinationsstrahlung des Plasmas auf. Werden die freien Elektronen durch das elektrische Feld der positiv geladenen Plasmateilchen abgebremst, emittieren sie elektromagnetische Bremsstrahlung. Gleichzeitig rekombinieren die freien Elektronen mit den Ionen des Plasmas während der Abkühlungsphase und senden dabei Rekombinationsstrahlung aus. In dieser Zeit können ausschließlich sehr starke Emissionslinien beobachtet werden. Die Emission charakteristischer elektromagnetischer Strahlung, die für die Spektralanalyse relevant ist, resultiert aus Übergängen der Plasmateilchen aus energetisch höheren Zuständen in energetisch niedrigere Zustände. Die Relaxation der Atome führt dabei zu diskreten Linienspektren, zunächst von Ionen und Atomen, später auch von Molekülen (Abbildung 3). Die unspezifische Brems- und Rekombinationsstrahlung weist eine kürzere Abklingzeit auf, als die charakteristischen Elementemissionen. Dadurch kann eine zeitaufgelöste Detektion das wenig aussagekräftige Weißlicht von den Messungen der charakteristischen Linienspektren separiert werden. Oft verbessern sich bereits nach etwa einer Mikrosekunde das Signal zu Untergrund- und das Signal zu Rauschen-Verhältnis um ein Vielfaches, sodass die analytisch relevanten Emissionslinien wesentlich schärfer werden. Typischerweise erfolgt die Detektion bei einem LIBS-Experiment etwa 1 µs nach dem Laserpuls (Abbildung 3).^{41, 42, 49, 52, 53}

2.4.3 Emissionsspektrum

In einem Emissionsspektrum wird die detektierte Signalintensität gegen die Wellenlänge in Nanometer aufgetragen. Ein durch die LIB-Spektroskopie erhaltenes Emissionsspektrum weist typischerweise eine Vielzahl an Spektrallinien auf. Die Auswertung eines solchen Spektrums gewährt Informationen zum untersuchten Probenmaterial sowie zum Plasma.

Informationen über die qualitative und quantitative elementare Zusammensetzung des untersuchten Probenmaterials, erfordern zunächst eine Identifikation der detektierten Emissionslinien. Über die Lage der Emissionslinien und ihre Intensität können den Linien die entsprechenden Elemente mit Hilfe von Datenbanken zugeordnet werden. Eine der am häufigsten genutzten und gleichzeitig vollständigsten Datenbanken für Emissionslinien ist die Datenbank des Nationalen Instituts für Normen und Technologien (NIST, von engl.: national institute of standards and technology).^{42, 54} Die Lage einer Emissionslinie im Spektrum resultiert unmittelbar aus dem elektronischen Übergang eines angeregten Teilchens im Plasma in einen energetisch niedrigeren Zustand. Die bei dem Übergang emittierte Strahlung weist eine definierte Frequenz auf, die durch die Energiedifferenz der Zustände bestimmt ist. Die Frequenz und damit auch die Wellenlänge ist charakteristisch für den Übergang und kann einem chemischen Element zugeordnet werden. Um ein chemisches Element anhand eines LIB-Spektrums sicher zu identifizieren, ist typischerweise das Vorhandensein von mindestens zwei Linien hoher Intensität dieses Elements erforderlich.⁴² Durch die Zuordnung der Elemente zu den Spektrallinien werden Informationen über die in der untersuchten Probe enthaltenen Elemente gewonnen, die qualitative Aussagen über die elementare Zusammensetzung des Materials erlauben.

Angaben zur quantitativen Zusammensetzung eines Probenmaterials können ebenfalls mittels der LIBS erhalten werden. Dies kann entweder durch den Einsatz von Kalibrationskurven oder durch die kalibrationsfreie Methode (CF-LIBS, von engl.: calibration-free LIBS) realisiert werden. Die in dieser Arbeit verwendete Methode der Kalibrationskurven ist in der analytischen Chemie auch als Standardadditionsmethode bekannt. Die Methode erfordert Referenzproben, die das zu analysierende Element in bekannten unterschiedlichen Konzentrationen beinhalten. Die Auftragung der gemessenen Intensität oder -fläche des Elementemissionssignals als Funktion der Konzentration erlaubt eine Anpassung der Messdaten. Die Konzentration der unbekannt Probe kann anschließend durch Interpolation ermittelt werden.⁴² Die CF-Methode benötigt dahingegen keine Referenzmaterialien. Sie basiert auf dem Ansatz, dass das LIB-Spektrum alle zur Bestimmung der Elementzusammensetzung notwendigen Informationen enthält. Mit Hilfe eines mathematischen Modells, das einige Annahmen das Plasma betreffend erfordert, kann die Konzentration einer unbekannt Probe ermittelt werden. Dabei werden die Spektrallinien aller Probenbestandteile detektiert und Boltzmann-Plots (Gleichung 1) entwickelt. Die Schnittpunkte der Boltzmann-Plots erlauben die Berechnung der Konzentrationen der Probenbestandteile.^{42, 55}

Die quantitative LIB-Spektroskopie wird vor allem durch sogenannte Matrixeffekte erschwert. Diese beeinflussen die Kalibrierung und damit die Ermittlung der Konzentration und sind für die Verschlechterung der Messgenauigkeit in LIBS verantwortlich. Matrixeffekte treten auf, wenn die physikalischen und chemischen Eigenschaften das Elementsignal beeinflussen. Die Veränderung des Elementsignals resultiert in Ungenauigkeiten der Elementkonzentration, obwohl diese unverändert ist. Die physikalischen Matrixeffekte resultieren aus den physikalischen Eigenschaften des Probenmaterials. Die quantitative LIBS kann beispielsweise durch Probenbeschaffenheiten, wie verschiedene Korngrößen, Inhomogenitäten oder Feuchtigkeit beeinflusst werden. Die physikalische Matrixeffekte stehen zudem in enger Korrelation mit dem Ablationsprozess der LIB-Spektroskopie. Sie können zu unterschiedlich starken Ablationen des Materials führen, obwohl die Charakteristika des Ablationsprozesses an sich gleich bleiben. Chemische Matrixeffekte betreffen die chemischen Eigenschaften, wie zum Beispiel die Zusammensetzung der Probe. Sie bezeichnen das Phänomen, dass Elemente die Emissionen eines anderen Elements beeinträchtigen können. Die Kalibrierung und damit auch die fehlerfreie Bestimmung von Elementgehalten im untersuchten Probenmaterial ist

hierdurch erschwert. In der kalibrierungsfreien LIB-Spektroskopie werden die Matrixeffekte durch das angewendete mathematische Modell intern korrigiert.^{41, 56, 57, 58}

Neben den qualitativen und quantitativen Daten zum Probenmaterial enthält das Emissionsspektrum Informationen zur Plasmacharakteristik. Die wichtigsten Parameter zur Beschreibung des Plasmas sind die Plasmatemperatur, die Plasmaelektronendichte und die Linienprofile der Emissionslinien. In einem Plasma laufen zahlreiche Prozesse unter Beteiligung von Elektronen und Photonen ab: Emission, Photoionisation, Rekombination, Absorption, Relaxation und Anregung. Fänden diese Prozesse in einem thermodynamischen Gleichgewicht statt, würden die Prozesse genauso oft auftreten, wie ihre umgekehrten Prozesse. Dadurch wären alle Prozesse ausgeglichen. Damit wäre es möglich, das Plasma anhand der Plasmatemperatur und -elektronendichte vollständig zu beschreiben. Kann keine konstante Temperatur erreicht werden, wie bei einem laserinduzierten Plasma, existiert ein Temperaturgradient. Das Plasma ist inhomogen und die Betrachtungen des gesamten Plasmas im thermodynamischen Gleichgewicht sind unbrauchbar. Denkbar ist jedoch, dass sich sehr kleine räumliche Bereiche im Plasma in einem Gleichgewicht befinden. Für diese Bereiche wird ein lokales thermodynamisches Gleichgewicht (LTE, von engl.: local thermodynamic equilibrium) angenommen. Bedingung für diese Annahme ist, dass die Kollisionsrate innerhalb des Plasmas größer ist als die Strahlungsrate. Innerhalb dieser Bereiche kann von einer konstanten Temperatur und Elektronendichte ausgegangen werden.^{42, 53, 59}

Innerhalb eines Plasmas existieren verschiedene Temperaturen, die aber im Fall des LTE lokal ausgeglichen sind. Um die Plasmatemperatur zu bestimmen, ist die Betrachtung einer dieser Temperaturen ausreichend. Üblicherweise wird in der LIB-Spektroskopie die Anregungstemperatur des Plasmas stellvertretend für die Plasmatemperatur ermittelt. Mittels der Boltzmann-Plot-Gleichung (Gleichung 1)⁶⁰ kann die Temperatur T des Plasmas bestimmt werden.

$$I = h \cdot \nu_{ij} \cdot n \cdot l \cdot \frac{g_i A_{ij}}{Q(T)} \exp\left(-\frac{E_i}{k_B \cdot T}\right) \quad (1)$$

Hier ist I (Wm^{-2}) die Intensität, h (Js) die Planck-Konstante, ν_{ij} (s^{-1}) die Übergangsfrequenz, n (m^{-3}) die Dichte der Spezies, l (m) die optische Weglänge, A_{ij} (s^{-1}) die Übergangswahrscheinlichkeit nach Einstein, g_i (dimensionslos) die Entartung des oberen Energieniveaus und $Q(T)$ (dimensionslos) ist die Zustandssumme der betrachteten Spezies. Die Energie des oberen Energieniveaus E_i (J), die Boltzmann-Konstante k_B (JK^{-1}) und die Temperatur T (K) sind Teil des letzten Terms der Gleichung 1. Die Indizes i und j beziehen sich auf das obere und untere Energieniveau des Übergangs. Die lineare Form der Gleichung erlaubt die Temperaturermittlung aus der Steigung mit Hilfe einer linearen Regression. Anhand der Intensitätsverhältnisse zweier Linien im Emissionsspektrum kann die Plasmatemperatur bestimmt werden. Betrachtet werden typischerweise zwei Linien eines Elements. Bei der Betrachtung zweier neutraler Linien kann die obige Gleichung 1 wie beschrieben zur Temperaturermittlung herangezogen werden. Werden jedoch eine ionische und eine neutrale Linie genutzt, muss die Boltzmann-Plot-Gleichung um geladene Spezies zur Saha-Boltzmann-Gleichung (Gleichung 2)^{39, 52, 57, 61} erweitert werden.

$$\frac{I_{Ion}}{I_{Atom}} = \frac{2(2\pi m_e k_B T)^{3/2}}{n_e h^3} \cdot \left(\frac{g_A}{\lambda}\right)_{Ion} \cdot \left(\frac{\lambda}{g_A}\right)_{Atom} \exp\left(-\frac{U^+ + E_{Ion} + E_{Atom}}{k_B T_{Ion}}\right) \quad (2)$$

Beschrieben wird hier die Intensität I für ionische (I_{Ion}) und atomare (I_{Atom}) Spezies. Hierbei ist λ (nm) die Wellenlänge, m_e (kg) die Elektronenmasse, n_e (cm^{-3}) die Elektronendichte und E_{Ion} (J) sowie E_{Atom} (J)

stehen für die Anregungsenergien. U^+ (J) ist das Ionisationspotential des Atoms und T_{ion} (K) ist die Ionisationstemperatur. Da die Elektronendichte und die Plasmatemperatur nicht im Voraus bekannt sind, muss die Berechnung iterativ erfolgen. Eine Plasmatemperatur liegt typischerweise im Bereich von $0,5 < T < 2$ eV.^{49, 55, 59}

Die Elektronendichte kann direkt aus den Linienprofilen bestimmter spektraler Linien des Emissionsspektrums bestimmt werden. Die Emissionslinien sind aufgrund von Effekten, die die Spektrallinien verbreitern, nicht unendlich schmal. Statt diskreter Linien bei einer definierten Wellenlänge, treten im Spektrum schmale Banden über mehrere Pikometer auf. Diese verbreiterten Linienprofile lassen sich auf verschiedenen Mechanismen zurückführen. Zu den wichtigsten zählen die natürliche Linien-, die Doppler- und die Druckverbreiterung, sowie die sogenannte Selbstabsorption. Die Linien- und Dopplerverbreiterung haben die geringsten Einflüsse auf Veränderung des Linienprofils in der LIB-Spektroskopie. Der dominierende Prozess der Linienverbreiterung in der LIBS ist die Stark-Verbreiterung, die Teil der Druckverbreiterung ist. Sie beruht auf Wechselwirkungen zwischen den emittierenden Atomen und Ionen und den umgebenden Partikeln. Im Plasma hängt die Auswirkung der Interaktion zwischen den Teilchen von der Art der beteiligten Spezies ab. Neben der van der Waals- und der Resonanzverbreiterung, die bei Kollisionen neutraler Teilchen auftreten, überwiegt in heißen Plasmen mit hoher Elektronendichte der Stark-Effekt. Die emittierenden Ionen erfahren durch die umgebenden Elektronen und Ionen des Plasmas ein elektrisches Feld. Für die emittierenden Ionen variiert dieses Feld statistisch und es schwankt mit der Zeit. Die durchschnittliche Stärke der Beeinflussung in den gebundenen Zuständen bedingt die Gesamtbreite der resultierenden Linienprofile. Die Linienverbreiterung durch den linearen Stark-Effekt tritt beim Wasserstoffatom und wasserstoffähnlichen Atomen auf. Bei allen anderen Teilchen tritt ein quadratischer Stark-Effekt auf. Zusätzlich zur Linienverbreiterung kommt es hierbei zu einer Linienverschiebung zu längeren Wellenlängen.^{49, 52, 53, 59}

Unter der Annahme, dass die Stark-Verbreiterung die Mechanismen dominiert und alle anderen Verbreiterungsmechanismen vernachlässigbar sind, ist die Elektronendichte n_e des Plasmas aus der Stark-Verbreiterung zugänglich. Hierfür wird die Linienbreite der ersten Wasserstofflinien der Balmerreihe ($H\alpha$ 656,28 nm) ermittelt. Aus der Halbwertsbreite der $H\alpha$ -Linie kann mittels Gleichung 3^{42, 59} die Elektronendichte iterativ bestimmt werden.

$$n_e = 8,02 \cdot 10^{12} \left(\frac{\Delta \lambda_{1/2}}{\alpha_{1/2}(T, n_e)} \right)^{3/2} \text{ cm}^{-3} \quad (3)$$

Hier ist $\Delta \lambda_{1/2}$ (Å) die ermittelte Halbwertsbreite der gemessenen $H\alpha$ -Linie und $\alpha_{1/2}(T, n_e)$ (Å) ist die tabellierte Halbwertsbreite der $H\alpha$ -Linie⁶² in Abhängigkeit von der Plasmatemperatur T und der Elektronendichte n_e . Beachtet werden muss, dass die $H\alpha$ -Linie nur dann für die Berechnung der Elektronendichte brauchbar ist, wenn die Elektronendichte nicht zu hoch ist ($n_e < 10^{17} \text{ cm}^{-3}$), da sie bei höheren Elektronendichten einer zusätzlichen Verbreiterung durch die Selbstabsorption unterliegen kann.⁴⁹ Typische Elektronendichten in der LIB-Spektroskopie liegen im Bereich von $10^{15} < n_e < 10^{18} \text{ cm}^{-3}$.^{49, 52, 59}

Bei Plasmabedingungen unter denen sehr hohe Elektronendichten auftreten, absorbiert das Plasma seine eigene Emission. Durch die Absorption kommt es zu einem verbreiterten Linienprofil der Spektrallinien. Ein Spezialfall der Selbstabsorption kann zusätzlich das Profil der Linien beeinflussen, wenn die Emission vom heißen Inneren des Plasmas auf dem Weg nach außen, eine kalte Zone unter

Absorption passiert. Spektrallinien, die davon betroffen sind, zeigen in der Linienmitte eine Einwölbung oder ein Plateau. Dieser Effekt wird Selbstumkehr genannt und kann dazu führen, dass nicht zweifelsfrei bestimmt werden kann, ob es sich um eine oder zwei Spektrallinien handelt. Für die Charakterisierung des Plasmas muss gewährleistet sein, dass es sich um ein optisch dünnes Plasma mit geringer Teilchendichte handelt.^{49, 53, 52}

2.4.4 LIBS-Instrumente

Ein LIBS-System besteht, wie oben beschrieben, aus den drei Hauptkomponenten Lasersystem, Lichtleitungs- und Sammeloptiken sowie der optischen Detektionseinheit. Mit steigendem Bekanntheitsgrad der LIBS-Technik wurden zahlreiche LIBS-Systeme mit unterschiedlichen Konfigurationen und Spezifikationen entwickelt.⁶³ In der vorliegenden Arbeit wurden zwei verschiedene Systeme verwendet: Es kamen ein stationäres LIB-Spektrometer und ein mobiles Handheld-LIB-Spektrometer zum Einsatz.

Die optische Detektionseinheit des stationären LIB-Spektrometers besteht aus einem Echelle-Spektrometer in Kombination mit einer ICCD-Kamera. Echelle-Spektrometer werden seit den 1990er Jahren in der LIB-Spektroskopie genutzt. Sie erfüllen simultan zwei Anforderungen der LIBS-Technik, die aus den hochkomplexen Spektren resultieren: Sie decken einen weiten spektralen Bereich ab und haben gleichzeitig eine hohe spektrale Auflösung. Ein Echelle-Spektrometer beruht typischerweise auf zwei kombinierten Dispersionselementen, einem sogenannten Echelle-Gitter und einem Prisma. Das Echelle-Gitter ist ein spezielles Beugungsgitter. Die Gitterfurchen des Echelle-Gitters sind als unsymmetrische, dreieckige Stufen ausgebildet. An den schmalen Vorderseiten, die verspiegelt sind, wird der einfallende Strahl optimal reflektiert. Es ergibt sich ein großer nutzbarer Spektralbereich und zahlreiche Beugungsordnungen. Genutzt werden die höheren Beugungsordnungen, da hier eine effektive Zerlegung des einfallendes Lichts in die Spektrallinien erfolgt und damit eine hohe Auflösung im Spektrum erreicht werden kann.^{47,64} Es kommt jedoch bei hohen Ordnungen zu einem Überlapp zwischen den langen Wellenlängen einer Ordnung mit den kurzen Wellenlängen der nächsten Ordnung. Um diese Überlappung zu vermeiden, wird bei Echelle-Spektrometern ein zweites optisches Element orthogonal zur Beugungsrichtung des Echelle-Gitters eingesetzt. Dies ist üblicherweise ein Prisma, das die Spektren verschiedener Ordnungen trennt. Es resultiert eine zweidimensionale Auftrennung der spektralen Information. Die einzelnen Ordnung des Spektrums können anschließend zeilenweise durch eine ICCD-Kamera ausgelesen werden.^{63, 65, 66, 67, 68} Durch diesen Aufbau der Echelle-Spektrometer können Wellenlängenbereiche von 200 bis 1000 nm mit einer hohen Auflösung untersucht werden.

Mobile, genauer tragbare LIBS-Instrumente erfordern eine Miniaturisierung der Komponenten eines LIBS-Systems. Dies gelang erstmals in den 1990er Jahren, als die ersten tragbaren LIBS-Geräte für die Detektion von Metallkontaminationen auf Oberflächen zum Einsatz kamen. Der entscheidende Vorteil von sogenannten Handheld-LIB-Spektrometern gegenüber stationären LIBS-Systemen ist, dass das Messinstrument zur Probe kommt. Dadurch kann die Probennahme, ein kritischer Schritt in der Analytik, entfallen.⁶⁹ Zudem kann durch den Einsatz eines Handheld-LIB-Spektrometers ein zu untersuchendes Areal zügig zweidimensional kartiert werden. In Verbindung mit einer Probenahme aus dem darunterliegenden Gestein, wie beispielsweise durch Bohrungen, können Elementverteilungen auch in der Tiefe analysiert werden, sodass sich dreidimensionale Elementprofile erstellen lassen. Ein typisches Handheld-LIBS-Gerät weist einen spektralen Bereich

von einigen hundert Nanometern mit einer Auflösung zwischen 0,1 und 0,3 nm auf. Es sind üblicherweise Laser der Klasse 3B mit einer Wellenlänge von 1064 nm, sogenannte Nd:YAG-Laser, verbaut. Im Schnitt wiegt ein Handheld-LIB-Spektrometer etwa 2 kg.^{41, 42, 70}

2.5 Chemometrie

Bei der Analyse von Probenmaterial mit der LIB-Spektroskopie werden komplexe Spektren erhalten, die zahlreiche Informationen über das untersuchte Probenmaterial beinhalten. Um bei der Analyse der chemischen Daten alle relevanten Informationen zu gewinnen, werden in der Chemometrie mathematische und statistische Methoden eingesetzt. Dabei kommen univariate Methoden bei Datensätzen zum Einsatz, deren Objekte lediglich ein Merkmal aufweisen. Beispielsweise kann einer Stoffkonzentration einer untersuchten Substanz (Objekt) genau eine Peakintensität (Merkmal) zugeordnet werden. Sind einem Objekt zwei oder mehrere Merkmale zugehörig, erfolgt die Datenanalyse mittels bi- oder multivariater Techniken.^{71, 72, 73} Im Folgenden werden die in dieser Arbeit genutzten uni- und multivariaten Analysemethoden kurz vorgestellt.

2.5.1 Univariate Analyse

Die univariate Analyse (UVA) beschreibt den Zusammenhang zwischen einem Merkmal und den Objekten des Datensatzes. Das Merkmal wird als abhängige Variable oder auch response Variable (Antwortvariable) bezeichnet. Das Objekt hingegen ist die unabhängige oder predictor Variable (Vorhersagevariable). Ein Teil der UVA ist die Regressionsanalyse, die die Beziehung zwischen der abhängigen und unabhängigen Variable beschreibt sowie eine Abschätzung des Wertes der abhängigen Variable aufgrund der unabhängigen Variable erlaubt. Für die Beschreibung der Beziehung der Variablen entwickelt die UVA ein Regressionsmodell. Im einfachsten Fall kann die Korrelation der Variablen mit einem univariaten linearen Regressionsmodell erklärt werden. Durch eine Regressionsgerade der Form $y = mx + n$ wird hierbei die lineare Beziehung der Variablen widerspiegelt, wobei n der Schnittpunkt mit der y-Achse und m der Anstieg der Geraden ist. Die Regressionsgeraden erlaubt die Vorhersage des Wertes der abhängigen Variable y anhand des Wertes des unabhängigen Variable x . Wie gut das Regressionsmodell die Korrelation beschreibt, kann mit Hilfe des Bestimmungskoeffizienten R^2 (Gleichung 4) ermittelt werden.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

Hier sind y_i die abhängigen Variablen, \hat{y}_i sind die berechneten y_i -Werte, \bar{y}_i ist der Mittelwert von y_i und n ist die Probenanzahl. Der Bestimmungskoeffizient gibt den durch die unabhängigen Variablen x erklärten Anteil der Variation der Antwortvariablen y an. Die Variabilität wird hierbei umso genauer durch das lineare Regressionmodell beschrieben, je größer R^2 ist. Die Werte des Bestimmtheitsmaßes liegen typischerweise zwischen $0 < R^2 < 1$.^{73, 74, 75, 76}

Um die Anpassungsgüte des Regressionsmodells vollständig zu beschreiben, wird neben dem Bestimmungskoeffizienten R^2 die Wurzel der mittleren Fehlerquadratsumme (RMSE, von engl.: root mean square error) (Gleichung 5) angegeben, wobei $\hat{y}_i - y_i$ die Differenz zwischen den vorhergesagten abhängigen Variablen und den abhängigen Variablen ist.⁷⁷

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

2.5.2 Hauptkomponentenanalyse, eine multivariate Analyseverfahren

Die Auswertung von Datensätzen, insbesondere komplexer und großer Datenmengen, die beispielsweise durch den Einsatz von Spektrometern generiert werden, ist mit univariaten Analysemethoden häufig nicht zu bewältigen. Hierfür werden multivariate Analysetechniken eingesetzt. Eines der gängigsten statistischen Verfahren zur Analyse komplexer Datensätze ist die Hauptkomponentenanalyse (PCA, von engl.: principal component analysis). Sie wird hauptsächlich zur Reduktion der Datendimensionalität und zum grafischen Erkennen von Mustern, Gemeinsamkeiten und Unterschieden in Datensätzen genutzt. 1901 von Karl Pearson entwickelt, kommt die PCA heute in zahlreichen Anwendungen in Industrie und Forschung zum Einsatz. Sie ist Teil der explorativen Datenanalyse und dient der Extraktion von Informationen und dem Entfernen von Rauschen aus den experimentell gewonnenen Daten. Die Methoden der explorativen Datenanalyse können in zwei Kategorien eingeteilt werden. Eine Gruppe umfasst die sogenannten unüberwachten (unsupervised) Methoden. Für die Entwicklung der multivariaten Modelle, die diesen Methoden zugrundeliegen, werden Daten ohne Kennzeichnung genutzt. Das Modell hat keine Informationen, wie zum Beispiel Klassenzuordnungen, zu den Daten. Solche Methoden finden beispielsweise bei der Analyse unklassifizierter Daten Anwendung. Im Gegensatz dazu stehen der Gruppe der überwachten (supervised) Methoden Kennzeichnungen der Daten zur Verfügung, die in die Entwicklung des Modells einfließen. Die Hauptkomponentenanalyse ist eine unüberwachte Methode der explorativen Datenanalyse. Die PCA wird vorrangig mit den Zielen der Datenvereinfachung und -reduktion, Ausreißerermittlung, Variablenselektion sowie Klassifikation angewendet.^{78, 79, 80, 81, 82, 83}

Die Hauptkomponentenanalyse wird vor allem im mehrdimensionalen Raum durchgeführt. Das Prinzip der PCA kann vereinfacht am Beispiel eines zweidimensionalen Raums gezeigt werden (Abbildung 4). Ein experimentell erhaltener Datensatz X von N Proben mit M Variablen kann durch Aufspannen eines zweidimensionalen Raumes zwischen zwei Achsen dargestellt werden (Abbildung 4, links). Geometrisch betrachtet wird eine Linie durch das Zentrum dieses Datenraumes gelegt. Diese Linie läuft entlang der Dimension mit der größten Varianz im Datenraum. Sie wird als erste Hauptkomponente $PC1$ (PC, von engl.: principal component) bezeichnet. Eine zweite Linie durchkreuzt den Datenraum entlang der Dimension mit der zweitgrößten Varianz. Die sogenannte zweite Hauptkomponente $PC2$ wird in einem neuen Koordinatensystem, das durch die Hauptkomponenten $PC1$ und $PC2$ aufgespannt wird, orthogonal zur ersten Hauptkomponente projiziert. Dieses neue Koordinatensystem wird als Score-Plot bezeichnet und visualisiert die Korrelation zwischen den Variablen und den Proben (Abbildung 4, rechts). Datenpunkte, die sich räumlich nahe sind, weisen ähnliche Eigenschaften auf. Sie bilden ein sogenanntes Cluster innerhalb des Score-Plots. Die neuen Dimensionen resultieren aus einer Linearkombination der originalen Variablen und werden im Loadings-Plot dargestellt.^{77, 81, 82, 84}

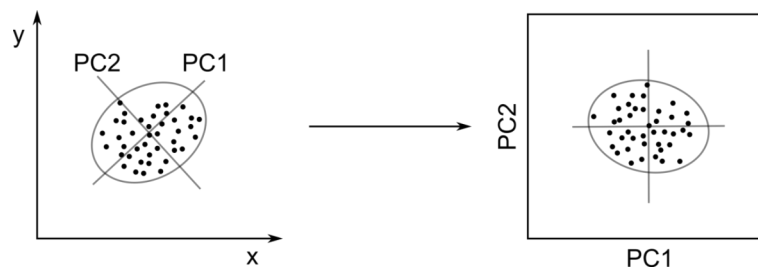


Abbildung 4: Schematische Darstellung des Prinzips der Hauptkomponentenanalyse (PCA) im zweidimensionalen Raum mit den Hauptkomponenten PC1 und PC2 (links) und im Score-Plot (rechts), nach Gauglitz *et al.*⁸³

Etwas formaler betrachtet, wird die Originalmatrix der Daten X , bestehend aus Anzahl der Proben N und Anzahl der Variablen M , in die drei Matrices S , L^T und F zerlegt (Abbildung 5). Die $N \times K$ Score-Matrix S wird mit der $K \times M$ Loadings-Matrix L^T multipliziert. Addiert wird die $N \times M$ Fehlermatrix F . Dabei ist K die Anzahl der Hauptkomponenten.

$$\begin{array}{|c|c|} \hline & M \\ \hline X & \\ \hline N & \end{array} = \begin{array}{|c|c|} \hline & K \\ \hline S & \\ \hline N & \end{array} \cdot \begin{array}{|c|c|} \hline & M \\ \hline L^T & \\ \hline K & \end{array} + \begin{array}{|c|c|} \hline & M \\ \hline F & \\ \hline N & \end{array}$$

Abbildung 5: Originalmatrix X der Hauptkomponentenanalyse mit N Proben und M Variablen zerlegt in die $N \times K$ Score-Matrix S , $K \times M$ Loadingsmatrix L^T und die $N \times M$ Fehlermatrix F , nach Frisvad.⁸⁰

Die Hauptkomponentenanalyse ist eine relativ einfache multivariate Methode, die es erlaubt mehrdimensionale Daten direkt in einem niedrigdimensionalen Raum mit nur geringem Verlust der Originalinformation abzubilden. Dadurch führt sie häufig auch zu einer vereinfachten Interpretierbarkeit der Daten. Typischerweise kann die Varianz im Datensatz zu einem großen Teil durch wenige Hauptkomponenten erfasst werden. Die Auftragung der Daten im Score-Plot der Scores PC1 und PC2 enthält daher bereits zahlreiche Informationen zum untersuchten Datensatz. Die Hauptkomponenten werden auch genutzt, um Klassifizierungs- und Regressionsprobleme wie beispielsweise in der PLS-Regression (2.5.3) zu lösen.⁷⁷

Robuste Hauptkomponentenanalyse

Ein Ziel der klassischen Hauptkomponentenanalyse ist die Ermittlung von Ausreißern. Wie oben beschrieben, spiegelt die erste Hauptkomponente die Dimension mit der größten Varianz der Daten wieder. Die zweite Hauptkomponente, orthogonal zur ersten, entspricht der zweitgrößten Varianz im Datenraum. Bei der PCA wird dieses Vorgehen solange fortgeführt, bis die Hauptkomponenten mit den Eigenvektoren der empirischen Kovarianzmatrix übereinstimmen. Der Bezug zur Kovarianzmatrix ist für die Beobachtung von Ausreißern mittels der PCA problematisch, da diese empfindlich auf Ausreißer reagiert. Infolgedessen sind die Varianzen der regulären Beobachtungen in den ersten Hauptkomponenten verzerrt. Eine Möglichkeit die PCA dennoch zur Ermittlung von Ausreißern einsetzen zu können, ist die sogenannte robuste Hauptkomponentenanalyse (ROBPCA, von engl.: robust principal component analysis). Sie wird in der vorliegenden Arbeit zum Zweck der Ermittlung von Ausreißern genutzt. Die ROBPCA wurde 2005 von Hubert *et al.* vorgestellt. Dabei werden die Projektionsnachverfolgung (PP, von engl.: projection pursuit) und die robuste Kovarianzschätzung basierend auf der minimalen Kovarianzdeterminante (MCD, von engl.: minimum covariance determinant) in einem niedrigdimensionalen Raum kombiniert. Die mittels ROBPCA

ermittelten Datenpunkte, können anhand ihres Typs klassifiziert und anschließend in einer sogenannten Ausreißerkarte dargestellt werden (Abbildung 6). Diese Karte wird durch den Abstand der Scores (SD, von engl.: score distance) auf der x-Achse und den orthogonalen Abstand (OD, von engl.: orthogonal distance) auf der y-Achse aufgespannt. Die resultierende Karte kann durch die cutoffs des SD und des OD in 4 Quadranten aufgeteilt werden.

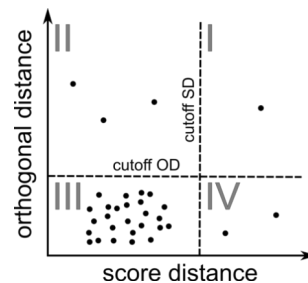


Abbildung 6: Ausreißerkarte der robusten Hauptkomponentenanalyse (ROBPCA). Die Datenpunkte werden im Diagramm in Abhängigkeit vom orthogonalen Abstand (orthogonal distance) und dem Score-Abstand (score distance) aufgetragen. Durch die cutoffs SD und OD resultieren 4 Quadranten, in denen die Punkte klassifiziert werden: I bad leverage-Punkte, II orthogonale Ausreißer, III reguläre Beobachtungen und IV good leverage-Punkte, nach Hubert *et al.*⁸⁴

Jeder der Quadranten enthält einen Typ der Datenpunkte. Im ersten Quadranten werden die sogenannten bad leverage-Punkte aufgetragen. Diese haben einen großen orthogonalen Abstand und sind weit entfernt von den regulären Beobachtungen des Datensatzes. Im zweiten Quadranten werden die orthogonalen Ausreißer gezeigt, die ebenfalls einen großen orthogonalen Abstand zum PCA-Raum haben. Der dritte Quadrant repräsentiert die regulären Beobachtungen. Sie sind eng gruppiert und nahe dem PCA-Raum. Im vierten Quadranten finden sich die good leverage-Punkte, die zwar nahe am Raum der PCA, aber weit entfernt von den regulären Beobachtungen liegen. Als Ausreißer des untersuchten Datensatzes werden die orthogonalen Ausreißer und die bad leverage-Punkte aus dem Datensatz für weitere Analysen eliminiert. Die ROBPCA ist eine schnelle Methode, die einerseits auf symmetrische und andererseits auf unsymmetrische Daten angewendet werden kann.^{85, 86}

2.5.3 PLS Regression, eine multivariate Analyseverfahren

Die partielle Regression der kleinsten Quadrate (PLSR, von engl.: partial least squares regression) ist eine Technik der multivariaten Analyse, die die Beziehung zwischen hochkomplexen Daten untersucht. 1975 von Herman Wold eingeführt, hat sich die Methode zu einer, der am häufigsten genutzten multivariaten Analysetechniken entwickelt. Die PLS-Regression stellt einen Zusammenhang zwischen den Matrizes der unabhängigen X-Daten ($N \times M$) und den abhängigen Y-Daten ($N \times P$) durch ein lineares, multivariates Modell her (Abbildung 7). Gleichzeitig werden die Strukturen der Matrizes X und Y neu modelliert. Die Methode konstruiert durch eine Linearkombination neue Komponenten aus den originalen Variablen M bzw. P. Dazu greift die PLSR auf die Hauptkomponentenanalyse zurück. Zunächst wird je eine PCA der X- und Y-Daten durchgeführt, mit dem Ziel die Hauptkomponenten K der beiden Matrizes zu bestimmen. Dabei wird die Dimension des X-Datenraums ermittelt, die im Y-Datenraum die maximale Varianz erklärt. Dadurch wird die Kovarianz der Matrizes identifiziert, die mittels der PLSR maximiert werden soll. Mit den aus der PCA erhaltenen $N \times K$ Score-Matrizes S und U wird im Anschluss eine lineare Regression durchgeführt. Dabei wird die $N \times P$ -Fehlermatrix der Y-Daten G minimiert. Die optimale

Anzahl an Hauptkomponenten für die Regression wird durch eine Kreuzvalidierung (CV, von engl.: cross-validation) empirisch bestimmt. Durch die Beziehung der Score-Matrices ($U=B \cdot S$) resultiert zusätzlich eine Korrelation zwischen den Originalmatrices X und Y . Die Stärke der PLSR ist es, Daten mit stark korrelierten, verrauschten und vielen X -Variablen zu analysieren und gleichzeitig mehrere Y -Variablen zu modellieren. Dabei gewinnen die Modellparameter an Präzision, je größer die Zahl an relevanten Variablen und Beobachtungen ist. Dadurch ist das PLSR-Modell, insbesondere für große Datensätze, gegenüber Matrixeffekten und äußeren Einflüssen robuster als beispielsweise die univariate Regression.^{71, 77, 87, 88, 89}

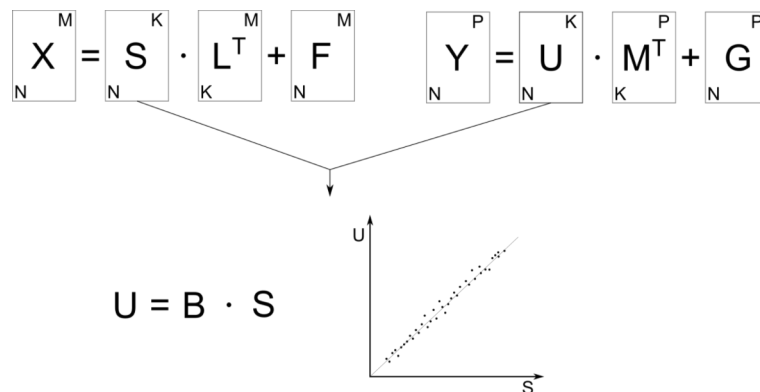


Abbildung 7: Prinzip der PLSR. Schematisch dargestellt sind die Matrices der unabhängigen X -Daten und der abhängigen Y -Daten. Beide Matrices sind aus einer Score-Matrix S bzw. U , Loadingsmatrix L^T bzw. M^T und einer Fehlermatrix F bzw. G aufgebaut. Durch eine lineare Regression der Score-Matrices S und U resultiert eine neue Matrix B , die den Zusammenhang beider Matrices wiedergibt, nach Lohninger.⁸⁸

Interval partial least squares-Regression

Um die Vorhersagekraft der PLS-Regression zu erhöhen, kann statt dem gesamten Datensatz nur der Teil in das Modell einfließen, der die wichtigsten Informationen enthält. Dadurch wird die Komplexität des Datensatzes reduziert, während die Analyse beschleunigt wird. Eine solche Methode ist die interval partial least squares-Regression (*i*PLSR). Sie wurde von Nørgaard *et al.*⁹⁰ entwickelt. Der zu analysierende Datensatz wird dabei in gleichgroße Datenbereiche, die Intervalle, geteilt. Durch eine Variablenselektion werden die Datenbereiche mit den wichtigsten Informationen herausgefiltert und spektrale Interferenzen aus anderen Bereichen entfernt. An den ausgewählten Teilbereichen wird die PLSR durchgeführt. Die Breite der Intervalle innerhalb eines Datensatzes ist gleich. Die Intervallbreite verschiedener Datensätze kann variieren. Im Fall von sehr breiten Bereichen liegt eine normale PLSR vor. Für sehr schmale Bereiche kann die PLSR auch als eine automatisierte univariate Regression betrachtet werden. Die *i*PLSR ist eine attraktive Methode, insbesondere für die Analyse spektraler Daten, da sie einen Überblick über analytisch interessante Spektralbereiche bietet und neben der Vorhersagekraft auch die Interpretierbarkeit der PLSR verbessert.^{90, 91}

Die hier vorgestellten chemometrischen Methoden wurden mit Hilfe der Computerprogramme MATLAB^{®92} und OriginLab (OriginPro^{®93}) auf die experimentell erhobenen LIBS-Daten angewendet.

3. Publikationen

3.1 Darstellung des eigenen Anteils an den einzelnen Publikationen

In diesem Kapitel werden die vier dieser kumulativen Dissertation zugrunde liegenden Publikationen kurz vorgestellt und der Eigenanteil an den Manuskripten 1 bis 4 dargestellt. Die publizierten Manuskripte sind in den Kapiteln 3.2, 3.3, 3.4 und 3.5 zu finden. Die zusätzlichen Informationen zu den Publikationen finden sich im Anhang in den Abschnitten S1.1 und S1.2.

Manuskript 1

D. Riebe, A. Erler, P. Brinkmann, T. Beitz, H.-G. Löhmannsröben, R. Gebbers, „Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture“, *Sensors* **19** (23), 5244 (2019).

Inhalt dieser Publikation ist der Vergleich unterschiedlicher Kalibrierungsansätze und die Evaluierung verschiedener Quantifizierungsmethoden. Dies erfolgte anhand von Spektren, die mittels laserinduzierter Breakdownspektroskopie an Agrarböden aufgenommen wurden. Ich habe experimentelle Untersuchungen durchgeführt und die Hauptkomponentenanalyse sowie die entsprechende Grafik erstellt. Die Auswertung der Messungen, die multivariate Datenanalyse und die Anfertigung des Manuskripts erfolgten durch Daniel Riebe und die Koautoren. Kapitel 3.2 beinhaltet das publizierte Manuskript.

Manuskript 2

N. Rethfeldt, P. Brinkmann, D. Riebe, T. Beitz, N. Köllner, U. Altenberger, H.-G. Löhmannsröben, "Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS", *Minerals* **11** (12), 1379 (2021).

Ziel dieser Publikation war die Detektion von Elementen der Seltenen Erden in Mineralen mit Hilfe der laserinduzierten Breakdownspektroskopie. Die Publikation basiert auf der Bachelorarbeit von Nina Rethfeldt. Die gemahlten Feldproben wurden von apl. Prof. Dr. Uwe Altenberger zur Verfügung gestellt. Die pulverförmigen synthetischen Proben wurden durch Nina Rethfeldt hergestellt. In beiden Fällen wurden die Pulverproben homogenisiert und anschließend zu Presslingen geformt. Die Presslinge wurden an einem stationären LIB-Spektrometer vermessen. Die Auswertung mittels univariater und iPLS-Regression erfolgte mit Hilfe der Software Matlab. Anhand der durchgeführten Experimente und der erhaltenen Ergebnisse hat die Bachelorstudentin ihre Bachelorarbeit geschrieben. Basierend auf dieser Arbeit habe ich in Zusammenarbeit mit den Koautoren das Manuskript verfasst. Im Rahmen dessen habe ich die Hauptkomponentenanalyse durchgeführt und ausgewertet, sowie Abbildungen mit Hilfe des Programms OriginLab erstellt. Die Publikation ist in Kapitel 3.3 gezeigt. Kapitel S1.1 beinhaltet das ergänzende Material der Publikation.

Manuskript 3

P. Brinkmann, N. Köllner, S. Merk, T. Beitz, U. Altenberger, H.-G. Löhmannsröben, „Comparison of handheld and Echelle spectrometer to assess copper in ores by means of Laser-Induced Breakdown Spectroscopy (LIBS)“, eingereicht bei *Minerals* (2022).

In dieser Publikation wurden kupferhaltige Gesteinsproben mittels laserinduzierter Breakdownspektroskopie und multivariater Datenanalyse untersucht, um den Kupfergehalt zu ermitteln. Die genutzten kupferhaltigen Erzproben wurden zum einen durch Dr. Friederike Körting und zum anderen durch den Projektpartner LTB Lasertechnik Berlin GmbH zur Verfügung gestellt. Die Proben der Begleitgesteine wurden von apl. Prof. Dr. Uwe Altenberger bereitgestellt. Die gemahlten Pulverproben habe ich homogenisiert und zu Presslingen geformt. Die hergestellten Presslinge wurden durch mich einerseits mit dem Handheld-LIB-Spektrometer und andererseits mit dem stationären LIB-Spektrometer vermessen. Die anschließende Auswertung der Spektren und die Erzeugung von auswertbaren Daten aus den Rohdaten realisierte ich mit Hilfe der Software Matlab und OriginLab. Die multivariate Datenanalyse mittels univariater und multivariater Regression sowie der Hauptkomponentenanalyse habe ich mit dem Programm Matlab durchgeführt. Die Abbildungen habe ich mit der Software OriginLab erstellt. Die erhaltenen Ergebnisse wurden durch mich diskutiert. Das Manuskript ist in Abstimmung mit den Koautoren von mir verfasst. Das eingereichte Manuskript ist in Kapitel 3.4 zu finden.

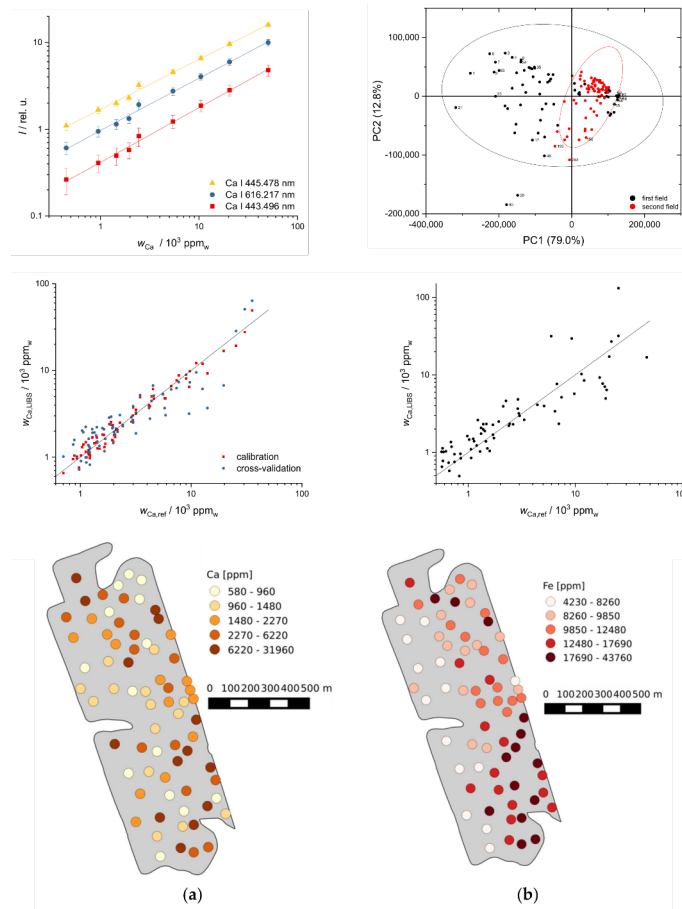
Manuskript 4

M. Wójcik, P. Brinkmann, R. Zdunek, D. Riebe, T. Beitz, S. Merk, K. Cieřlik, D. Mory, A. Antończak, "Classification of Copper Minerals by Handheld Laser-Induced Breakdown Spectroscopy and Nonnegative Tensor Factorisation", *Sensors* **20** (18), 5152 (2020).

Diese Publikation beinhaltet die Klassifikation von Kupfermineralen durch laserinduzierte Breakdownspektroskopie und multivariate Datenanalyse. Die der Klassifizierung zugrunde liegenden Kupferminerale habe ich zusammengetragen und deren Charakteristika, wie Name und Formel des Minerals, Herkunft und Mineralklasse katalogisiert. Die Messung der Kupferminerale an mehreren Messpunkten pro Mineral habe ich mit Hilfe des Handheld-LIB-Spektrometers durchgeführt. Die Minerale habe ich fotografisch erfasst und die unterschiedlichen Messpunkte auf den Mineralen anhand der fotografischen Aufnahmen dokumentiert. Die Auswertung der Messergebnisse, die Durchführung der multivariaten Datenanalyse und die Erstellung des Manuskripts erfolgte durch Michał Wójcik und die Koautoren. In Kapitel 3.5 ist das publizierte Manuskript, in Kapitel S1.2 das ergänzende Material gezeigt.

3.2 Vergleich von Kalibrierungsansätzen in der laserinduzierten Breakdownspektroskopie für die proximale Bodenerfassung in der Präzisionslandwirtschaft (*Manuskript 1*)

„Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture“



Sensors **19** (23), 5244 (2019)

<https://doi.org/10.3390/s19235244>



Article

Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture

Daniel Riebe ¹, Alexander Erler ¹, Pia Brinkmann ¹, Toralf Beitz ¹, Hans-Gerd Löhmannsröben ^{1,*} and Robin Gebbers ²

¹ Physical Chemistry, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; riebe@uni-potsdam.de (D.R.); aerler@uni-potsdam.de (A.E.); pbrinkma@uni-potsdam.de (P.B.); beitz@uni-potsdam.de (T.B.)

² Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee 100, 14469 Potsdam, Germany; rgebbers@atb-potsdam.de

* Correspondence: loeh@chem.uni-potsdam.de

Received: 28 October 2019; Accepted: 25 November 2019; Published: 28 November 2019



Abstract: The lack of soil data, which are relevant, reliable, affordable, immediately available, and sufficiently detailed, is still a significant challenge in precision agriculture. A promising technology for the spatial assessment of the distribution of chemical elements within fields, without sample preparation is laser-induced breakdown spectroscopy (LIBS). Its advantages are contrasted by a strong matrix dependence of the LIBS signal which necessitates careful data evaluation. In this work, different calibration approaches for soil LIBS data are presented. The data were obtained from 139 soil samples collected on two neighboring agricultural fields in a quaternary landscape of northeast Germany with very variable soils. Reference analysis was carried out by inductively coupled plasma optical emission spectroscopy after wet digestion. The major nutrients Ca and Mg and the minor nutrient Fe were investigated. Three calibration strategies were compared. The first method was based on univariate calibration by standard addition using just one soil sample and applying the derived calibration model to the LIBS data of both fields. The second univariate model derived the calibration from the reference analytics of all samples from one field. The prediction is validated by LIBS data of the second field. The third method is a multivariate calibration approach based on partial least squares regression (PLSR). The LIBS spectra of the first field are used for training. Validation was carried out by 20-fold cross-validation using the LIBS data of the first field and independently on the second field data. The second univariate method yielded better calibration and prediction results compared to the first method, since matrix effects were better accounted for. PLSR did not strongly improve the prediction in comparison to the second univariate method.

Keywords: laser-induced breakdown spectroscopy; LIBS; proximal soil sensing; soil nutrients; elemental composition

1. Introduction

Precision agriculture (PA) requires reliable, affordable, immediately available soil data with sufficient spatial and temporal resolution [1]. Soil maps for PA are typically derived from soil sampling with subsequent laboratory analysis or from mapping with automated mobile proximal soil sensors. Soil sampling and laboratory analysis is time consuming and becomes prohibitively expensive if conducted on a fine grid [2]. As an alternative, mobile proximal soil sensors can measure several hundred points (different locations) per hectare [3–7]. However, most of the current soil sensors

neither cover the whole range of soil fertility parameters nor do they directly measure them. The automated mobile proximal soil sensing systems frequently used in practical PA as well as research include geoelectrical, potentiometric pH, gamma-ray and spectral-optical sensors. Among these, pH electrodes most directly access a soil fertility parameter, namely active acidity [8]. Gamma-ray sensors detect radiation from K decay and from other isotopes. This can be correlated with K in clay minerals and with plant available K^+ [9,10]. Spectral-optical sensors include spectrometers and multi-wavelength sensors in the visible and near-infrared region. Correlations with several soil fertility parameters were observed, particularly with organic matter [11]. In research, visible and near-infrared spectrometers were used to map several fertility parameters at the same time [12,13]. For practical applications, cheaper and more robust dual or multi-wavelength sensors for organic matter and soil moisture were commercialized (e.g., by Veris Technologies and Precision Planting). However, the relationships between optical soil properties and soil fertility parameters are variable due to large overlaps of absorption bands. Thus, optical online sensors require careful calibration for each field [11,14]. Geoelectric sensors traditionally formed the backbone of PA soil analysis efforts and they are still widely used due to their robust nature making them both dependable and suitable for field applications [15–17]. However, apparent soil electrical conductivity (ECa) is affected by many soil parameters, including water content, texture, salinity, bulk density and temperature [18]. Therefore, reference sampling in each field is required [19]. In a comparative study, Piiki et al. [20] recently addressed the issue of obtaining as much direct information with as little calibration as possible. In that study, X-ray fluorescence (XRF) spectroscopy produced the most reliable predictions of soil parameters due to its direct detection of elemental compositions. XRF has gained interest in recent years due to the availability of handheld sensors [21–23]. The drawbacks of XRF include the harmful X-rays, long measurement times and the restriction to heavier elements.

Laser-induced breakdown spectroscopy (LIBS) is a promising alternative to XRF for determining element mass fractions in soils. In this method, an intense pulse of laser radiation is focused onto the soil, where it ablates material from the surface and creates a microplasma. Subsequently, excited atoms and ions in the plasma emit specific radiation which can be analyzed to elucidate the elemental composition of the sample [24–28]. Ablation and plasma excitation are both highly complex phenomena. Since the interaction of the laser radiation with the sample surface is influenced by its composition and structure, a matrix-dependence of the signal response is observed. Matrix effects result from the light-to-sample coupling, collisional interactions within the plasma, and plasma temperature, among others, all of which influence the ratio of neutral and ionized species and self-absorption. These matrix effects and spectral interferences were already investigated in different types of soil [29,30]. In LIBS soil analyses, the whole spectrum of elements can be accessible. Depending on the calibration effort, elements can be determined qualitatively or quantitatively. This allows the direct analysis of macro and micro nutrients since no or only minimal sample preparation is necessary. In order to achieve the power density required for plasma generation, the laser light pulse, of typically nanosecond duration, is usually focused to a spot of about 10–500 μm diameter. As a consequence, the soil micro-heterogeneity has to be considered in order to obtain representative results. This can be achieved by averaging multiple spectra.

Additional advantages of LIBS include the measurement speed, safety, as well as the portability of the technique. These attributes make the method particularly interesting for on-site soil mapping. While large-scale soil mapping applications of LIBS have not been reported yet and early LIBS investigations of soils were focused on pollutants [31], the detection of nutrients in soils was already demonstrated in some publications. Diaz et al. [32] were able to determine detection limits of P, Fe, Mg, Ca and Na by univariate calibration in fertilizer/soil mixtures. Yongcheng et al. [33] could improve the prediction of the Mg mass fraction by using a multivariate regression model that incorporates the lines of other metals present in the soil. Nicolodelli et al. [34] investigated the feasibility of measuring C in soils with a low resolution spectrometer. Rühlmann et al. [35] compare univariate and multivariate data

evaluation approaches for the quantification of the nutrient Ca in reference soils. A review of recent work on the application of LIBS for investigating agricultural materials can be found in [36].

In this work, LIBS spectra of soils were measured in the laboratory as a fundamental study and to provide a basis for the future application of LIBS directly on agricultural fields. The aim of the study was the evaluation of different quantification approaches for LIBS data which consider the matrix effect. Another goal was to explore how a calibration obtained for one field can be transferred to another one. The accuracy of standard-free LIBS approaches [31,37] for the analysis of the very complex matrix soil is still unsatisfactory. Therefore, the focus of this work was to compare different univariate or multivariate calibration methods for the determination of nutrient mass fractions. The target parameters of the investigation were the major nutrients Ca and Mg as well as the minor nutrient Fe. One aim of this work was to examine whether the quantification of nutrients on the size scale of a field can be carried out by univariate calibration or whether a multivariate method, namely partial least squares regression (PLSR) has to be applied. Different calibration strategies, such as the generation of calibration standards by standard addition to a single reference soil sample and the use of multiple reference soil samples, were compared. The heterogeneity of the soil samples was characterized by principal component analysis (PCA). The plasma was generated using UV radiation (355 nm), which was recommended for soil investigations [30] in contrast to the widely applied NIR radiation (1064 nm).

2. Materials and Methods

Soil samples and reference analysis. A total of 139 samples from two agricultural fields near the village of Wilmersdorf in Northeast Germany (53°06' N, 13°54' E) were investigated. The regional soilscape was formed by the last glaciation about 10,000 years ago and the following postglacial processes. The parent material of the soil consists of calcareous glacial till covered by sandy deposits. Soil texture varies between sand, loamy sand, and sandy loam in the topsoil. The main soil types are alfisols. The samples were collected in 2011 to obtain reference data for proximal soil sensing as published by Schirrmann, Gebbers and Kramer [14]. Reference analysis for elements was carried out by inductively coupled plasma optical emission spectroscopy (ICP-OES) after aqua regia extraction by a certified laboratory. As target parameters, Ca and Mg were selected as examples for major nutrients and Fe as an example for minor nutrients.

ECa mapping: ECa [mS/m] was mapped with a Veris 3100 system (Veris Technologies, Salina, KS, USA) in 2011. Only data from the shallow measurement were used. Data were interpolated by block kriging (a) on a regular grid for visualization and (b) on the sampling locations for correlation analysis.

LIBS setup and measurement parameters. The plasma was created using a Nd:YAG laser (Quanta-Ray, Spectra-Physics, Santa Clara, CA, USA, $\lambda = 355$ nm, $E = 90$ mJ). Emissions were collected by a concave mirror, coupled into an optical fiber and guided to an echelle spectrometer (Aryelle Butterfly, LTB, Berlin, Germany) equipped with an ICCD camera (iStar, AndorTechnology, Belfast, UK). The spectrometer has two separate wavelength ranges (UV range: 190–330 nm, VIS range: 275–750 nm) and a resolution of 20–30 pm. A total of 200 single shot spectra were recorded per sample in the UV as well as in the VIS range. The sample holder was rotated and linearly translated during measurements forming a spiral-like trace of ablation events. Optimization of LIBS spectra led to the following measurement parameters: a detection delay of 2 μ s, a measurement window of 10 μ s as well as a constant amplification factor of the iCCD camera.

Sample treatment for LIBS. The soil samples were mixed with starch (19 wt% final mass fraction), ground in an agate ball mill and subsequently pressed into pellets. One pellet was created for each field sample point. For standard addition, the sample point of the first field with the lowest intensity of the respective element peak was chosen for each investigated element. To that end, the soils were mixed with the respective amounts of target elements added as salts (CaCO_3 , MgCl_2 and FeS, the amount of starch was reduced in order to keep soil fraction constant) and also formed into pellets. The salts were obtained from Sigma Aldrich (St. Louis, MO, USA).

Data pretreatment. Pretreatment of the LIBS spectra consisted of outlier (e.g., spectra where the laser did not fire) removal using the following procedure. The total intensity of each spectrum for a given sample was calculated. The median of these values was derived and only the spectra in the range $0.75 \times \text{median} \leq \text{total intensity} \leq 1.25 \times \text{median}$ were averaged to yield a single spectrum for each soil sample. For PLSR calibration (third method, see below) spectra were mean centered. Logarithms of the known element mass fractions were used as y variables when the data distribution was strongly skewed. For univariate calibration, the peaks of investigated elements were integrated for each averaged spectrum individually. For multivariate analysis, the entire spectra of either the UV or the VIS region were used.

Calibration. *Standard addition univariate calibration.* The first calibration method was based on standard addition using just one soil sample and applying the derived univariate calibration model to the LIBS data of both fields. Using the standard addition method potentially has several advantages. First, the technique can be used to determine the mass fraction of one or several nutrients in one soil sample by extrapolating the mass fraction of the base sample from the samples with added known quantities of an element. Second, as applied in this work, reference analysis of the pure soil sample by the traditional digestion method and in combination with samples where increasing amounts of the nutrients were added can provide LIBS calibration curves. One additional benefit is that standard addition yields response curves where the matrix effect of the local soil type is accounted for. The nutrients are advantageously added as salts. However, care must be taken that the salts are not hygroscopic. The third advantage is that the large range of nutrient mass fractions covered for the determination of calibration curves facilitates assigning lines to elements and finding the best lines to use for the quantification of each element. Potentially, different lines could be chosen for different mass fraction ranges. The most intense lines might be useful at low element mass fractions, while at high element mass fractions, these intense lines can be self-absorbed. Therefore, choosing weaker lines can become beneficial.

Reference univariate calibration. The second method derived a univariate calibration model from the reference analytics of all samples from one field. The prediction is validated using the second field. This approach accounts for the matrix effects even better, as the variation of the matrix across a field is also reflected. Furthermore, the error in the laboratory-based reference analytics of one soil sample has a large influence on the LIBS-based results. Therefore, a larger data pool with known reference values and a broader variety of matrices should be regarded. In our case, the data of the first field is used for the calibration and the data of the second field is used for the validation. The concept of this procedure is that reference data gained in one year could be used to build a calibration for a specific field. In subsequent years, it will not be necessary to take new samples and the calibration can therefore also be applied to fields in the near surrounding and to fields with a similar soil type. Furthermore, a much larger number of data points can be evaluated, as LIBS can be employed directly on the field, allowing a closer spatial mapping of the fields.

Reference multivariate calibration. The third approach was a multivariate calibration based on PLSR. PLSR was done using the kernel algorithm and a maximum of 7 components. The LIBS spectra of the first field were used for training. Validation was carried out by 20-fold cross-validation using the data of the first field and independently by testing the model trained on the data of the first field with the data of the second field.

Software: Origin (OriginLab, Northampton, MA, USA) was used for PCA. Unscrambler X (Camo Analytics, Oslo, Norway) was used to perform PLSR.

3. Results and Discussion

3.1. LIBS Spectra

LIBS spectra of soils are rich in lines, mostly due to the presence of Fe and other transition metals (Figure 1). Additionally, lines of most minor and major mineral nutrients are found, the lines of metal nutrients are particularly intense.

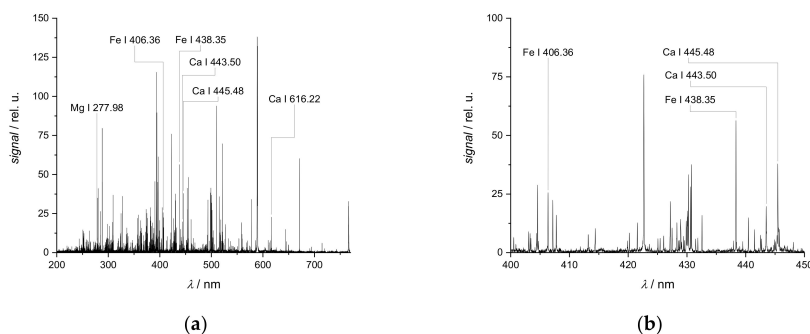


Figure 1. Typical soil laser-induced breakdown spectroscopy (LIBS) spectrum (composite of UV and VIS part of the LIBS spectrum), (a) full wavelength range, (b) detail view of wavelength range 400–450 nm.

But the plethora of lines complicates their assignment to elements, as peak overlapping is frequently observed. The line identification is not straightforward as the total elemental composition of soils is generally not known. Otherwise, the limitation of the measurement campaign on one or two fields is advantageous for peak assignment as the soil heterogeneity is not as great as on a larger scale. Most peaks are generally found in spectra of all samples (different points on the field) but their intensities still differ substantially on the scale of one field. This observation is due to variations of the element mass fractions across the field, which is also the origin of the strong matrix effects encountered and necessitates the determination of element distributions in precision agriculture.

3.2. Standard Addition Calibration

Ca mass fractions vary greatly (from 500 ppm_w to 5 wt%) in the investigated soils. A corresponding calibration curve has to reflect a mass fraction range of nearly three orders of magnitude and is best presented in a double-logarithmic diagram. The fitting function $I = a w^c$, where I is line intensity, w the mass fraction and a and c are the optimized constants, essentially equivalent to taking the logarithm of the data prior to linear fitting, was used. The calibration curves for three different Ca lines (at 443.496 nm, 445.478 nm, 616.217 nm) are shown in Figure 2. These lines provided best results in terms of repeatability (precision), linearity ($R^2 > 0.993$) and usability over the entire range of relevant mass fractions.

The Fe mass fraction in the investigated soil (from 0.5 wt% to 6 wt%) is on a higher mass fraction level but covers a narrower range than the Ca mass fraction. Thus, a linear calibration plot was selected (Figure 3a). The LIBS spectrum contains a very high number of Fe lines. Due to the high mass fraction, many of them show strong effects of self-absorption. This requires a careful selection of lines. The lines at 406.359 nm and 438.354 nm (see Figure 3a) are best suited for calibration and yield $R^2 > 0.988$ for $\lambda = 406.359$ nm and $R^2 > 0.992$ for $\lambda = 438.354$ nm. The repeatability is not as good as for Ca as the error bars show, but quantification is still possible.

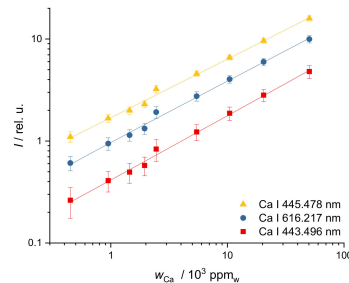


Figure 2. Calibration curves of different Ca lines in soil, integrated line intensity over calculated Ca mass fraction: R^2 (443.496 nm) = 0.994, R^2 (445.478 nm) = 0.996, R^2 (616.217 nm) = 0.993 (individual wavelengths are offset for visibility).

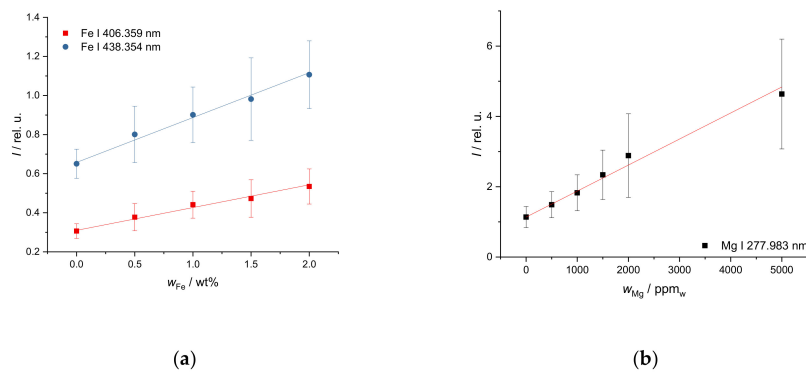


Figure 3. Response curve of (a) FeS and (b) $MgCl_2$ in soil, integrated line intensity over calculated (a) Fe and (b) Mg mass fractions, (a) R^2 (406.359 nm) = 0.988, R^2 (438.354 nm) = 0.992, (b) R^2 (277.983 nm) = 0.989.

Mg appears in the soils at low to intermediate mass fractions (from 700 ppm_w to 4000 ppm_w) and covers a narrow mass fraction range. Therefore, a linear calibration curve was selected as well. Useful Mg lines are only found in the UV region. The repeatability is worse at higher mass fractions as evidenced by higher error bars. The line at 277.983 nm shown in Figure 3b yields the best result in terms of linearity over a range of mass fractions ($R^2 = 0.989$).

For all chosen elements, lines could be found in the available spectral range that are free from overlap, less prone to self-absorption effects and provide good dynamic ranges. The detection limits usually are in the range of 100 ppm which is adequate for the soils in question. Lower detection limits would only be needed for micronutrients.

3.3. Application of Standard Addition Calibration to the First Field

In order to obtain the mass fraction distribution of the nutrients on a field, the standard addition calibration was applied to LIBS spectra of all samples (points on the field) for predicting the element mass fractions. A comparison of the values thus obtained from the LIBS measurements and the reference values obtained by ICP-OES yields insights into the performance of the calibration method. This is shown for the nutrient Ca in Figure 4a, where the Ca mass fractions measured by LIBS are the averages of the results for the three lines in the LIBS spectrum. The line in Figure 4a presents the

identity of values obtained by LIBS and ICP-OES. The Ca mass fractions obtained by LIBS scatter around this line and show a very good agreement over the range of mass fractions investigated. Thus, a univariate calibration, built using only one reference sample, can be used for predicting the Ca mass fraction for the entire field. Absolute deviations for the different sample points on the field are shown in Figure 4b with the highest errors in the range of 1%. The deviations seem accumulated around certain points of the field as sample numbers indicate proximity. Because samples were taken along predefined lines across the field, this could be an indication of different soil types and thus matrix effects at those positions.

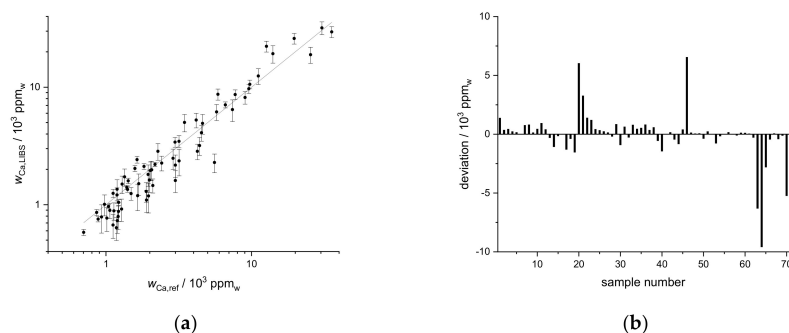


Figure 4. (a) Mass fraction of Ca predicted by LIBS univariate calibration based on standard addition ($w_{Ca,LIBS}$) compared to reference values obtained by ICP-OES upon aqua regia extraction ($w_{Ca,ref}$) for the first field, $R^2 = 0.91$, (b) deviation of predicted from reference value sorted by sample number.

Results were not as promising for Fe (Figure 5a). Although a rough agreement is present in values around 1 wt%, lower and especially higher Fe mass fractions are overestimated by as much as twice the reference value. Reasons could be the relatively narrow range of mass fractions and especially the high overall Fe mass fractions, leading to e.g., self-absorption. The results of the Mg prediction are shown in Figure 5b. Agreement of the results is much better, although in the low mass fraction region, a larger spread of the data points is observed. Overall, this kind of calibration is satisfactory for the major nutrients Ca and Mg but the predictive power is limited for the minor nutrient Fe.

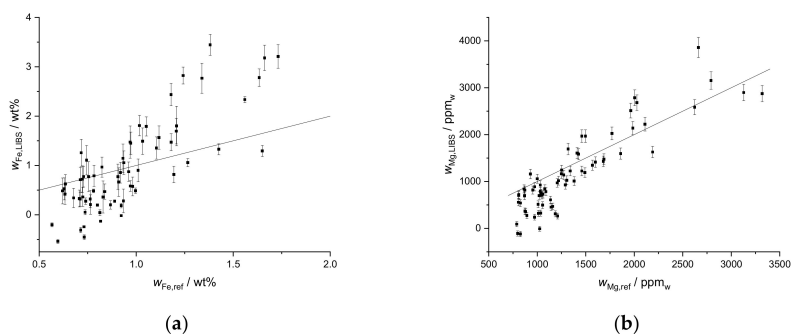


Figure 5. Mass fractions of (a) Fe and (b) Mg predicted by LIBS univariate calibration based on standard addition compared to reference values obtained by ICP-OES upon aqua regia extraction for the first field.

3.4. Application of Standard Addition Calibration to Second Field

As a further validation, the same calibration curves were also applied to the second field data, which does not contain the soil sample used for obtaining the standard addition calibration. Therefore, the results of this second field are a measure of the predictive power for local agriculturally-used areas in the surrounding of the calibration field. It is likely that the soil types on these fields will have a similar characteristic.

The prediction results of the three elements for both fields are similar (Figures 6 and 7). The best agreements are again achieved for Ca and Mg, especially the correlation for Ca is excellent. The prediction for Ca seems to be robust and can be also used for different fields of similar soil types. The deviation of the Mg prediction (Figure 7) is a bit stronger than for the first field. While it could still be usable in practice to get a rough estimation of the Mg mass fraction, a better method is desirable. The agreement of the values of the second field in the Fe plot (Figure 6b) is actually better than for the first field from which the soil for the calibration was taken, although the range of mass fractions is larger. However, the deviation is still present, and the calibration can only be used as a rough estimation.

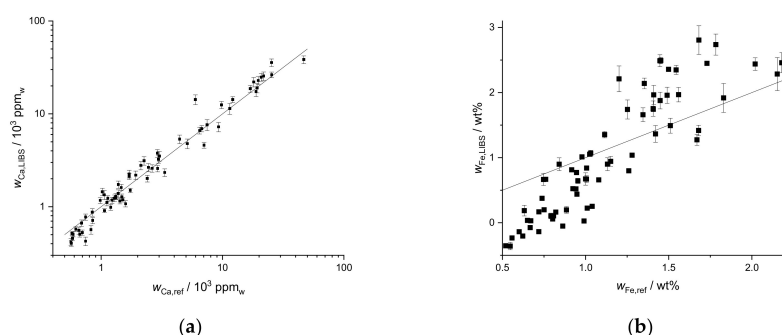


Figure 6. Mass fraction of (a) Ca and (b) Fe predicted by LIBS univariate calibration based on standard addition compared to reference values obtained by ICP-OES upon aqua regia extraction for the second field, R^2 (Ca) = 0.60.

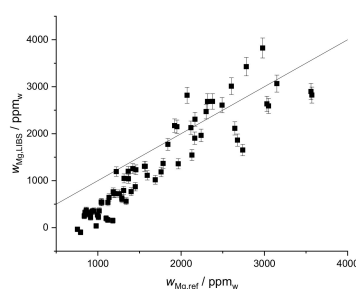


Figure 7. Mass fraction of Mg predicted by LIBS univariate calibration based on standard addition compared to reference values obtained by ICP-OES upon aqua regia extraction for the second field.

Although the calibration method established from one sample works fine for the element Ca, even on different fields, the applicability to other nutrients is limited. The most important reason for this finding is the matrix effect, which is not sufficiently taken into account by this method. On a future field campaign, the discussed peaks in the LIBS spectra can, however, be used to find points of extreme element mass fractions, which are then crucial sampling points from which a better calibration can be built.

3.5. Reference Univariate Calibration

The results of the reference univariate calibration using all samples from the first field and its application to the second field are shown below. The calibration data of the LIBS response for three different wavelengths as a function of the Ca mass fraction obtained by the reference analysis are shown in Figure 8a. For all wavelengths, a good regression (e.g., R^2 (445.478 nm) = 0.94) between data of the LIBS and reference analytical method is obtained, the coefficients of determination for the lines at 443.496 nm and 445.478 nm are better than those for the line at 616.217 nm. The calibration curves are applied to the LIBS data of the second field (Figure 8b), the Ca mass reported as measured by LIBS are the averages of the results for the three lines. This procedure results in predicted values that are in very good agreement with the reference data ($R^2 = 0.93$).

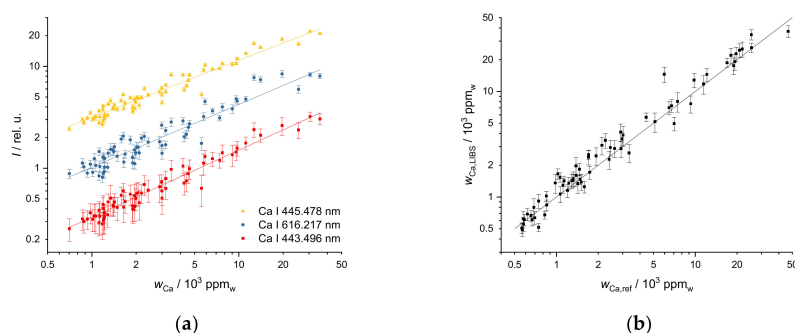


Figure 8. (a) Univariate calibration of Ca mass fraction calculated for data of the first field, R^2 (443.496 nm) = 0.93, R^2 (445.478 nm) = 0.94, R^2 (616.217 nm) = 0.86 (individual wavelengths are offset for visibility), (b) Calibration applied to the second field, $R^2 = 0.93$.

The same procedure was applied to Fe (Figure 9). The calibration plots (Figure 9a) took two signals at the wavelengths at 406.359 nm and 438.354 nm into account but yielded a poorer prediction than that found for Ca. However, the larger amount of data used for the calibration compared to the standard addition method results in a better prediction for the second field (Figure 9b). Here, the matrix effects are better accounted for. Although the result is acceptable, a univariate calibration might not be ideal for Fe.

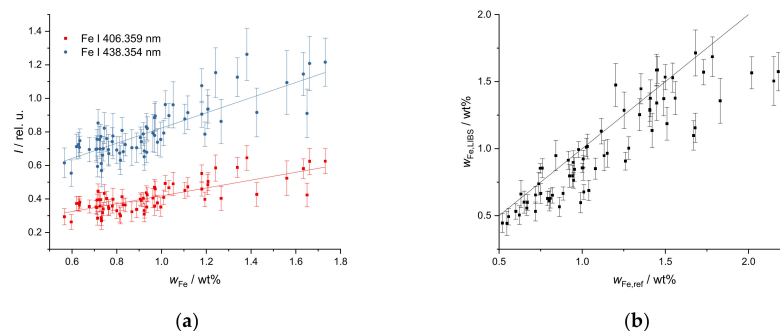


Figure 9. (a) Univariate calibration of Fe mass fraction calculated for data of the first field, R^2 (406.359 nm) = 0.53, R^2 (438.354 nm) = 0.61; (b) Calibration applied to the second field, $R^2 = 0.51$, without two most obvious outliers (same samples as for Ca) $R^2 = 0.68$.

While the calibration plot for Mg is characterized by a linear relationship but relatively large error bars (Figure 10a), the validation plot (Figure 10b) obtained on the second field actually has a better coefficient of determination than the calibration fit itself. Although this can be explained by a few high error intervals of the intensities used for the calibration, it is still a promising result.

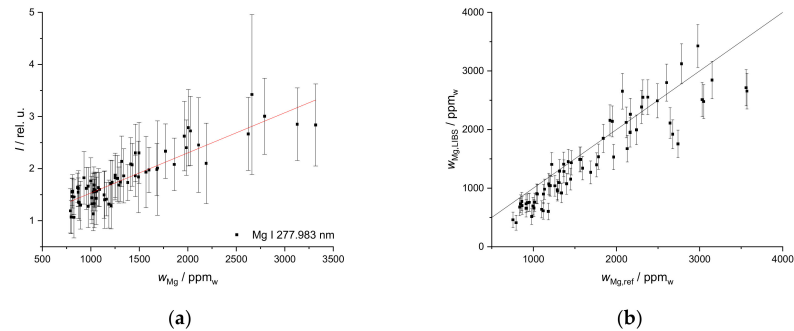


Figure 10. (a) Univariate calibration of Mg mass fraction calculated for data of the first field, the line used is Mg I 277.983 nm, $R^2 = 0.71$, (b) Calibration applied to the second field, $R^2 = 0.51$, without two most obvious outliers (same samples as for Ca) $R^2 = 0.76$.

Overall, it can be concluded that using actual field samples for building the calibration improves the performance for the univariate calibration. Such univariate methods can thus be used for predicting values. In the future, online LIBS measurements on the field can be used to determine which samples to use for laboratory analyses. If these samples represent a good range of the values of interest, a strong calibration can be built, which can in turn be used to predict values for all points of the online measurements.

3.6. Multivariate Analysis

One possibility for characterizing the heterogeneity of soil compositions of the individual fields, and of the first field in relation to the second field, is to apply PCA. For that purpose, the complete LIBS spectra were introduced into PCA. In the score plot (Figure 11) the first two components, which explain 93% of the variance, are presented. Grouping the data points by fields implies that the heterogeneity of the second field (red points) is lower than that of the first (black points). This also explains why the univariate calibrations were largely successful and justifies using the first field as training data and the second field as validation data in multivariate analyses.

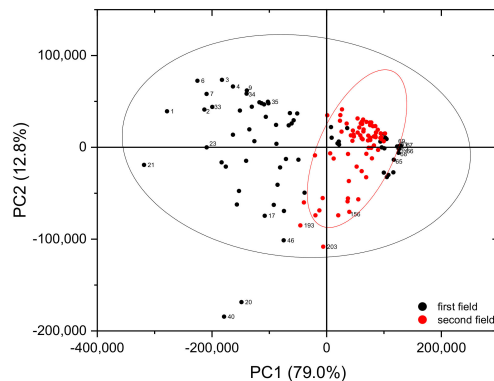


Figure 11. PCA of all LIBS spectra of the first (black points) and the second field (red points).

As a multivariate regression method, PLSR was performed on our data. Validation of the results of the regression model was taken into account in two different ways. First, 20-fold cross-validation was applied on the data of the first field. This is demonstrated for Ca data of the first field in Figure 12a. Correlation plots of Ca mass fractions predicted by LIBS and by ICP-OES are shown. The red squares are the result of the PLSR calibration, and the blue points represent 20-fold cross-validation. For PLSR, the logarithms of Ca values were taken, and 5 components were used.

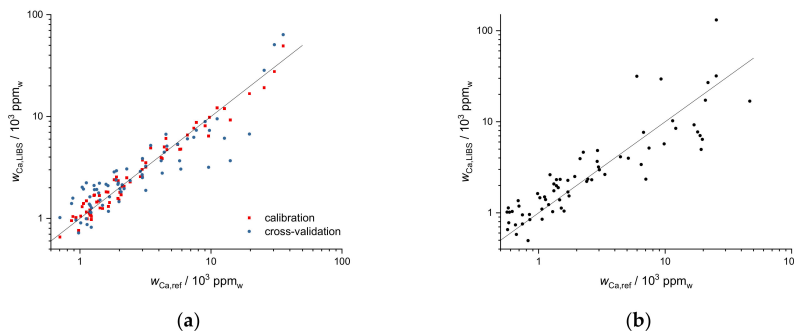


Figure 12. (a) Partial least squares regression (PLSR) of Ca data for the first field (UV range), red squares represent calibration data ($R^2 = 0.96$), blue points represent 20-fold cross-validation ($R^2 = 0.83$), (b) application of PLSR of Ca data to the second field ($R^2 = 0.58$).

The PLSR cross-validation data ($R^2 = 0.83$) demonstrate the potential of multivariate regression for predicting mass fractions on the same field. A second, additional validation is obtained by applying the calibration obtained of the first field to the second field, demonstrating how universal a calibration could become. This is shown in Figure 12b for Ca and gives an impression of the prediction quality of PLSR calibrations for additional fields with similar soil types ($R^2 = 0.58$).

PLSR was also applied to the elements Fe and Mg (Figure 13). For both elements, the logarithms of the mass fractions were used for PLSR. Seven components were used for PLSR of the Fe data, and 6 components were used for PLSR of the Mg data. The cross-validation plots of both elements scatter a bit stronger than the Ca plot ($R^2(\text{Fe}) = 0.70$, $R^2(\text{Mg}) = 0.79$). The application of the calibration data of the first to the second field results in a similar scattering of the predicted values. With increasing

element mass fractions, the deviation of predicted values from reference values becomes systematically larger. LIBS underestimates high Fe and Mg mass fractions.

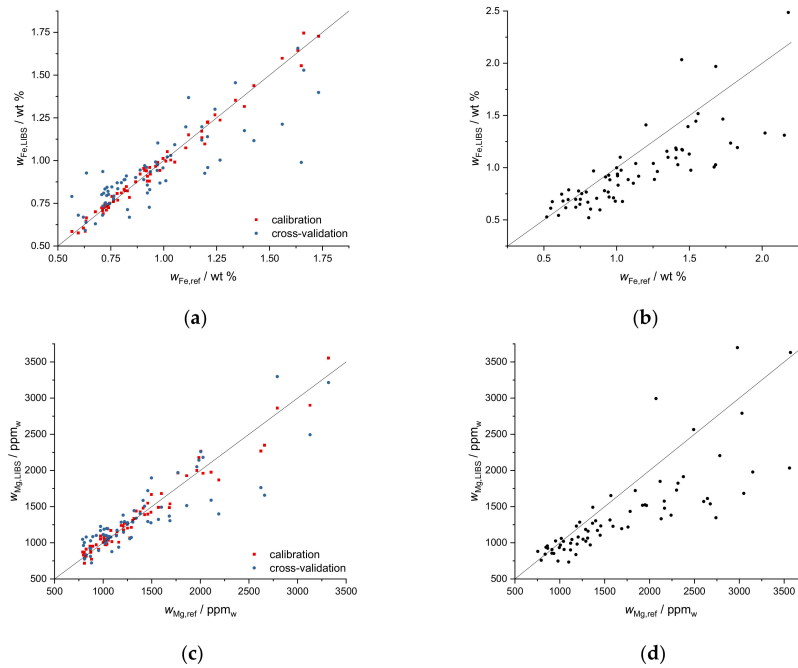


Figure 13. (a) PLSR of Fe data for the first field (UV range), red points represent calibration data ($R^2 = 0.986$), blue diamonds represent 20-fold cross-validation ($R^2 = 0.70$), (b) application of PLSR of Fe data to the second field $R^2 = 0.37$, (c) PLSR of Mg data for the first field (UV range), red points represent calibration data ($R^2 = 0.96$), blue diamonds represent 20-fold cross-validation $R^2 = 0.79$, (d) application of PLSR of Fe data to the second field $R^2 = 0.38$.

3.7. Field Maps

One main aim in precision farming is the mapping of nutrient distributions on agricultural fields. In connection with fertilizer recommendations based on measured data as well as models, these maps help the farmers identify nutrient-depleted areas. The change between nutrient-rich and depleted areas can be seen on the three maps of Figure 14. Thus, the eastern areas are richer in all three elements while the western areas are depleted.

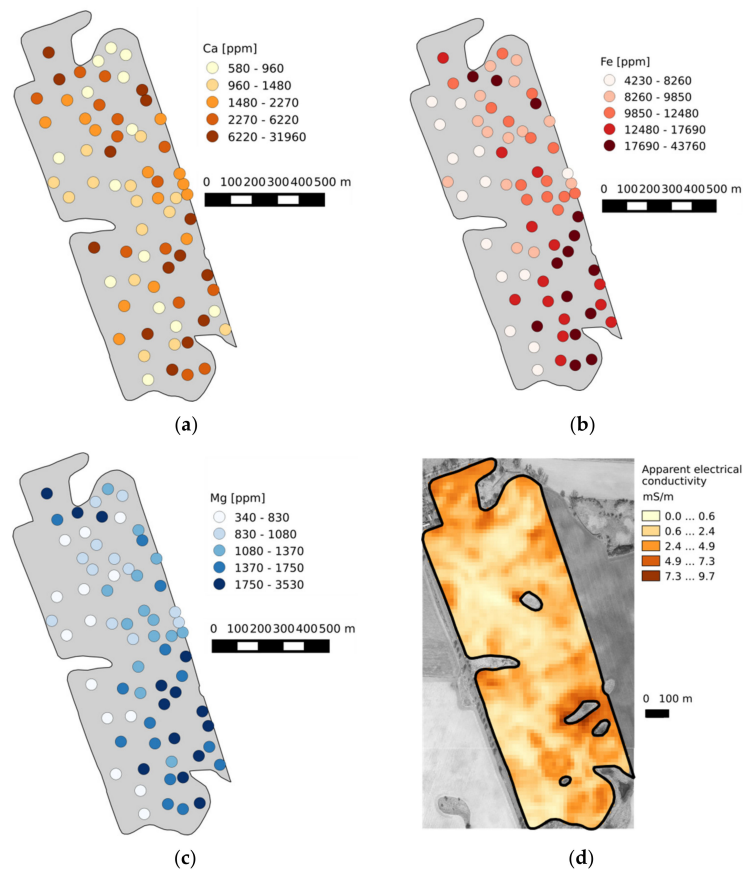


Figure 14. Maps showing the element distribution of (a) Ca, (b) Fe, (c) Mg and (d) ECa on the first field (north is up).

This spatial trend was further explained by the evaluation of the ECa map (Figure 14d), which was recorded directly on the field. The ECa values were positively correlated with all elements as determined by the reference method and by LIBS.

4. Summary and Conclusions

LIBS is a promising method for efficient soil data collection at high spatial resolution, as required for precision agriculture. This is due to the capability of LIBS to rapidly measure mass fractions of many elements simultaneously without or only very little sample preparation. These advantages are contrasted by a strong matrix dependence of the LIBS signal which requires careful calibration and data evaluation methods.

In this work, different approaches for the evaluation of LIBS data were investigated. The first univariate method was based on standard addition to just one soil sample, the second was a univariate method based on a larger number of samples characterized by reference analytics. This second method accounted for matrix effects and thus yielded better predictions. The third method was a multivariate method (PLSR) which yielded better calibration curves, but did not substantially improve

the prediction compared to the reference univariate method. Alternative and maybe better suited multivariate methods can potentially provide better results.

The work begun here should be continued by expanding the LIBS measurements to a larger database, containing more and especially different soil types. It could also be interesting to obtain soil samples from the same field in different years and verify if the same calibrations can be used. The classification of soil samples based on different soil types and the application of tailored calibration methods for these soil types can improve the accuracy and repeatability of measuring results in the future. Additionally, the amount of sample preparation performed in this work has to be reduced in order to transfer the method to the field.

Author Contributions: Conceptualization: D.R., T.B., H.-G.L. and R.G.; Data curation: D.R. and P.B.; Formal analysis: D.R.; Funding acquisition: D.R., T.B., H.-G.L. and R.G.; Investigation: D.R., A.E., P.B. and T.B.; Methodology: D.R. and T.B.; Project administration: D.R., T.B., H.-G.L. and R.G.; Resources: H.-G.L. and R.G.; Supervision: H.-G.L.; Visualization: D.R.; Writing—Original draft: D.R. and T.B. Writing—Review & editing: H.-G.L. and R.G.

Funding: The authors gratefully acknowledge the financial support for this research received from the German Ministry of Education and Research (BMBF) in the framework of the BonaRes project Integrated System for Site-Specific Soil Fertility Management (I4S, grant no. 031A564H).

Acknowledgments: The authors would also like to thank Sascha Hons for support during the LIBS measurements. We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stafford, J.V. Remote, non-contact and in-situ measurement of soil moisture content: A review. *J. Agric. Eng. Res.* **1988**, *41*, 151–172. [[CrossRef](#)]
2. Viscarra Rossel, R.A.; McBratney, A.B. Soil chemical analytical accuracy and costs: Implications from precision agriculture. *Aust. J. Exp. Agric.* **1998**, *38*, 765. [[CrossRef](#)]
3. Hummel, J.W.; Gaultney, L.D.; Sudduth, K.A. Soil property sensing for site-specific crop management. *Comput. Electron. Agric.* **1996**, *14*, 121–136. [[CrossRef](#)]
4. Viscarra Rossel, R.A.; McBratney, A.B.; Minasny, B. (Eds.) *Proximal Soil Sensing*; Springer: Dordrecht, The Netherlands, 2010.
5. Viscarra Rossel, R.A.; Adamchuk, V.I.; Sudduth, K.A.; McKenzie, N.J.; Lobsey, C. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In *Advances in Agronomy*; Academic Press: Cambridge, MA, USA, 2011; pp. 243–291.
6. Viscarra Rossel, R.A.; Bouma, J. Soil sensing: A new paradigm for agriculture. *Agric. Syst.* **2016**, *148*, 71–74. [[CrossRef](#)]
7. Gebbers, R. Proximal soil surveying and monitoring techniques. In *Precision Agriculture for Sustainability*; Stafford, J., Ed.; Burleigh Dodds Science Publishing: Cambridge, UK, 2018; pp. 1–50.
8. Adamchuk, V.I.; Morgan, M.T.; Ess, D.R. An automated sampling system for measuring soil pH. *Trans. ASAE* **1999**, *42*, 885–891. [[CrossRef](#)]
9. Hyvönen, E.; Turunen, P.; Vanhanen, E.; Arkimaa, H.; Sutinen, R. Airborne Gamma-ray Surveys in Finland. In *Aerogeophysics in Finland 1972–2004: Methods, System Characteristics and Applications*; Special Paper, 39; Airo, M.-L., Ed.; Geological Survey of Finland: Espoo, Finland, 2005; pp. 119–134.
10. Heggemann, T.; Welp, G.; Amelung, W.; Angst, G.; Franz, S.O.; Koszinski, S.; Schmidt, K.; Pätzold, S. Proximal gamma-ray spectrometry for site-independent in situ prediction of soil texture on ten heterogeneous fields in Germany using support vector machines. *Soil Tillage Res.* **2017**, *168*, 99–109. [[CrossRef](#)]
11. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*; Academic Press: Cambridge, MA, USA, 2010; pp. 163–215.
12. Shibusawa, S.; Sato, H.; Sasao, A.; Hirako, S.; Otomo, A. A Revised Soil Spectrophotometer. *IFAC Proc. Vol.* **2000**, *33*, 231–236. [[CrossRef](#)]
13. Schirrmann, M.; Gebbers, R.; Kramer, E.; Seidel, J. Soil pH mapping with an on-the-go sensor. *Sensors* **2011**, *11*, 573–598. [[CrossRef](#)]

14. Schirrmann, M.; Gebbers, R.; Kramer, E. Performance of Automated Near-Infrared Reflectance Spectrometry for Continuous in Situ Mapping of Soil Fertility at Field Scale. *Vadose Zo. J.* **2013**, *12*, 1. [[CrossRef](#)]
15. Corwin, D.L. Past, present, and future trends of soil electrical conductivity measurement using geophysical methods. In *Handbook of Agricultural Geophysics*; Allred, B.J., Daniels, J.J., Ehsani, M.R., Eds.; CRC Press: Boca Raton, FL, USA, 2008; pp. 17–44.
16. Gebbers, R.; Lück, E.; Dabas, M.; Domsch, H. Comparison of instruments for geoelectrical soil mapping at the field scale. *Near Surf. Geophys.* **2009**, *7*, 179–190. [[CrossRef](#)]
17. Heil, K.; Schmidhalter, U. The application of EM38: Determination of soil parameters, selection of soil sampling points and use in agriculture and archaeology. *Sensors* **2017**, *17*, 2540. [[CrossRef](#)] [[PubMed](#)]
18. Corwin, D.L.; Lesch, S.M. Application of Soil Electrical Conductivity to Precision Agriculture. *Agron. J.* **2003**, *95*, 455. [[CrossRef](#)]
19. Adamchuk, V.I.; Viscarra Rossel, R.A.; Marx, D.B.; Samal, A.K. Using targeted sampling to process multivariate soil sensing data. *Geoderma* **2011**, *163*, 63–73. [[CrossRef](#)]
20. Piikki, K.; Söderström, M.; Eriksson, J.; John, J.M.; Muthee, P.I.; Wetterlind, J.; Lund, E. Performance evaluation of proximal sensors for soil assessment in smallholder farms in Embu County, Kenya. *Sensors* **2016**, *16*, 1950. [[CrossRef](#)] [[PubMed](#)]
21. Zhu, Y.; Weindorf, D.C.; Zhang, W. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. *Geoderma* **2011**, *167–168*, 167–177. [[CrossRef](#)]
22. Gebbers, R.; Schirrmann, M. Potential of using portable x-ray fluorescence spectroscopy for rapid soil analysis. In *Precision Agriculture '15*; Stafford, J.V., Ed.; Wageningen Academic Publishers: Wageningen, The Netherlands, 2015; pp. 27–34.
23. Nawar, S.; Delbecque, N.; Declercq, Y.; De Smedt, P.; Finke, P.; Verdoodt, A.; Van Meirvenne, M.; Mouazen, A.M. Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? *Geoderma* **2019**, *350*, 29–39. [[CrossRef](#)]
24. Fortes, F.J.; Moros, J.; Lucena, P.; Cabalin, L.M.; Laserna, J.J. Laser-Induced Breakdown Spectroscopy. *Anal. Chem.* **2013**, *85*, 640–669. [[CrossRef](#)]
25. Cremers, D.A.; Chinni, R.C. Laser-Induced Breakdown Spectroscopy—Capabilities and Limitations. *Appl. Spectrosc. Rev.* **2009**, *44*, 457–506. [[CrossRef](#)]
26. Hahn, D.W.; Omenetto, N. Laser-Induced Breakdown Spectroscopy (LIBS), Part I: Review of Basic Diagnostics and Plasma—Particle Interactions: Still-Challenging Issues within the Analytical Plasma Community. *Appl. Spectrosc.* **2010**, *64*, 335A–336A. [[CrossRef](#)]
27. Hahn, D.W.; Omenetto, N. Laser-Induced Breakdown Spectroscopy (LIBS), Part II: Review of Instrumental and Methodological Approaches to Material Analysis and Applications to Different Fields. *Appl. Spectrosc.* **2012**, *66*, 347–419. [[CrossRef](#)]
28. Zorov, N.B.; Popov, A.M.; Zaytsev, S.M.; Labutin, T.A. Qualitative and quantitative analysis of environmental samples by laser-induced breakdown spectrometry. *Russ. Chem. Rev.* **2015**, *84*, 1021–1050. [[CrossRef](#)]
29. Popov, A.M.; Zaytsev, S.M.; Seliverstova, I.V.; Zakuskin, A.S.; Labutin, T.A. Matrix effects on laser-induced plasma parameters for soils and ores. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *148*, 205–210. [[CrossRef](#)]
30. Zaytsev, S.M.; Krylov, I.N.; Popov, A.M.; Zorov, N.B.; Labutin, T.A. Accuracy enhancement of a multivariate calibration for lead determination in soils by laser induced breakdown spectroscopy. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *140*, 65–72. [[CrossRef](#)]
31. Ciucci, A.; Corsi, M.; Palleschi, V.; Rastelli, S.; Salvetti, A.; Tognoni, E. New Procedure for Quantitative Elemental Analysis by Laser-Induced Plasma Spectroscopy. *Appl. Spectrosc.* **1999**, *53*, 960–964. [[CrossRef](#)]
32. Díaz, D.; Hahn, D.W.; Molina, A. Evaluation of Laser-Induced Breakdown Spectroscopy (LIBS) as a Measurement Technique for Evaluation of Total Elemental Concentration in Soils. *Appl. Spectrosc.* **2012**, *66*, 99–106. [[CrossRef](#)]
33. Yongcheng, J.; Wen, S.; Baohua, Z.; Dong, L. Quantitative Analysis of Magnesium in Soil by Laser-Induced Breakdown Spectroscopy Coupled with Nonlinear Multivariate Calibration. *J. Appl. Spectrosc.* **2017**, *84*, 731–737. [[CrossRef](#)]
34. Nicolodelli, G.; Marangoni, B.S.; Cabral, J.S.; Villas-Boas, P.R.; Senesi, G.S.; dos Santos, C.H.; Romano, R.A.; Segnini, A.; Lucas, Y.; Montes, C.R.; et al. Quantification of total carbon in soil using laser-induced breakdown spectroscopy: A method to correct interference lines. *Appl. Opt.* **2014**, *53*, 2170. [[CrossRef](#)]

Sensors 2019, 19, 5244

16 of 16

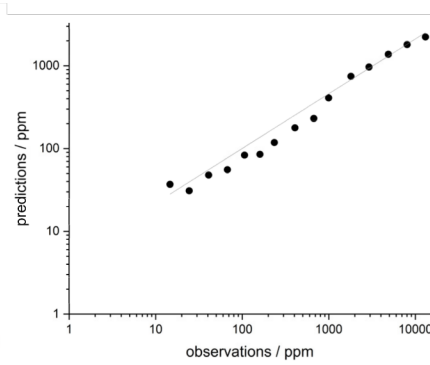
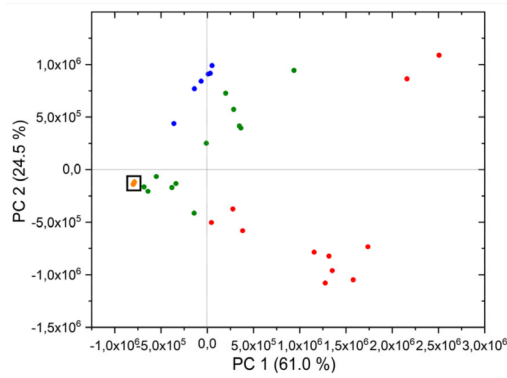
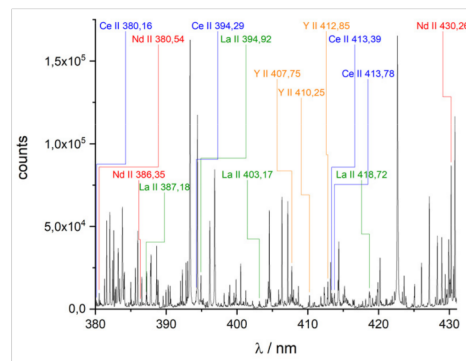
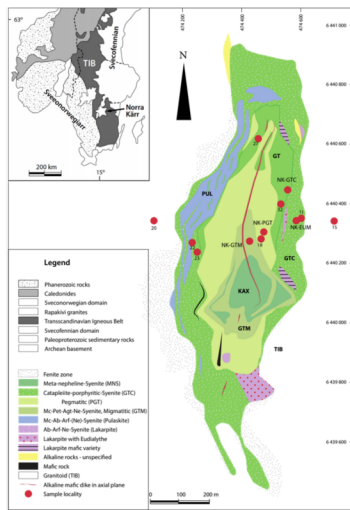
35. Rühlmann, M.; Büchele, D.; Ostermann, M.; Bald, I.; Schmid, T. Challenges in the quantification of nutrients in soils using laser-induced breakdown spectroscopy—A case study with calcium. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *146*, 115–121. [[CrossRef](#)]
36. Nicolodelli, G.; Cabral, J.; Menegatti, C.R.; Marangoni, B.; Senesi, G.S. Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers. *TrAC Trends Anal. Chem.* **2019**, *115*, 70–82. [[CrossRef](#)]
37. Herrera, K.K.; Tognoni, E.; Omenetto, N.; Smith, B.W.; Winefordner, J.D. Semi-quantitative analysis of metal alloys, brass and soil samples by calibration-free laser-induced breakdown spectroscopy: Recent results and considerations. *J. Anal. At. Spectrom.* **2009**, *24*, 413–425. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.3 Detektion von Seltenen Erdelementen in Mineralen und Böden mittels laserinduzierter Breakdownspektroskopie und Interval-PLSR (*Manuskript 2*)

„Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS“



Minerals **11** (12), 1379 (2021)

<https://doi.org/10.3390/min11121379>



Article

Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS

Nina Rethfeldt ¹, Pia Brinkmann ¹, Daniel Riebe ¹ , Toralf Beitz ^{1,*}, Nicole Köllner ² , Uwe Altenberger ² and Hans-Gerd Löhmansröben ¹

¹ Physical Chemistry, University of Potsdam, 14476 Potsdam, Germany; rethfeldt@uni-potsdam.de (N.R.); pbrinkma@uni-potsdam.de (P.B.); daniel.riebe@alumni.uni-potsdam.de (D.R.); loeh@chem.uni-potsdam.de (H.-G.L.)

² Structural Petrology, University of Potsdam, Karl-Liebknecht-Street 24-25, 14476 Potsdam, Germany; nicole.koellner@gfz-potsdam.de (N.K.); uwe@geo.uni-potsdam.de (U.A.)

* Correspondence: beitz@uni-potsdam.de



Citation: Rethfeldt, N.; Brinkmann, P.; Riebe, D.; Beitz, T.; Köllner, N.; Altenberger, U.; Löhmansröben, H.-G. Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS. *Minerals* **2021**, *11*, 1379. <https://doi.org/10.3390/min11121379>

Academic Editor: Bernard Hubbard

Received: 20 September 2021

Accepted: 28 November 2021

Published: 7 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The numerous applications of rare earth elements (REE) has lead to a growing global demand and to the search for new REE deposits. One promising technique for exploration of these deposits is laser-induced breakdown spectroscopy (LIBS). Among a number of advantages of the technique is the possibility to perform on-site measurements without sample preparation. Since the exploration of a deposit is based on the analysis of various geological compartments of the surrounding area, REE-bearing rock and soil samples were analyzed in this work. The field samples are from three European REE deposits in Sweden and Norway. The focus is on the REE cerium, lanthanum, neodymium and yttrium. Two different approaches of data analysis were used for the evaluation. The first approach is univariate regression (UVR). While this approach was successful for the analysis of synthetic REE samples, the quantitative analysis of field samples from different sites was influenced by matrix effects. Principal component analysis (PCA) can be used to determine the origin of the samples from the three deposits. The second approach is based on multivariate regression methods, in particular interval PLS (iPLS) regression. In comparison to UVR, this method is better suited for the determination of REE contents in heterogeneous field samples.

Keywords: LIBS; rare earth elements; minerals; PCA; iPLS regression

1. Introduction

Rare earth elements (REE) are used in a wide range of modern technologies, which include, for instance, renewable energy technologies, communication technologies and petrochemistry [1–3]. Lanthanum and neodymium are relevant for electric mobility, for example, as well as for optimizing the material properties of steel alloys in conjunction with cerium. In the future, zirconium dioxide stabilized by yttrium will be applied as electrolyte material in fuel cells [2]. These diverse applications, especially in high-tech-industries, drive the global demand for REE. While in 2013 the need for REE in important future technologies accounted for about 30,900 t, studies predict an increasing demand of up to 70,900 t by 2035 [4]. Due to this increasing demand, REE are important resources [5]. Contrary to their name, they are widely distributed in the earth's crust as trace elements. Due to the chemical similarity, elements of neighboring atomic numbers can substitute each other in the crystal structures of various minerals [1,6]. The most economically important REE-containing minerals are carbonates and phosphates, such as monazite, xenotime, bastnaesite or parisite. Within the minerals, the proportions of the various REE vary. Light REE (LREE), all elements from lanthanum to europium, predominate mainly in bastnaesite and monazite, whereas xenotime is a source of heavy REE (HREE), from gadolinium to lutetium and yttrium. This distribution depends on the deposit, its

formation and type [1,3,6–8]. Although there is a large number of REE-rich minerals worldwide, mining is currently focused on a few deposits and minerals only. The mining of these resources is challenging, but the recycling of REE-containing products is also complicated [3]. Besides the aforementioned issues and the depletion of current deposits, bottlenecks in the availability of REE are expected in the future [9]. The world's largest deposits are located in China. Currently, China is the world market leader in terms of REE production. Besides other Asian countries, there are several thousand tons of REE mined per year by Australia, the USA and Russia. To date, Europe has not mined any of the existing REE deposits. Economically relevant deposits can be found mainly in Scandinavia, especially in Sweden and Finland. These potential mining areas have elevated contents of REE and could cover the supply of Europe [3,6,10,11]. At the same time, the exploration of further REE deposits is growing in importance to ensure the supply in the future. Exploration requires fast, ecological and cost-efficient analytical methods, on the one hand to minimize the amount of sample material and, thus, the chemical and geological laboratory work, and to reduce the associated costs on the other hand [12]. The most common techniques to detect and analyze REE are inductively coupled plasma mass spectrometry (ICP-MS), inductively coupled plasma optical emission spectrometry (ICP-OES), X-ray fluorescence (XRF) and neutron activation analysis (NAA). However, most of these techniques are time-consuming and therefore expensive, for example due to sample preparation. In addition, almost all of these methods require prior sample collection for analysis in the laboratory [5,13–15]. Laser-induced breakdown spectroscopy (LIBS) is a method that allows rapid on-site analysis in real time. In addition, almost no sample preparation is required, as solid, liquid and gaseous samples can be analyzed in a wide variety of matrices. In comparison to XRF, all elements of the periodic table can be measured, including the lighter elements [5,12,13,16].

In LIBS, laser radiation is used to generate a plasma on the sample surface. The resulting emission allows the analysis of the investigated samples based on the spectral signatures of the characteristic elements [13]. Although the REE have numerous emission lines, the detection and analysis of the elements is difficult due to spectral interferences of the matrices and the low concentrations of the REE [5,16]. LIBS is often used for qualitative analysis of elements or for classification of minerals. Quantitative analysis is based on univariate and/or multivariate methods. Quantification by univariate regression (UVR) is often difficult, especially in heterogeneous materials, because numerous chemical and physical effects of the complex matrix can occur [12,17,18]. In contrast to univariate methods, the use of multivariate methods for quantification can remedy this situation. While univariate regression only includes individual peaks of the spectrum in the quantification, multivariate analysis techniques, such as partial least squares (PLS) regression, can include the entire spectrum or, through variable selection, just a partial spectrum. In the case of the application of the entire spectrum in PLS regression, the regression model contains noise and redundant information. This can have a negative impact on the prediction results. An alternative is to reduce the total spectrum to selected partial spectra, which contain the most important information by variable selection. This can potentially improve the predictive power of the model [19]. One such method is interval partial least squares (*i*PLS) regression. In this method the entire spectrum is divided into equally sized spectral intervals. From the resulting spectral sub-regions, the relevant information is provided and summarized, which facilitates focusing on important spectral regions and removing interference from other regions. PLS regression is performed on one spectral interval for each element [19,20]. The width of these intervals can vary. In the case of broad intervals, *i*PLS regression resembles a normal multivariate PLS regression. However, if the intervals become smaller, the *i*PLS regression could be considered an automated UVR. *i*PLS regression was first used in IR spectroscopy, but also finds application in other spectroscopic techniques, such as LIBS on soils [20,21].

The present work focuses on the utilization of LIBS for the exploration of new REE deposits. The aim is the quantitative screening of REE in minerals, rocks and soils in the

surrounding area of a potential REE deposit using LIBS and univariate and multivariate *i*PLS regression. In this study, mineral, rocks and soil samples from the Scandinavian deposits Norra Kärr and Olserum in Sweden and Fen-Complex in Norway are investigated. Their elemental contents were determined by reference analysis (ICP-OES and XRF). The relatively small number of samples examined is a result of the high costs of sampling, transport and reference analysis. This problem accompanies geologists in their daily work and is used in this study to demonstrate the potential of our methods. In order to characterize the influence of matrix effects, synthetic REE samples produced from REE chloride salts in serpentinite and a model soil as well as the field samples are compared. All samples are analyzed with respect to the elements cerium (Ce), lanthanum (La), neodymium (Nd) and yttrium (Y). The results of the analysis with LIBS are evaluated with UVR and *i*PLS regression.

2. Materials and Methods

2.1. Samples and Reference Analysis

A total of 167 samples containing REE were analyzed. Of these, 122 were artificial samples (hereafter referred to as synthetic) and 45 were natural samples (hereafter referred to as field). In both cases, a further distinction was made between rock and soil samples. All samples were available as ground powders. The concentrations of REE in the synthetic samples are summarized in Table S1 in the Supplementary Materials. Table S2 contains the reference data of the composition of the field samples.

2.1.1. Geological Overview of Deposits

The field samples were collected at three different deposits in Sweden and Norway. 19 of these samples were taken in the Olserum-Jupval REE deposit in south east Sweden. Sampled outcrops are located ca. 7 km NW of Gambleby Västervik region (Figure 1a). REE-phosphates-rich dikes cross cut metasedimentary wall rocks (cordierite-bearing gneisses). The phosphates comprise monazite (LREE), xenotime (Y, HREE) and fluor-apatite (LREE). In addition, allanite (REE-silicate), and the REE carbonates bastnaesite and synchisite-Ce occur. The origin of the REE-deposit was studied in detail by Andersson et al. [22], Fullerton [23], Russel [24] and Gavelin [25]. REE-phosphate-rich dikes of up to 2 m thickness occur near the contact to granitoids of the Transscandinavian Igneous Belt (ca. 1.85–1.65 Ga [26]) in the metasedimentary Västervik Formation. The samples are from two dikes located at a distance of ca. 150 m (from 57.94839N to 57.96364N, from 16.35213E to 16.32522E) [22,23,27].

The Norra-Kärr Complex is a small, magmatic intrusion with an “agpaitic composition”, i.e., $(Na + K) > Al$. The complex has an elliptical shape with a north–south extension of approx. 1300 m and an east–west extension of approx. 500 m (Figure 1a,b). The Norra Kärr Complex (ca. 1.5 Ga old [28]) intruded into igneous rocks of the Trans-Scandinavian Magmatic Belt (ca. 1.8–1.7 Ga [29]). It essentially consists of various types of nepheline syenites [26]. The original plutonic rocks show evidence of strong deformation and metamorphic overprint [30]. The tectonometamorphic overprinting is assigned to the Sveconorwegian orogeny (ca. 1048 Ga [31]). The rock sequence is known as a zirconium deposit due to the high abundance of Zr-rich minerals [26,32]. The major REE-bearing minerals are the silicates eudialyte, catapleiite and cerite. Eudialyte occurs as a REE-poor magmatic type and a HREE-rich secondary (metamorphic) type, which is the major REE-bearing phase. The analyzed rock and soil samples are from different parts of the intrusion, with slightly different compositions. The REE concentrations of the rocks are up to 1 wt.%. In addition, samples were taken from outside the intrusion as a reference (subsoil and topsoil Nr. 20). Another sample obtained to the east of the intrusion was contaminated by hydrothermal fluids, a so called fenitized rock. A total of 17 samples from Norra Kärr were examined [26–32].

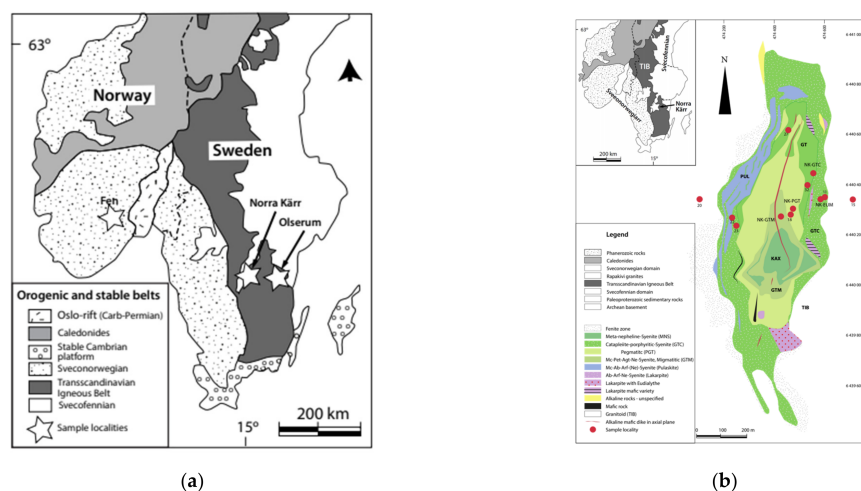


Figure 1. Maps of (a) all investigated deposits and (b) sample localities in Norra Kärr, Sweden.

The Fen Complex, located in the Telemark province of southern Norway, represents an early Cambrian magmatic intrusion complex. It is dominated by carbonatites, formed from carbonate melts [33–37]. The complex shows a surface outcrop of ca. 10 km². The previous carbonatitic melts intruded into Mesoproterozoic gneisses (Figure 1a). The carbonatites are interlayered by alkaline ultramafic silicate rocks, i.e., containing <40% SiO₂ [26]. All melts generated from a deeper mantle source. The carbonaceous intrusions modified the surrounding gneisses by CO₂-rich streams leading to alkaline metasomatic reactions. Therefore, the Fen Complex is the type locality of the “fenitisation-process”. The Fen Complex is known for its highly enriched thorium and rare earth concentrations [37]. The major REE-bearing minerals are monazite [(LREE,Th)PO₄], bastnaesite [(LREE)CO₃F] and parisite [Ca(LREE)₂(CO₃)₃F₂] [38]. Nine field samples for analyses of the Fen Complex were provided by the Federal Institute for Geosciences and Natural Resources (BGR) [38].

Reference analysis for the REE cerium, lanthanum and neodymium was carried out by inductively coupled plasma optical emission spectroscopy (ICP-OES) and X-ray fluorescence spectroscopy (XRF spectroscopy) for yttrium. The samples were available as ground powders. For LIBS measurement, all powders were pressed into pellets (TP 40, Herzog Maschinenfabrik, Osnabrück, Germany).

2.1.2. Preparation of the Synthetic Samples

The synthetic samples were prepared by standard addition. REE used as chloride hydrates (CeCl₃ · 7H₂O, LaCl₃ · 7H₂O, NdCl₃ · 6H₂O, YCl₃ · 6H₂O) were mixed with the respective matrices, soil and rock. For the samples in rock, a milled mixture of serpentinite and weathered crust was used. The matrix used in the soil-based samples is a soil containing loess from central Europe. The salts were obtained from Sigma Aldrich (St. Louis, MO, USA). The weighed rock samples were mixed with 3 wt.% distilled water, homogenized in a mill (MM 400, Retsch, Düsseldorf, Germany) and pressed into pellets (TP 40, Herzog Maschinenfabrik, Osnabrück, Germany). The soil samples were only homogenized and pressed into pellets. For one pellet, 3 g of sample is used. Pellets were prepared separately for each REE and concentration: 16 pellets for both Ce and Nd and 15 pellets for both La and Y were prepared, respectively (see Table S1 in the Supplementary Materials). Also, the pellets containing blank rock and soil were prepared for preliminary measurements but were not involved in *i*PLS regressions. The REE concentrations in the serpentinite and the soil were much less than 10 ppm.

2.2. Instrumental Set up of LIBS and Measurement Parameters

The samples were positioned on a rotating and linearly translating sample desk, resulting in a spiral-like measurement pattern. This ensured a fresh surface on the pellet for each ablation event. The light of the Nd:YAG laser (Bernoulli LIBS, Litron Lasers, Rugby, England, Great Britain, $\lambda = 1064$ nm, $E = 20$ mJ, repetition rate 10 Hz) was focused by a lens (focal length 50 mm) and created the plasma on the sample surface. Emissions were collected by a concave mirror (ME-OPT-0007, Andor Technology, Belfast, Northern Ireland, UK, focal length 52 mm, $\lambda = 200$ –1100 nm) and focused on an optical fiber to guide the radiation to an echelle spectrometer (Aryelle Butterfly, LTB, Berlin, Germany). The two wavelength ranges (UV range: 190–330 nm, VIS range: 275–750 nm) of the spectrometer were measured separately with a resolution of 20–30 pm (delay time 2 μ s). The spectrometer was equipped with an ICCD camera (iStar, Andor Technology, Belfast, UK) as detector (gate width 10 μ s). A total of about 230 spectra per sample were recorded in both the UV and VIS regions. For each spectrum, 10 single shots were accumulated.

2.3. Processing and Analysis of Data by Univariate and Multivariate Methods

Origin (OriginLab, Northampton, MA, USA) was used for univariate analysis. Matlab (Version 2019b, MathWorks, Natick, MA, USA) was used for the preprocessing and the multivariate methods PCA and *i*PLS regression.

Spectra were preprocessed by background correction and standard normal variate (SNV) normalization. A top-hat filter was used for the background correction. The utilized structure element length of 20 data points corresponds to a filter width of approximately 0.26 nm. The SNV normalization followed the background correction. For SNV normalization, the mean value of the entire spectrum was subtracted from the spectrum and the difference was divided by the standard deviation. After processing the data using background correction and SNV normalization, the 230 spectra were averaged so that one spectrum was finally obtained for each sample examined.

For the univariate regression, one element-specific peak was identified for each REE examined, and the peak area (A) was determined by integration of each of the 230 spectra and subsequent calculation mean and standard deviation. The peak area was plotted against the concentration. Principal component analysis was performed on the raw spectra of the field and synthetic samples.

For the *i*PLS regression, we used the *i*Toolbox from Nørgaard [20]. The 230 spectra were averaged prior to *i*PLS and a separate *i*PLS was performed for each element, with the only input variables being the concentrations of the element in question. Cross validation was performed internally in the *i*PLS algorithm, as well as validation externally with test data excluded from the algorithm. The *i*PLS regression was based on variable selection in spectral intervals. The number of intervals was varied, with a maximum of 5000 intervals. Interval separation was performed on raw wavelength data. Due to the way echelle data is transformed from the 2D to the (standard) 1D representation, this resulted in intervals with different spectral widths (but constant numbers of sample points). The exact number of intervals and the number of components were determined individually for each sample (Tables 4 and 5) by systematic variation of the interval length. The interval number that resulted in an interval with the smallest RMSE was selected. The PLS regression was performed at this interval. Validation of the *i*PLS regression results was performed by 5-fold cross-validation with systematic exclusion.

3. Results

3.1. Structure of REE Spectra

LIB spectra of REE-bearing minerals are highly complex due to the line-rich REE [12,13]. The NIST LIBS database displays, e.g., 7723 lines for cerium. Furthermore, REE often occur associated with other REE. In addition, strong emission lines of other elements of the surrounding matrix contribute to the complexity of the spectra. A typical LIB

spectrum of a REE-bearing mineral and an enlarged partial spectrum are displayed in Figure 2.

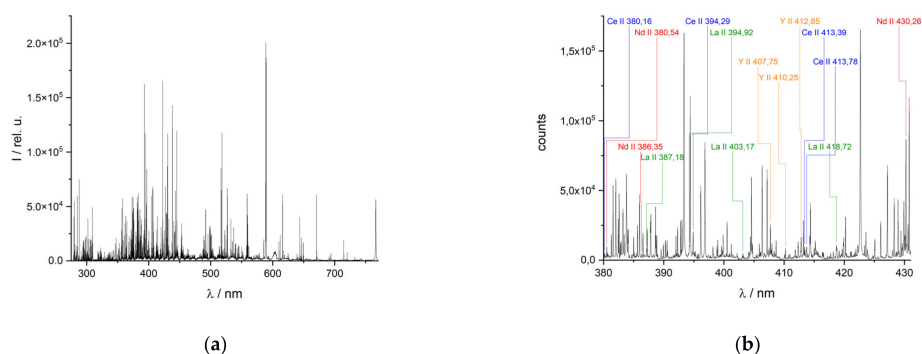


Figure 2. Representative LIB spectrum of a monazite-xenotime-apatite-bearing vein: (a) the entire spectrum ($\lambda = 275\text{--}750$ nm) and (b) a partial spectrum ($\lambda = 380\text{--}430$ nm) with important REE emission lines.

As a consequence of the large number, emission lines from different elements can be partially or even completely superimposed. This complicates the assignment of the lines to the numerous elements, especially in spectrometers with a lower resolution. The use of high-resolution spectrometers can reduce these problems in line evaluation (Figure 3). Compared to spectra with a resolution of ≥ 100 pm, spectra with a higher resolution can be obtained by spectrometers based on echelle gratings, allowing better resolved peaks to be observed. The echelle spectrometer used here has a spectral resolution of about 25–30 pm in the VIS range.

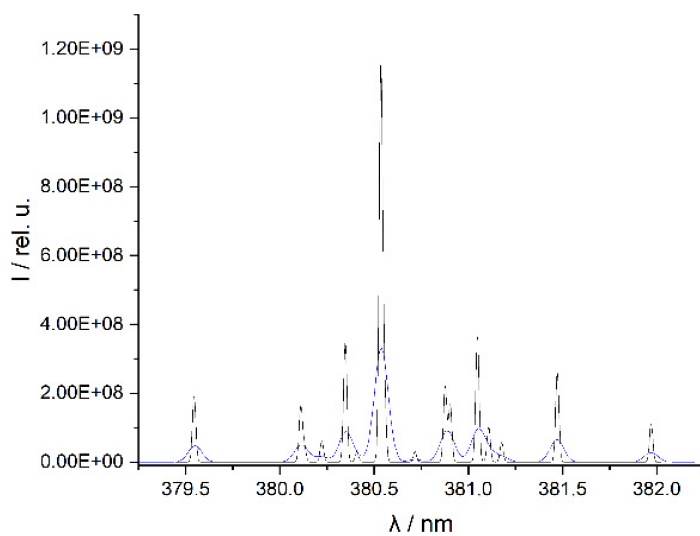


Figure 3. Spectra of Nd (element peak at 380.5 nm) simulated on the basis of the NIST database [39] with a spectral resolution of 100 pm (blue line) and 25 pm (black line).

The improved resolution of the spectra enables the separation of partially superimposed lines in lower-resolution instruments and yields a more exact line position. This allows the identification of numerous lines of REE using the NIST database (Figure 2b) [39]. The identified emission lines can then be used in quantitative elemental analysis with LIBS. Two examples of unresolved features in the simulated low-resolution spectrum (resolution of 100 pm) of Nd in Figure 3 are the line triplet at 381.05 nm, which can be completely resolved in the simulated high-resolution spectrum which mimics the spectrum of an echelle spectrometer (resolution of 25 pm), and the duplet at 380.9 nm which can only be partially resolved even in the echelle spectrometer. REE lines that overlap lines of the matrix elements are unsuitable for the investigation. These lines, marked with an asterisk in Table 1 below, are not included in the analyses performed in this work. The other spectral lines listed were used in the univariate and multivariate regression.

Table 1. Experimental spectral lines of analyzed REE (shown in Figure 2b) used in this work collected with the echelle spectrometer. Emission wavelengths marked with an asterisk (*) are not used in this work.

Species	Spectral Lines
Ce II	380.15 nm, 394.27 nm *, 413.38 nm, 413.76 nm
La I	418.73 nm *
La II	387.16 nm *, 394.91 nm, 403.16 nm
Nd II	380.53 nm, 386.34 nm, 430.22 nm *
Y I	410.24 nm, 407.73 nm *, 412.82 nm

3.2. Univariate Regression

REE do not occur as native metals in nature. Rather, they occur in various rock-forming minerals [6]. This results in field samples being complex mixtures of a wide variety of elements [13]. To evaluate the influence of the matrix elements on LIB spectra of REE, univariate regression analyses were performed on the data sets of the synthetic and field samples experimentally obtained. While the matrix of the synthetic REE samples (one rock or soil) is constant, REE in field samples are embedded in various minerals, rocks or soils. The samples in this work contain Ce, La, Nd and Y at different concentrations. For each REE investigated, a regression analysis was performed at up to 3 emission lines. The corresponding coefficients of determination (R^2) and limits of detection (LOD) are reported. The LOD is the quotient of three times the standard deviation of the baseline and the slope of the obtained regression line. The resulting values of LOD are shown in Table 2.

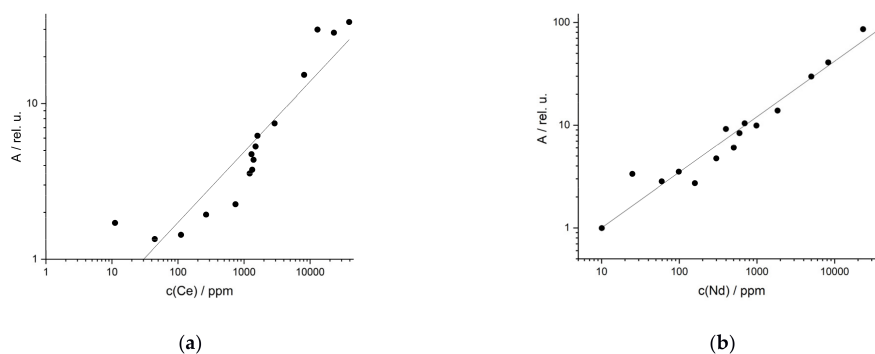
Only spectral lines which occur in the entire examined concentration range were used. Since unsuitable lines also occur for each element (Table 1), different emission lines were used for the analyses in the different matrices. In the investigated field samples, the REE content varies over several orders of magnitude (from 1 ppm up to 3%). Therefore, the presentation of the regression lines is best done by a double logarithmic plot to cover the whole concentration range. All lines investigated are summarized in Table 2 with the corresponding coefficients of determination and LOD.

3.2.1. Univariate Regression of Synthetic Samples

For synthetic REE samples of rocks, the UVR functions of the Ce line at 413.76 nm and the Nd line at 386.34 nm are shown in Figure 4 as examples. Both elements show the best regression results at these emission lines in terms of linearity, which is $R^2 = 0.91$ for Ce and $R^2 = 0.98$ for Nd. The other studied REE, La and Y, also have good coefficients of determination at the selected wavelengths: R^2 (La at 379.08 nm) = 0.99 and R^2 (Y at 412.82 nm) = 0.97.

Table 2. Results of univariate regression of synthetic samples of REE (Ce, La, Nd, Y) in rock and soil: wavelength, LOD and R^2 of best and other analyzed REE lines.

Species	λ /nm	LOD/ppm	R^2	Species	λ /nm	LOD/ppm	R^2
in rocks				best results			
Ce II	413.76	115	0.91	Ce II	413.38	120	0.90
				Ce II	456.23	130	0.90
La II	379.08	3180	0.99	La II	399.57	7700	0.99
				La II	433.37	3500	0.99
Nd II	386.34	158	0.98	Nd II	380.53	28500	0.91
Y I	412.82	61	0.97	Y I	410.24	65	0.97
				Y II	488.36	50	0.97
in soils				best results			
Ce II	413.76	285	0.99	Ce II	413.38	290	0.98
				Ce II	456.23	270	0.99
La II	404.29	160	0.97	La I	550.13	210	0.95
				La I	624.99	200	0.95
Nd II	386.34	414	0.96	Y II	363.31	270	0.95
Y II	488.36	227	0.98	Y II	437.49	290	0.94

**Figure 4.** Univariate regression of synthetic samples in rock from (a) Ce (R^2 (413.76 nm) = 0.91) and (b) Nd (R^2 (386.34 nm) = 0.98).

The results of the UVR obtained for the synthetic samples in soils were similar to the synthetic minerals although the emission lines varied. Examples are the results for the REE Ce and Y in soils (Figure 5). For Ce the emission line at 413.76 nm, and for Y the line at 488.36 nm were used, respectively. Both have high coefficients of determination (R^2 (Ce) = 0.99, R^2 (Y) = 0.98). Coefficients of $R^2 > 0.96$ were also obtained for the UVR of the elements La (404.29 nm) and Nd (386.34 nm). The other reported lines of the REE in rock and soil also yield good results, but they show reduced linearity as indicated by smaller R^2 ($R^2 = 0.90$ – 0.99 , Table 2). The detection limits of the REE in rock and soil are similar for most elements. Their average LOD is 200 ppm (except for La in rock, which exhibited an unusually high LOD as a result of the large spread of the data points, very likely due to matrix effects) with a standard deviation of 110 ppm. Nevertheless, the values of the LOD vary between elements and even between the matrices for one element, sometimes significantly.

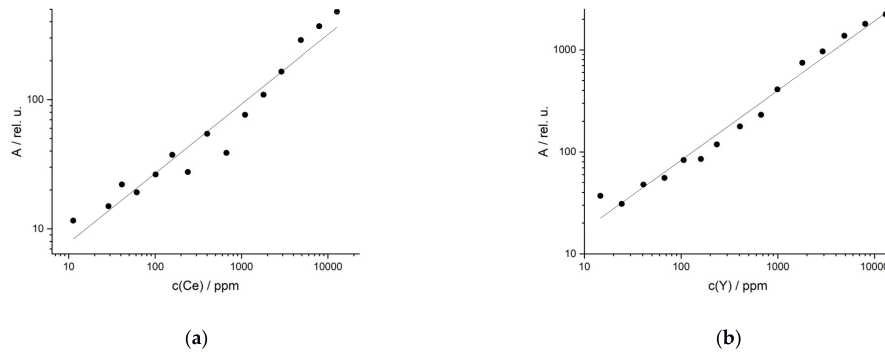


Figure 5. Univariate regression of synthetical samples in soil of (a) Ce (R^2 (413.76 nm) = 0.99) and (b) Y (R^2 (488.36 nm) = 0.98).

3.2.2. Univariate Regression of Field REE Samples

In comparison to the synthetic samples, the UVR of the field samples of REE-bearing rocks yielded worse results. This is due to two reasons: first, the matrix, which is significantly more complex than the matrix of the synthetic samples, which consists of only one rock (serpentinite). Second, there are great matrix variations between the different sample locations. The UVR results were again evaluated on the basis of the coefficients of determination. These range from 0.43 to 0.85 for field rock samples (Table 3). An example is the best La line at 408.67 nm ($R^2 = 0.82$, Figure 6a). Except for Nd, the best emission lines of the other REE investigated yield coefficients of determination around 0.80 (Table 3).

Table 3. Results of univariate regression of field samples of REE (Ce, La, Nd, Y) in rocks and soils: wavelength, LOD and R^2 of best and other analyzed REE lines.

Species	λ /nm	LOD/ppm	R^2	Species	λ /nm	LOD/ppm	R^2
in rocks				best results			
Ce II	413.76	218	0.79	Ce II	380.15	290	0.75
La II	408.67	199	0.82	Ce II	446.02	210	0.7
Nd II	325.91	1090	0.56	La II	404.29	180	0.79
Y II	437.49	117	0.85	La II	433.37	220	0.77
				Nd II	386.34	250	0.41
				Nd II	325.91	1230	0.47
				Y I	412.82	160	0.82
				Y II	321.68	140	0.80
in soils				other lines			
Ce I	594.08	359	0.51	Ce I	560.12	500	0.41
La II	404.29	294	0.52	Ce I	571.9	470	0.41
Nd II	325.91	592	0.43	La II	433.37	540	0.48
Y I	297.45	296	0.43	La I	624.99	860	0.44
				Nd II	380.53	410	0.16
				Nd II	386.34	800	0.34
				Y II	321.68	350	0.33
				Y I	410.24	600	0.20

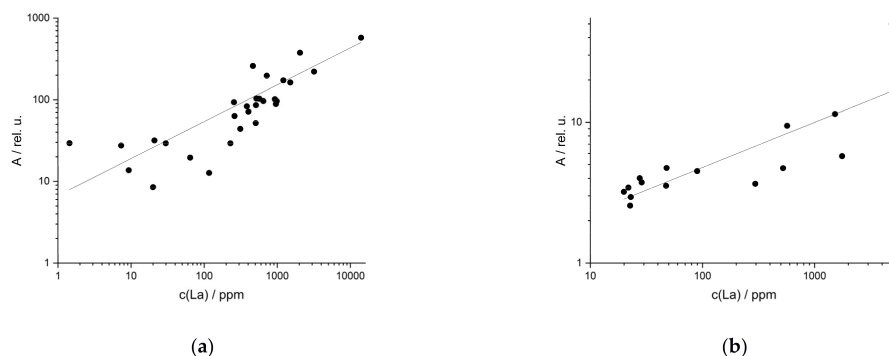


Figure 6. Univariate regression of field samples (a) in rocks from La (R^2 (408.67 nm) = 0.82) and (b) in soils from La (R^2 (404.29 nm) = 0.52).

Stronger matrix effects can be observed in the UVR of field soil samples. These samples have an even higher variance in composition, which affects the LIB spectra. UVR of the best lines yields low coefficients of determination between 0.16 and 0.52 (Table 3). This is shown in Figure 6b for the regression lines of La (404.29 nm).

The average LOD for the investigated REE in field samples (soils and rocks) is 300 ppm (except for Nd in rock) with a standard deviation of 140 ppm. Thus, the LOD values are slightly higher than the LOD of the synthetic samples. Better results of the UVR analysis were obtained for those samples that have the same and a less complex matrix composition, i.e., for the synthetic samples of rock or soil. In natural field samples with strong matrix variations, multivariate methods have the potential to compensate the variations in spectral interference caused by the differences in matrix compositions and can be an alternative to the UVR.

3.3. Characterization of Matrix Effects by Means of PCA

One way to investigate these influences of the matrices is to characterize the differences in the chemical composition of soil and rock. This is possible by means of unsupervised methods such as principal component analysis (PCA). For this purpose, the LIB spectra of the soils and rocks collected near the three REE deposits are investigated by PCA. The first two principal components are shown in the score plot (Figure 7). For the reference rocks, 82.8% of the variance can be explained by the first two components (Figure 7a). In the case of the reference soils, 96.7% of the variance is explained by the two components (Figure 7b). In the score plot of the rock samples (Figure 7a), the samples of the three deposits form well separated clusters, which are displayed in different colors. The samples of the individual deposit clusters also show a considerable scattering, especially in the case of the Norra Kärr samples. This allows the determination of the provenance of the samples and also indicates the strong differences in the chemical composition of the matrices within and between the deposits, which is caused by the types of rocks present in the individual deposits. All three deposits are composed of siliceous or carbonaceous rocks. However, the included minerals influence the composition and thus allow a differentiation of the deposits by PCA. The igneous rock of the Swedish Norra Kärr deposit (red) is a nepheline-bearing syenite with apatitic composition. These samples are particularly rich in sodium and potassium. In contrast, the Olserum deposit (green) has a mainly cordierite-bearing gneiss. As a result, the siliceous rock is magnesium- and aluminum-bearing. The rocks of the Fen complex (blue) are carbonatites. Due to chemical and physical weathering processes and accompanying degradation, the data points are partly scattered over large areas even within the clusters. In addition, the widely spaced sampling locations (>50 m) contribute

to the splitting of the data points. The points from the Norra Kärr deposit show the largest scatter. Here the sampling locations are up to 200 m apart. The sampled veins in Olserum are about 150 m apart.

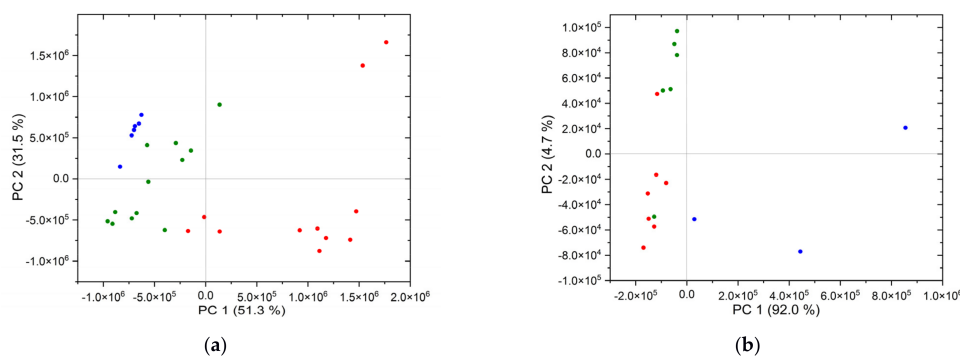


Figure 7. (a) PCA of reference rocks (VIS, raw), (b) PCA of reference soils (VIS, raw): blue—Fen complex, green—Olserum, red—Norra Kärr.

For the soils, the interpretation of the PCA (Figure 7b) is more complex, since the soils consist not only of physically weathered rock, but also of organic matter from the biosphere and atmospheric dust [40]. Nevertheless, as in the principal component analysis of the reference rocks, a clustering of the data points according to the deposits and their separation from each other can be observed. This probably results from the rocks present. Although the sampling locations in the deposits are far apart, the data points within the PCA are not as split up as in the case of the reference rocks. This could be due to incoming organic matter and atmospheric dust during formation. Thus, the samples are possibly more homogeneous. In contrast, the strong splitting of the data points from the Fen complex (blue) is particularly striking. This is probably due to the different weathering grade of the soils.

In comparison, both PCA show similar results: data points are grouped by deposit and separated from each other. All clusters show an internal separation of the data points. The explained variance of both PCA analyses is already >82% for two principal components. Thus, it is possible to assign the present samples to a single deposit. Surprisingly, the explained variance of the reference soils is larger than that of the rocks. The reason for this could be the nature of the samples. The soil samples are in powder form and thus show a greater macroscopic homogeneity than the investigated rocks. As a result, the properties are more evenly distributed, so that the components of the PCA capture more properties. Thus, the reference soils can be explained with fewer principal components.

However, the differences in the chemical composition of the samples are also evident in both analyses. This heterogeneity of the field samples is striking in the PCA of field and synthetic samples (Figure 8a,b). In addition to the LIB spectra of the reference rocks, the spectra of the synthetic samples (Ce) were included in the analysis. From the principal component analysis (Figure 8a) it can be seen that the synthetic samples are more homogeneous. These data points (orange) are concentrated in a small area of the score plot. Nevertheless, they are not located at exclusively one point in the score plot, as the magnification in Figure 8b shows. In comparison, the data points of the field samples are distributed over a wide area in the plot. This supports our observation that the separation between the samples from different sites is caused by the differences in the chemical composition of the matrix and not by the variations in REEs concentration.

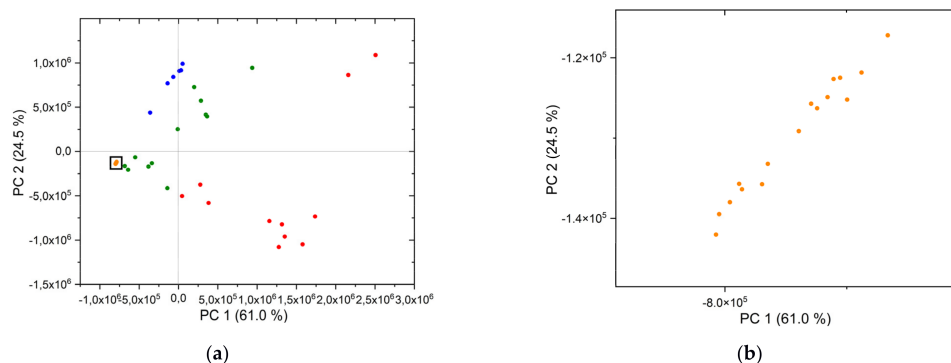


Figure 8. (a) PCA of reference rock and synthetic samples (Ce), (b) magnification of area with synthetic samples (Ce): blue Fen complex, green Olserum, red Norra Kärr, orange synthetic samples (Ce).

After performing PCA, it is clear that UVR is well suited for the synthetic samples because they have a more homogeneous composition. For field samples, on the other hand, univariate regression is unsuitable due to the heterogeneous matrix. For heterogeneous samples, the use of multivariate regression models is therefore examined.

3.4. Interval Partial Least-Squares (iPLS) Regression

The multivariate method used was interval partial least-squares (iPLS) regression based on variable selection. In this method, the spectra are divided into intervals of equal spectral width. The number of intervals and the number of components were determined based on the lowest RMSE individually for each sample and are reported in Tables 4 and 5.

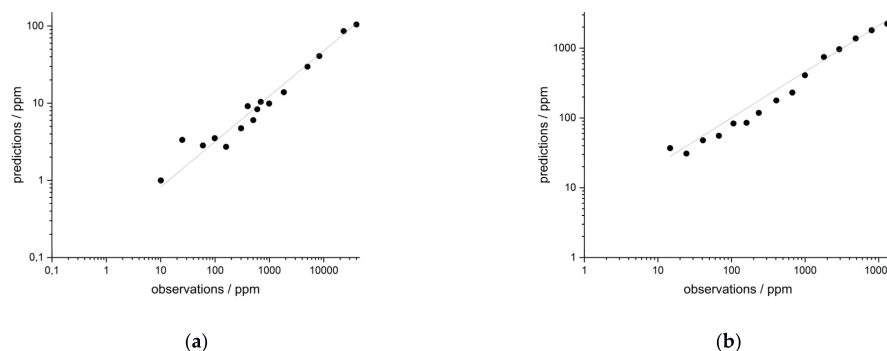
Table 4. Results of iPLS regression of synthetic samples in rocks and soils.

Species	# Intervals	Interval Width/nm	iPLS: R ²	PLS: R ²
	# Components	Element Line/nm	RMSECV	RMSECV
synthetical samples in rocks				
Ce II	3000 intervals	0.02	0.98	0.88
VIS	10 components	507.50	1500 ppm	3719 ppm
La II	20 intervals	20.35	0.99	0.99
VIS	4 components	398.8, 399.57, 403.2, 404.29, 407.7, 408.67	490 ppm	391 ppm
Nd II	1 interval	490.56	0.98	0.96
VIS	3 components	380.53	1650 ppm	2250 ppm
Y II	3000 intervals	0.05	0.99	0.73
UV	10 components	324.20	590 ppm	5130 ppm
synthetical Samples in Soils				
Ce II	2000 intervals	0.20	0.99	0.93
VIS	12 components	407.90, 408.00	280ppm	840 ppm
La II	2000 intervals	0.19	0.96	0.24
VIS	6 components	394.90	0.42%	3205 ppm
Nd II	3000 intervals	0.16	0.95	0.73
VIS	7 components	532.00	0.45%	1903 ppm
Y II / I	800 intervals	0.46	0.98	0.95
VIS	4 components	374.80, 374.90	480 ppm	827 ppm

Table 5. Results of *i*PLS regression of field samples in rocks and soils.

Species	# Intervals	Interval Width/nm	<i>i</i> PLS: R ²
	# Components	Element Line/nm	RMSECV
field samples in rocks			
Ce II	2000 intervals	0.288	0.92
VIS	18 components	639.30	1680 ppm
La I	5000 intervals	0.078	0.92
VIS	3 components	495.00	760 ppm
Nd I	400 intervals	1.30	0.89
VIS	15 components	521.30	1000 ppm
Y I	1000 intervals	0.52	0.91
VIS	4 components	552.76	870 ppm
field samples in soils			
Ce II	400 intervals	1.24	0.83
VIS	12 components	502.30	1000 ppm
La II	1800 intervals	0.269	0.93
VIS	11 components	497.00	350 ppm
Nd II	1000 intervals	0.164	0.83
VIS	12 components	307.50	420 ppm
Y II	2100 intervals	0.17	0.84
VIS	11 components	371.00	200 ppm

The method was first tested on the more homogeneous synthetic samples. As shown by the results in Table 4, the *i*PLS regression provides similarly good and, in some cases, even better results than the UVR for these samples. Thus, for the synthetic samples of cerium in rock, a better coefficient of determination of 0.98, compared to the UVR ($R^2 = 0.91$, Cf. Table 2), could be obtained. An improvement of the regression could also be accomplished for yttrium (R^2 (*i*PLS) = 0.99, R^2 (UVR) = 0.97). For lanthanum and neodymium (Figure 9a) the coefficients of determination of *i*PLS regression and UVR are in the same good range.

**Figure 9.** *i*PLS regression of synthetic samples (a) in rock of Nd and (b) in soils of Y.

Similar results were obtained for the synthetic soil samples (see univariate data in Table 2 and data of *i*PLS regression in Table 4). Figure 9b shows the *i*PLS results of yttrium in soils. For the investigated REE cerium, lanthanum, neodymium and yttrium, coefficients of determination between $R^2 = 0.95$ – 0.99 were achieved. These are similar to the R^2 of the UVR ($R^2 = 0.96$ – 0.99). The fact that the multivariate method performs as well or even better

for the synthetic samples suggest that their application to heterogeneous field samples could be beneficial.

The application of *i*PLS regression to the heterogeneous field samples yields significantly better regression results than the UVR in all cases investigated. For samples in rock, $R^2 > 0.91$ can be obtained (Table 5). For lanthanum (Figure 10a), in contrast to the UVR ($R^2 = 0.82$), a coefficient of determination of $R^2 = 0.92$ can be achieved by *i*PLS regression. For neodymium, a R^2 of 0.98 is found by using the multivariate regression method (Figure 10b), which is also significantly better than for UVR (R^2 (UVR) = 0.56). Somewhat less pronounced, but equally good, are the improvements in the regressions of cerium and yttrium in rock.

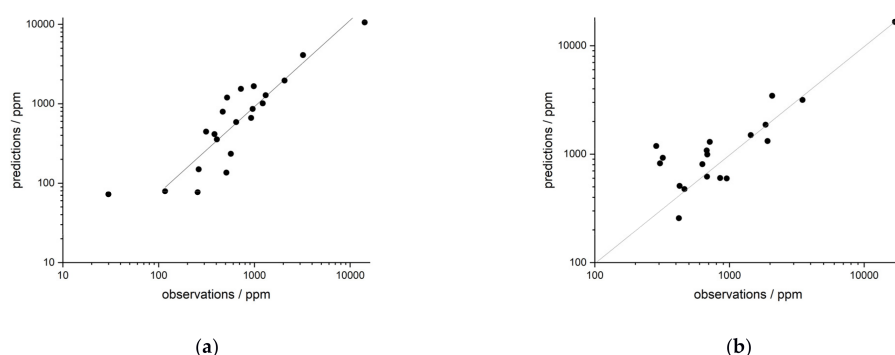


Figure 10. *i*PLS regression of field samples in rocks of (a) La and (b) Nd.

The improvement of the regression is most obvious for the field samples in soils. The best coefficient of determination achieved for a UVR was $R^2 = 0.52$. All coefficients of determination obtained for *i*PLS regression are better than $R^2 > 0.83$ (Table 5). For lanthanum, the univariate regression value $R^2 = 0.52$ can be increased to $R^2 = 0.93$ in the multivariate regression. The improvement is similarly strong in the case of neodymium. Comparable to this are the changes in regression for yttrium and cerium.

In the *i*PLS regression, only the intervals containing relevant information, such as the element lines, are considered (Tables 4 and 5). To that end, very small interval widths had to be used in most cases. The subsequent PLS regression could thus also be considered an automated UVR. Nevertheless, the *i*PLS regression has higher coefficients of determination than the UVR. In addition to *i*PLS, PLS regression of the entire spectra was also examined. However, this method was not able to provide a prediction better than guessing. This demonstrates the advantage of variable selection in *i*PLS. Due to the sometimes extremely narrow intervals, only parts of the relevant element lines are included in the regression. Negative effects, such as self-absorption and partial contamination of the flank of lines, are excluded, resulting in improved regressions with higher coefficients of determination [41].

In particular, the quality of the regression of the field soil samples shows the strength of the multivariate method *i*PLS regression chosen here. Whereas, in the univariate regression, coefficients of determination of $R^2 = 0.47$ are achieved on average, the regression can be improved to a mean $R^2 = 0.88$. In addition to selecting appropriate spectral regions that contain relevant information, variable selection also allows elimination of intervals that contain noise or other irrelevant data. This can consider the influence of matrix effects of the studied rocks and soils.

4. Summary and Conclusions

LIBS is a promising method for the detection of rare earth elements in rocks and soils, as this technique can be used to detect REE in various matrices with no or minimal sample

preparation. However, the spectral influence due to the matrices is problematic. In order to consider this influence and increase the power of the analysis, multivariate methods are required, especially for heterogeneous samples. In the present work, two approaches were used to analyze the LIBS data. The first approach is the UVR based on the peak area of the relevant LIBS signals. Here, the matrix effects of the heterogeneous field samples were identified as problematic and further investigated using PCA. The influence of these effects is considered in the second approach, a multivariate method. An *i*PLS regression was used in which only relevant regions of the spectrum were included in the regression. This improved the predictive accuracy of the regression. The investigations show that the detection and evaluation of REE in rocks and soils is possible using LIBS in combination with multivariate *i*PLS regression. The use of mobile LIB spectrometers can thus potentially enable the exploration of new REE deposits.

To increase the predictive power of the method, further investigations should be extended by samples from other deposits. Of particular interest here are REE-bearing samples, which have different accompanying matrices. For future studies, larger intervals and combinations of individual intervals could be used with a smaller number of components. For possible applications in the field, the use of a mobile spectrometer should be further investigated. Should these investigations prove successful, the transfer of the methods developed here into the field could be a promising possibility.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/min11121379/s1>, Table S1: Concentrations of the synthetic samples in rock and soil, Table S2. Reference data of the field samples.

Author Contributions: Conceptualization, D.R., T.B., U.A. and H.-G.L.; Data curation, N.R., P.B., D.R. and T.B.; Formal analysis, N.R. and P.B.; Funding acquisition, D.R., T.B. and H.-G.L.; Investigation, N.R. and D.R.; Methodology, D.R. and T.B.; Project administration, D.R., T.B. and H.-G.L.; Resources, T.B., N.K. and H.-G.L.; Supervision, H.-G.L.; Visualization, N.R. and P.B.; Writing—original draft, P.B. and T.B.; Writing—review & editing, D.R., N.K., U.A. and H.-G.L. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support of this research work within the framework of the ProFIT funding of the State of Brandenburg (ILB) in the LIBSqORE project (No. 80172489) and the InfraFEI grant of the State of Brandenburg (ILB) in the FuSeSE project (No. 85045759). We would like to thank the Federal Ministry of Education and Research (BMBF), which financed parts of the sampling campaign and ICP analysis of natural samples in the project “r4—Innovative Resource Efficiency Technologies Research for the Provision of Strategic Economic Raw Materials”. We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

Data Availability Statement: The concentration of the REE of the natural samples used for LIBS are all analyzed at the Institute of Geosciences, University Potsdam (UP). They are part of a diploma and a master thesis of the UP (Köllner, 2015; Sieg, 2016) and unpublished analyses by one author (U.A.). Some data are obtained by the project of the Federal Ministry of Education and Research (BMBF) “r4—Innovative Resource Efficiency Technologies Research on the provision of strategic economic raw material”. All data are available on request (uwe@geo.uni-potsdam.de).

Conflicts of Interest: The authors declare no conflict of interest.

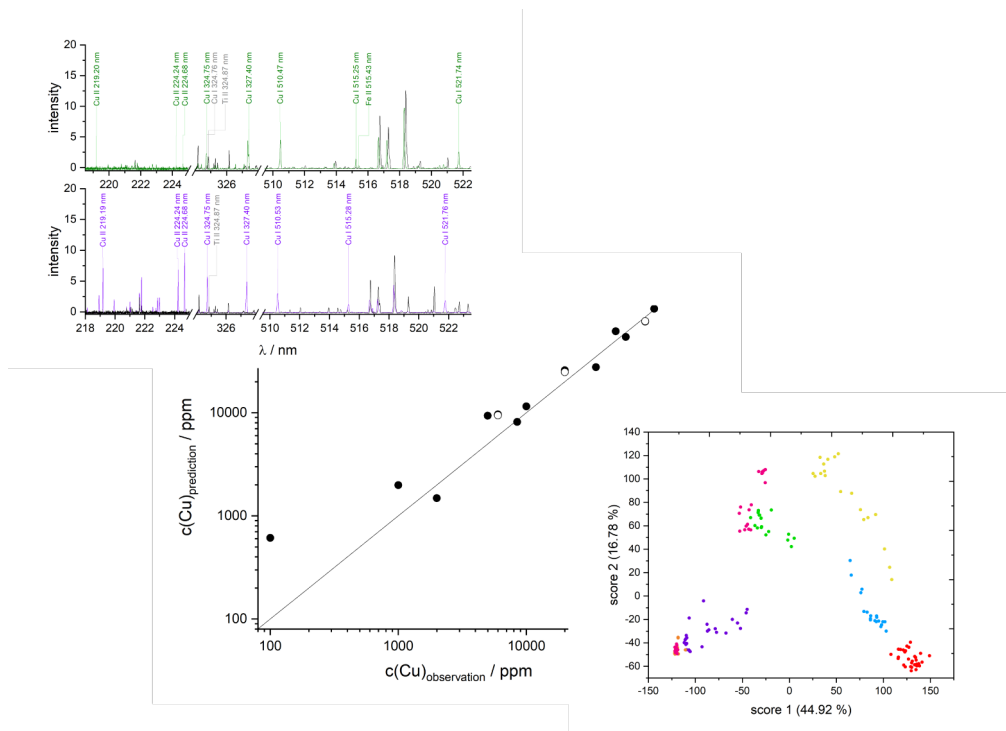
References

1. Altenberger, U.; Oberhänsli, R. Vom Atom Zum Hightec-Produkt. Minerale Der Seltenerdelemente Als Rohstoffe. *PdN Chem. Schule/Mineralien Erze* **2012**, *7*, 6–12.
2. Sicius, H. *Seltenerdmetalle: Lanthanoide Und Dritte Nebengruppe*; Essentials; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2015; ISBN 978-3-658-09839-1.
3. Balam, V. Rare Earth Elements: A Review of Applications, Occurrence, Exploration, Analysis, Recycling, and Environmental Impact. *Geosci. Front.* **2019**, *10*, 1285–1303. [[CrossRef](#)]
4. Marscheider-Weidemann, F.; Langkau, S.; Hummen, T.; Erdmann, L.; Tercero Espinoza, L.; Angerer, G.; Marwede, M.; Benecke, S. Rohstoffe für Zukunftstechnologien 2016.—DERA Rohstoffinformationen 28: 353 S., Berlin. Available online: <https://www.dera.de>

- deutsche-rohstoffagentur.de/DERA/DE/Downloads/Studie_Zukunftstechnologien-2016.pdf?__blob=publicationFile&v=3 (accessed on 6 November 2021).
5. Gaft, M.; Raichlin, Y.; Pelascini, F.; Panzer, G.; Motto Ros, V. Imaging Rare-Earth Elements in Minerals by Laser-Induced Plasma Spectroscopy: Molecular Emission and Plasma-Induced Luminescence. *Spectrochim. Acta Part B At. Spectrosc.* **2019**, *151*, 12–19. [CrossRef]
 6. Walters, A.; Lusty, P.; Hill, A. *Rare Earth Elements*; British Geological Survey: Nottingham, UK, 2011.
 7. Mehmood, M. Rare Earth Elements—A Review. *J. Ecol. Nat. Resour.* **2018**, *2*. [CrossRef]
 8. Romppanen, S.; Häkkinen, H.; Kaski, S. Singular Value Decomposition Approach to the Yttrium Occurrence in Mineral Maps of Rare Earth Element Ores Using Laser-Induced Breakdown Spectroscopy. *Spectrochim. Acta B* **2017**, *134*, 69–74. [CrossRef]
 9. Erdmann, L.; Behrendt, S. Kritische Rohstoffe Für Deutschland. *Izt* **2011**, 1–13. Available online: <https://www.izt.de/fileadmin/publikationen/54416.pdf> (accessed on 6 November 2021).
 10. Brown, T.J. *World Mineral Production*; British Geological Survey: Nottingham, UK, 2020; ISBN 9780511763175.
 11. Goodenough, K.M.; Wall, F.; Merriman, D. The Rare Earth Elements: Demand, Global Resources, and Challenges for Resourcing Future Generations. *Nat. Resour. Res.* **2018**, *27*, 201–216. [CrossRef]
 12. Müller, S.; Meima, J.A.; Rammlmair, D. Detecting REE-Rich Areas in Heterogeneous Drill Cores from Storkwitz Using LIBS and a Combination of K-Means Clustering and Spatial Raster Analysis. *J. Geochem. Explor.* **2021**, *221*, 106697. [CrossRef]
 13. Bhatt, C.R.; Jain, J.C.; Goueguel, C.L.; McIntyre, D.L.; Singh, J.P. Determination of Rare Earth Elements in Geological Samples Using Laser-Induced Breakdown Spectroscopy (LIBS). *Appl. Spectrosc.* **2018**, *72*, 114–121. [CrossRef]
 14. Zawisza, B.; Pytlakowska, K.; Feist, B.; Polowniak, M.; Kita, A.; Sitko, R. Determination of Rare Earth Elements by Spectroscopic Techniques: A Review. *J. Anal. At. Spectrom.* **2011**, *26*, 2373–2390. [CrossRef]
 15. Fabre, C.; Devismes, D.; Moncayo, S.; Pelascini, F.; Trichard, F.; Lecomte, A.; Bousquet, B.; Caudiz, J.; Motto-Ros, V. Elemental Imaging by Laser-Induced Breakdown Spectroscopy for the Geological Characterization of Minerals. *J. Anal. At. Spectrom.* **2018**, *33*, 1345–1353. [CrossRef]
 16. Labutin, T.A.; Zaytsev, S.M.; Popov, A.M.; Zorov, N.B. A Novel Approach to Sensitivity Evaluation of Laser-Induced Breakdown Spectroscopy for Rare Earth Elements Determination. *J. Anal. At. Spectrom.* **2016**, *31*, 2223–2226. [CrossRef]
 17. Anderson, R.B.; Morris, R.V.; Clegg, S.M.; Humphries, S.D.; Wiens, R.C.; Bell III, J.F.; Mertzman, S.A. Partial Least Squares and Neural Networks for Quantitative Calibration of Laser-Induced Breakdown Spectroscopy (LIBS) of Geologic Samples. In Proceedings of the Lunar and Planetary Science, The Woodlands, TX, USA, 1–5 March 2010.
 18. Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.G.; Gebbers, R. Soil Nutrient Detection for Precision Agriculture USING Handheld Laser-Induced Breakdown Spectroscopy (LIBS) and Multivariate Regression Methods (PLSR, Lasso and GPR). *Sensors* **2020**, *20*, 418. [CrossRef]
 19. Andersen, C.M.; Bro, R. Variable Selection in Regression—A Tutorial. *J. Chemom.* **2010**, *24*, 728–737. [CrossRef]
 20. Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J.P.; Munck, L.; Engelsen, S.B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54*, 413–419. [CrossRef]
 21. Sánchez-Esteva, S.; Knadel, M.; Kucheryavskiy, S.; de Jonge, L.W.; Rubæk, G.H.; Hermansen, C.; Heckrath, G. Combining Laser-Induced Breakdown Spectroscopy (LIBS) and Visible near-Infrared Spectroscopy (Vis-NIRS) for Soil Phosphorus Determination. *Sensors* **2020**, *20*, 5419. [CrossRef]
 22. Andersson, S.S.; Wagner, T.; Jonsson, E.; Fusswinkel, T.; Leijd, M.; Berg, J.T. Origin of the High-Temperature Olserum-Djupedal REE-Phosphate Mineralisation, SE Sweden: A Unique Contact Metamorphic-Hydrothermal System. *Ore Geol. Rev.* **2018**, *101*, 740–764. [CrossRef]
 23. Fullerton, W. REE Mineralisation and Metasomatic Alteration in the Olserum Metasediments. Master's Thesis, Lund University, Lund, Sweden, 2014.
 24. Russell, R.V. Porphyroblastic Differentiation in Fleck Gneiss from Västervik, Sweden. *Geologiska Föreningen i Stockholm Förhandlingar* **1969**, *91*, 217–282. [CrossRef]
 25. Gavelin, S. *The Västervik Area in South-Eastern Sweden. Studies in Proterozoic Sedimentation, High-Grade Metamorphism and Granitization*; Sveriges Geologiska Undersökning: Uppsala, Sweden, 1984; Volume 32.
 26. Adamson, O.J. The Petrology of the Norra Kärr District: An Occurrence of Alkaline Rocks in Southern Sweden. *GFF* **1944**, *437*, 113–255. [CrossRef]
 27. Sieg, L. Distribution of Rare Earth Elements in Rocks, Soils and Plants from Olserum and Norra Kärr, Sweden. Master's Thesis, University of Potsdam, Potsdam, Germany, 2016, unpublished.
 28. Sjöqvist, A.S.L.; Cornell, D.H.; Andersen, T.; Christensson, U.I.; Berg, J.T. Magmatic Age of Rare-Earth Element and Zirconium Mineralisation at the Norra Kärr Alkaline Complex, Southern SWEDEN, Determined by U–Pb and Lu–Hf Isotope Analyses of Metasomatic Zircon and Eudialyte. *Lithos* **2017**, *294–295*, 73–86. [CrossRef]
 29. Gorbatshev, R. *The Transscandinavian Igneous Belt—Introduction and Background*; Högdahl, K., Eklund, U.B., Andersson, O., Eds.; Geological Survey of Finland: Espoo, Finland, 2004; ISBN 9516908896.
 30. Atanasova, P.; Marks, M.A.W.; Heinig, T.; Krause, J.; Gutzmer, J.; Markl, G. Distinguishing Magmatic and Metamorphic Processes in Peralkaline Rocks of the Norra Kärr Complex (Southern Sweden) Using Textural and Compositional Variations of Clinopyroxene and Eudialyte-Group Minerals. *J. Petrol.* **2017**, *58*, 361–384. [CrossRef]

3.4 Vergleich von Handheld- und Echelle-Spektrometer zur Bestimmung von Kupfergehalten in Erzen mit Hilfe der laserinduzierten Breakdownspektroskopie (*Manuskript 3*)

„Comparison of Handheld and Echelle Spectrometer to assess Copper in Ores by means of Laser-Induced Breakdown Spectroscopy (LIBS)“



eingereicht bei

Minerals (2022)



Article

Comparison of handheld and Echelle spectrometer to assess copper in ores by means of laser-induced breakdown spectroscopy (LIBS)

Pia Brinkmann¹, Nicole Köllner², Sven Merk³, Toralf Beitz¹, Uwe Altenberger⁴, Hans-Gerd Löhmannsröben¹

¹ Physical Chemistry, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany; pbrinkma@uni-potsdam.de (P.B.); beitz@uni-potsdam.de (T.B.); loeh@chem.uni-potsdam.de (H.-G.L.)

² Remote Sensing, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany; nicole.koellner@gfz-potsdam.de (N.K.)

³ LTB Lasertechnik Berlin GmbH, Am Studio 2c, 12489 Berlin, Germany; sven.merk@ltb-berlin.de (S.M.)

⁴ Structural Petrology, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany; uwe@geo.uni-potsdam.de (U.A.)

* Correspondence: beitz@uni-potsdam.de

Abstract: Its properties make copper one of the world's most important functional metals. Numerous megatrends are increasing the demand for copper. This requires the prospection and exploration of new deposits, as well as the monitoring of copper quality in the various production steps. A promising technique to perform these tasks is Laser Induced Breakdown Spectroscopy (LIBS). Its unique feature, among others, is the ability to measure on-site without sample collection and preparation. In this work, copper-bearing minerals from two different deposits are studied. The first set of the field samples comes from a volcanogenic massive sulfide (VMS) deposit, the second part comes from a stratiform sedimentary copper (SSC) deposit. Different approaches are used to analyze the data. First, univariate regression (UVR) is used. However, due to the strong influence of matrix effects, this is not suitable for quantitative analysis of copper grades. Second, the multivariate method of partial least squares regression (PLSR) is used, which is more suitable for quantification. In addition, the effects of the surrounding matrices on the LIBS data are characterized by principal component analysis (PCA), alternative regression methods to PLSR are tested and the PLSR calibration is validated using field samples.

Keywords: LIBS; copper-bearing minerals; UVR; PCA; PLSR

1. Introduction

Non-ferrous metals (NF metals) are important materials in numerous areas of modern life. NF metals are light and noble metals, such as copper, aluminum and gold [1]. They are used in a wide range of applications, particularly in mechanical and vehicle engineering, the electronics and electrical engineering industries, and in the building industry. They are also a fundamental component of future technologies, such as electric traction engines and thin film photovoltaic. As a result, the demand for NF metals will increase in the future [2]. Various options are available to secure the supply of raw materials. In addition to raw material substitution and more efficient use of resources in production and application as well as in recycling, ore mining and thus metal extraction can be optimized [2]. The production of commodity metals comprises two main areas: mining, and thus also the prospection and exploration of deposits, and the metallurgical process of the extracted ores and concentrates.

Prospection and exploration of a potential deposit involves many steps and usually takes years. Numerous methods and techniques are used in locating and exploring new profitable deposits, including: remote sensing, geophysical techniques, mapping, drilling, and finally geochemical methods [3]. All these investigations are time-consuming and cost-intensive. In particular, the systematic drilling and sampling as well as their analysis in the laboratory are expensive [3,4]. Alternatives could be investigation methods that can be performed on site by using mobile sensor equipment. Examples of these include mobile X-ray fluorescence analysis (XRF) or laser-induced breakdown spectroscopy (LIBS), which allow to analyze the chemical (elemental) composition in situ [3]. An explored deposit becomes a mine, if the ore content and extent, as well as the infrastructure of the deposit make mining profitable [5]. This is done by open pit, underground mining or by leaching. In all cases, the ore is concentrated on site by separating ore minerals and gangue. Subsequently, metallurgical process of the extracted NF ores and NF concentrates takes place [1,3]. Pretreatment, e.g. by roasting and flotation, is followed by melting processes and further thermal treatments. This is followed by thermal and electrical refining and final processing [1]. This procedure leads to a gradual increase in the level of purity. Monitoring metal concentration before, during and after the production process is essential to ensure the quality of the product. Particular attention is paid to the incoming inspection of the ores. In the NF metals industry, various sensor-based techniques are used for incoming and process control. They provide information on properties and surface characteristics of particles and ores as well as on chemical and mineral composition [6]. The most commonly used sensor techniques in the NF metals industry include X-ray fluorescence (XRF), X-ray transmission (XRT), near-infra-red (NIR), as well as radiometric, optical and inductive sensors [6,7]. In particular, the use of LIBS would also be feasible for ore or concentrate input control. The method allows on-site, real-time analysis. In addition, sample material of all aggregate states can be examined almost without sample preparation and practically non-invasively. In LIBS, laser radiation induces a plasma on the sample surface, which excites the sample material. The elements can be identified from the spectral signature of the resulting emission [8,9]. In contrast to other spectroscopic methods, such as XRF, LIBS can be used to analyze all elements of the periodic table [9]. Thus, LIBS is suitable for the analysis of all NFs metals.

Copper is a superlative NF metal: highest electrical conductivity, highest thermal conductivity, highest ductility, only excelled by silver [10,11]. In addition to these properties, the good ductility, corrosion resistance, antibacterial effect and the possibility of recycling the raw material without loss of quality, make copper one of the world's most important functional metals [12,13]. The demand for copper is high and will continue to increase in the upcoming years, driven by megatrends such as digitalization, energy and transport transformation [13]. Copper occurs in nature in its native form or in minerals [14,15]. To date, about 630 copper-bearing minerals are known [16]. The copper ores of economic interest include, in addition to native copper, the sulfidic ores chalcocite (Cu_2S) and chalcopyrite (CuFeS_2), and the oxidic ores cuprite and malachite [3,13,14,17,18]. The most important ore for copper production is chalcopyrite due to its frequent occurrence [3]. Copper is mainly extracted by open pit mining, but also by underground mining and leaching [5,13]. However, copper mining has to accept new challenges. Environmental protection, land consumption and social responsibility are essential aspects that must be taken into account in modern mining [3]. At the same time, changes in ore quality are to be expected due to lower ore contents and toxic accompanying elements such as arsenic, mercury and bismuth [13,19]. The mineability of copper deposits has declined worldwide over the last 100 years, as the demand for copper is constantly increasing. Currently, raw material deposits with copper contents of 0.4 % are mineable [13]. In order to be able to meet the increasing

demand for copper and to replace depleted mines, new deposits are being prospected [3].

Copper mining is currently being carried out at numerous deposits worldwide. The main sources of copper, in addition to porphyritic, hydrothermally induced deposits, include volcanogenic massive sulfide deposits (VMS deposits) and stratiform sedimentary deposits (SSC deposits) [3,13,14,20]. VMS deposits are accumulations of sulfides previously formed by black smokers, i.e. geothermal ocean floor vents exhaling superheated metal containing water, in the deep sea. Besides copper, they are sources of several other metals. They are characterized by a wide distribution and a high quality of their ores. Based on mineralogy, several types can be distinguished. One representative of the VMS deposits is the so-called Cyprus type, which describes important copper deposits on Cyprus [3,12,14]. These are mostly small deposits that have medium contents of copper and zinc. Copper-rich veins are also found beneath the ophiolite-bound basalts. Cyprus was already a mining area and production site for copper in the early Bronze Age [3,21]. At its peak, 17 copper mines were active on Cyprus. Today, only the Skouriotissa mine is still in use. It is one of the largest deposits on the island, with 6 million tons of copper estimated total reserves [3,21]. In this "Cyprus-type" deposits, the copper-bearing minerals occur in various associated rocks, such as gabbros or basalts [12]. The main copper-bearing ore minerals of the SSC deposit are chalcocite, bornite and chalcopyrite. One of the best-known European deposits is the Polish Kupferschiefer (copper schist) in the Lubin district. It belongs to the Zechstein Kupferschiefer formation and is part of the world's most important copper deposits. About 500.000 tons of copper are extracted from the schist here annually [3,12,22].

The mined sulfide copper ore contains a maximum of 2 % copper. It is processed still on site. The resulting copper concentrate already has a copper content of up to 30 %. The sulfidic ore concentrates are heated to copper matte (~ 64 % copper), which is then roasted. In the process, sulfuric acid is recovered as a by-product of copper production. In further steps, accompanying elements, such as iron, have to be removed. These components are separated as slag. Through further roasting and heating, blister copper and finally coarse copper is obtained from the copper matte, which has a copper content of up to 98 %. After further oxidizing and reducing production steps, the obtained tough-pitch copper (99.5 % copper) contains only a few impurities and can be converted into cathode copper by electrolytic refining. At 99.99 %, cathode copper is the commercial metal with the highest purity. From this, numerous copper forms are produced, which are finally transferred to end-use products. Other important NF metals are recovered from the anode slime [3,5,16,17].

With regard to exploration and smelting, LIBS can be used to determine the qualitative as well as the quantitative elemental composition of copper-bearing ores and concentrates. In addition to the content of copper in the raw material, the determination of the accompanying elements is also relevant for smelting, since different processes are used for metal extraction from sulfide and oxide ores [5]. By controlling the copper quality in the individual production steps, of the cathode copper and of the final products, their quality grade can be assured. In this respect, LIBS is often used for qualitative analysis and classification. The determination of the content of the elements contained in a sample is difficult using LIBS because the method has a limited reproducibility due to the experimental conditions [23]. In addition, the surrounding matrix can exhibit high variability in its composition, resulting in numerous chemical and physical factors affecting in the spectra [23,24]. This poses a challenge on the use of univariate regression (UVR) for content determination. Nevertheless, quantification can be achieved by using multivariate methods that take matrix effects into account [23]. One of the most common multivariate analysis techniques is partial least squares

regression (PLSR). Here, in contrast to UVR, the entire spectrum is taken in account for the analysis, reducing the susceptibility of the results to matrix effects [23].

In the present work, on the one hand, the use of LIBS in the exploration of copper deposits by means of a handheld spectrometer was investigated. On the other hand, a stationary LIBS was used for the analysis of copper-bearing ores. Only ground ore and rock samples from a VMS deposit and an SSC deposit were used for this study. For calibration, commercially available chalcopyrite and chalcocite minerals were ground up and blended with basalt and schist rock powders, respectively, at various concentrations. UVR was used to construct calibration graphs of the two studied minerals were obtained in both analyzed rocks. In addition to UVR, PLS regression was used to quantify copper content and PCA was used to investigate matrix effects in geologic samples.

2. Materials and Methods

2.1. Samples and reference analysis

A total of 158 samples with different copper contents were analyzed in this work whereof 69 are artificial samples (referred to here as synthetic) and 82 are field samples. One subset of the field samples was taken from the Apliki and Skouriotissa VMS deposit of Cyprus. The other subset was taken from the SSC deposit of Lubin in Poland. The accompanying matrix for a sample is either igneous rock basalt in the case of the VMS deposit or schists in the case of the SSC deposit. All samples were in the form of homogenized ground powders.

Reference analysis for the geological samples from Cyprus was performed by aqua regia digestion followed by ultratrace ICP-MS analysis and by $\text{LiBO}_2/\text{LiB}_4\text{O}_7$ digestion followed by ICP-OES analysis, respectively. Sample copper concentrations range from 41 - 10000 ppm. The samples from Poland were referenced by X-ray fluorescence (XRF). Sample copper concentrations range from 0.8 - 2.4 %. For the LIBS analyses, 3 g of ground powders were pressed into pellets (P 40, Herzog Maschinenfabrik, Osnabrück, Germany).

2.2. Geological overview of deposits

The sample subset from Poland, where the copper deposits belong to the sediment-hosted stratabound copper deposits originates from Lower Silesia, the central area of the Pre-Sudetic Monocline and the North-Sudetic Basin. They are part of the Zechstein copper schist formation. The copper-bearing minerals occur in separate lithological layers, accompanied by dark gray sandstones in the bottom layer, black schists in the middle and carbonate rocks in the top part. The most important copper minerals of this deposit are chalcocite (Cu_2S), chalcopyrite (CuFeS_2) and bornite (Cu_5FeS_4) [22,25].

The second sample subset, which originates from Cyprus, are sulfide deposits that are part of the volcanogenic massive sulfide deposits. The Cyprus type of this deposit belongs to the Troodos ophiolite complex, whose copper-bearing stockwork is surrounded by basaltic pillow lavas. Besides the dominant minerals chalcopyrite (CuFeS_2), pyrite (FeS_2) and sphalerite ($(\text{Zn,Fe})\text{S}$), secondary copper minerals such as chalcocite (Cu_2S), covellite (CuS) and digenite ($\text{Cu}_{1.8}\text{S}$) occur [26,27].

2.3. Preparation of synthetic samples

Powders of the copper-bearing minerals chalcocite (minerals from the New Cornelia Mine, Ajo, Arizona, USA) and chalcopyrite (minerals from Füsseberg near Bierdorf, Siegerland, Germany) were used for the preparation of the synthetic samples. The samples were prepared by standard addition. In this process, the copper-bearing

minerals were mixed with the respective accompanying matrix of basalt or schist. The used basalt originate from Cyprus and the used schist sample originates from east of the village of Lehesten in the southeastern Thuringian Forest. The schists are from the Saxothuringian as part of the Variscan Mountains. The formerly marine, clayey sediments were deposited in the Mississippian (Lower Carboniferous, ca. 360 Ma). The weighed samples were homogenized and pressed into pellets (TP40, Herzog Maschinenfabrik, Osnabrück, Germany). For one pellet a total of 3 g of sample material was used. Here, two pellets were pressed for each concentration of copper, one in basalt the other in schist. A total of 35 samples were prepared for each mineral, 15 in schist and 20 in basalt.

2.4. LIBS setup and measurement parameters

The samples were analyzed by two different LIBS spectrometers. One LIBS spectrometer is a stationary laboratory benchtop instrument equipped with an Echelle spectrometer (Aryelle Butterfly, LTB, Berlin, Germany). For the study, the samples were positioned on a rotation and linear translation stage for enabling a spiral probing pattern during the measurement. This ensured that each ablation was performed on a fresh surface. Plasma at the sample surface was generated by a focused ($f = 50$ mm) Nd:YAG laser (Bernoulli LIBS, Litron Lasers, Rugby, England, Great Britain, $\lambda = 1064$ nm, $E = 80$ mJ, repetition rate 10 Hz, pulse duration 7 ns). Using a concave mirror (ME-OPT-0007, Andor Technology, Belfast, Northern Ireland, focal length 52 mm, $\lambda = 200 - 1100$ nm), the emission was collected and focused onto an optical fiber that guides the light to the spectrometer. The spectrometer has two wavelength ranges (UV: 190 - 330 nm, VIS: 278 - 769 nm). Both ranges were measured separately with a resolution of 25 pm. An ICCD camera (iStar, AndorTechnology, Belfast, UK) was used for detection. A total of approximately 200 spectra were recorded per sample in the UV and VIS regions, with 10 spectra averaged per shot.

The second LIBS spectrometer is a handheld instrument (Z-300, SciAps, WBN, MA). The plasma is generated by the integrated laser ($\lambda = 1064$ nm, $E = 7.5$ mJ, repetition rate 10 Hz, pulse duration 5 ns). The spectrometer has a wide wavelength range (180 - 960 nm) with a resolution of 100 pm. In order to amplify the LIBS signal, the measurements are additionally purged with argon gas before the measurement. The measurements were performed using the GeoChemPro app. Each sample probed by a grid of 8 by 8 spots, resulting in 64 spectra for each sample. The sampled area was approximately 1 mm². Each pellet was measured three times to obtain a representative set of spectra of the sample.

2.5. Data pretreatment

All spectra were prepared for analysis by background correction, standard normal variate-normalization, and final averaging of the data. A top-hat filter was used for the background correction. The structuring element length was 20 data points. This corresponds to a filter width of about 0.26 mm. Following the background correction, a SNV normalization was performed. For this, the mean value of the entire spectrum is subtracted from the spectrum. The difference is divided by the standard deviation of the spectrum. The resulting data were finally averaged to obtain a representative spectrum for each sample.

Linear regression was performed using two specific copper lines (324.75 nm and 327.40 nm). The area of each was obtained by numerical integration with the trapezoidal method. The sum of both peak areas represents the total peak area that is plotted against the copper concentration.

The PLSR was performed with a 10-fold cross-validation. It was performed for the number of components with the lowest estimated prediction error.

Outliers were identified using the robust PCA (ROBPCA) method of Hubert *et al.* [28]. Robust principal component analysis (ROBPCA) is an alternative to classical principal component analysis (PCA), which is very sensitive to outliers due to the empirical covariance matrix of the data. Robust PCA is designed for fast and robust observation of high dimensional data. It uses projection pursuit techniques in the original data space and projects the observations into a subspace with lower dimensions. Here, robust covariance estimates are made. This allows the identification of outliers in a data set and the classification of the data into four categories. Regular observations and good leverage points, which are classified as belonging to the data set, and orthogonal outliers and bad leverage points, which are designated as outliers. In the algorithm of Hubert *et al.* a so-called outlier map is generated by plotting the orthogonal distance against the robust score distance. In the resulting four quadrants the different data points are grouped. All data points in quadrants I or II are orthogonal outliers or bad leverage points, and are subsequently removed from the original data set [28].

For the principal component analysis (PCA), all spectra were fed after prior background correction, normalization and averaging.

The procedure for the determination of alternative regression models was analogous. In addition, the outliers were identified using ROBPCA and eliminated before applying the regression models.

The preparation of the data, as well as the analyses were carried out with Matlab (version 2021a, MathWorks, Natick, MA, USA). Origin (OriginLab, Northampton, MA, USA) was used to display the results.

3. Results

In the present work, synthetic and field samples of the copper-bearing minerals chalcopyrite (CuFeS_2) and chalcocite (Cu_2S) were analyzed by LIBS. The analyses were performed on two different LIBS spectrometers. First, the samples were analyzed using a handheld device. It is conceivable that mobile LIBS spectrometers could be used in the exploration and prospection of new potential copper deposits. Second, the analysis was carried out with an Echelle spectrometer. The high-resolution spectrometer could be used, for example, in the process analysis of the incoming copper ores and concentrates in the copper production.

3.1. LIB Spectra

The investigated copper-bearing field samples are from the basalt-related Cyprus-type deposit and from the schist-related SSC deposit from Lubin. A comparison of the obtained LIB spectra (Figure 1) highlights the differences and similarities between the host rocks, basalt and schist. One aim also was to identify the spectral lines of the host rock interfering with the copper lines in univariate regression and to estimate possible matrix effects.

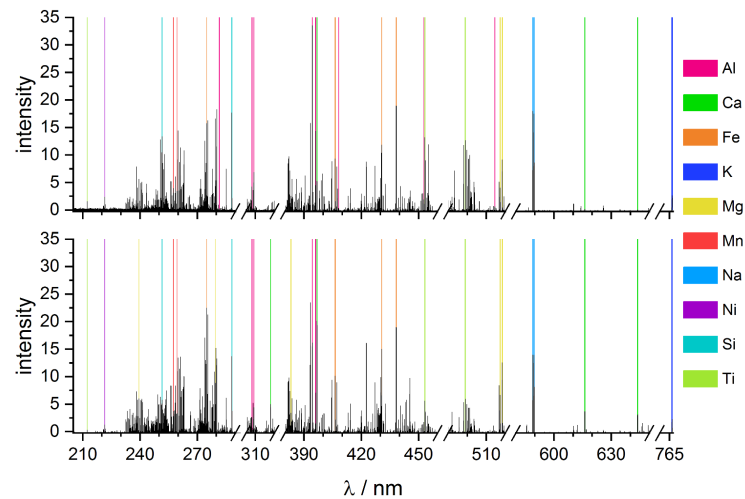


Figure 1. LIBS spectra measured with the Echelle spectrometer of the host rocks schist (top) and basalt (bottom) with marked lines of the relevant elements.

The line-rich LIBS spectra of the host rocks basalt (bottom, Figure 1) and schist (top, Figure 1) show a variety of elements present in both matrices. Differences between the host rocks are mainly in the contents of the individual elements. For example, the mafic minerals of basalt contain high amounts of magnesium, iron and calcium [29]. In contrast, the sedimentary rock of the SSC deposit is clay-bearing and chalky and contains mainly the elements aluminum, silicon and calcium [30,31]. In addition, similar contents of manganese are found in both matrices and titanium can be determined in schist. Similar results are also obtained from the reference analysis of the two host rocks (Table 1). Furthermore, nickel could be identified in basalt and schist.

Table 1. Average contents of the elements in the host rocks basalt and schist from the reference analyses.

average element content / %	Al	Ca	Fe	K	Mg	Mn	Na	Si	Ti
in schist	9.37	0.14	4.65	2.39	1.24	0.08	0.81	29.69	0.58
in basalt	7.07	7.71	2.47	0.77	5.56	0.09	0.90	22.34	0.23

In order to detect possible interferences by the surrounding matrix, the LIBS spectra of the field samples are compared with the spectra of the host rocks with respect to the copper lines (Figure 2).

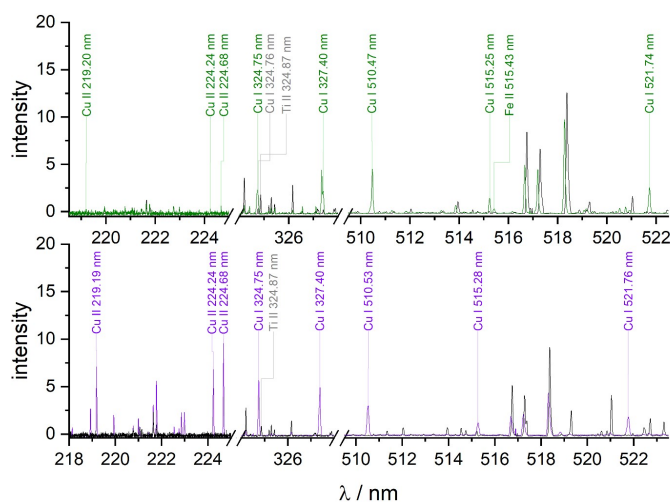


Figure 2. LIBS spectra of the field sample compared with the corresponding host rock. Top: Field sample from Cyprus compared with basalt; Bottom: Field sample from Poland compared with schist. Nine copper lines are identified in each of the field sample spectra. Lines in the rock spectra that are in the spectral vicinity of the copper lines are marked in grey. The measurements were performed with the Echelle spectrometer.

In schist, some iron lines are found in close proximity to the copper lines at 515.32 nm and 521.82 nm. Near to the Cu lines at 324.75 nm and 327.40 nm used in this work, titanium is found in schist and basalt at 324.87 nm. While the accompanying matrix basalt has traces of copper, no copper is found in the host rock schist.

However, since they do not show significant interfering lines, they are well suited as host for the copper minerals in the synthetic samples. The copper lines at 324.75 nm and 327.40 nm are also nearly undisturbed and can be used for univariate regression of the handheld and Echelle spectra.

3.2. Univariate Analysis

One of the most common methods to determine the element contents in different matrices is univariate regression (UVR), in which a calibration function is generated [32]. The calibration function describes a linear or nonlinear relationship between the element content and the area of the element line in the spectrum. Its quality is described by the coefficient of determination R^2 and the limit of detection (LOD). In the present case, the linear calibration function describes the correlation between the copper concentration and the peak area of the two copper lines at 324.75 nm and 327.40 nm in the spectrum, which could allow quantification of unknown copper concentrations. Other copper lines that could also be used for the analysis, e.g. at 521.82 nm [33], are not sensitive enough due to the low copper concentration in the field samples. This is also clear from the weighting of the spectral lines in the PLSR. In the case of the basalt-containing field samples from Cyprus, the lines at 521.82 nm are not detected. The coefficients of determination and the LOD of the calibration functions which are obtained from the Echelle and handheld spectra are summarized in Table 2.

Table 2. Coefficients of determination R^2 and LOD of the univariate regressions of the synthetic samples and the field samples in basalt and schist with the Echelle and handheld spectrometer

sample	basalt		schist		LIB spectrometer
	R^2	LOD / ppm	R^2	LOD / ppm	
Cu ₂ S	0.95	140	0.85	1000	Echelle
CuFeS ₂	0.98	110	0.55	1000	Echelle
Cypriot mines 1+2 (basalt)	0.73	110	-	-	Echelle
Polish mine	-	-	0.27	9000	Echelle
Cu ₂ S	0.98	1600	0.41	1000	handheld
CuFeS ₂	0.96	1100	0.39	1000	handheld
Cypriot mines 1+2 (basalt)	0.63	1000	-	-	handheld
Polish mine	-	-	0.31	9000	handheld

The LOD are roughly estimated values. With respect to the coefficient of determination three trends can be observed from the results:

1. The UVR obtained from the Echelle spectra have higher coefficients of determination than the UVR obtained from the handheld spectra.
2. The UVR of the samples in schist ($R^2 = 0.27 - 0.85$) have lower R^2 than the UVR of the samples containing basalt ($R^2 = 0.63 - 0.98$).
3. The synthetic samples ($R^2 > 0.95$ in basalt, $R^2 > 0.39$ in schist) have UVR with higher coefficients of determination than the corresponding field samples ($R^2 > 0.63$ in basalt, $R^2 > 0.27$ in schist).

The reasons for the lower linear correlation in the handheld spectra are probably the lower pulse energy, the lower spectrometer resolution and the purging with argon before and during the measurement. Compared to the Echelle spectrometer, where a pulse energy of 80 mJ was used, the handheld spectrometer has only about one tenth of this pulse energy (7.5 mJ). The low pulse energy results in a cooler plasma, which affects the overall line intensity. This change in line intensity can be partly compensated by using argon gas in the sample space. At the same time, the use of a purge gas is associated with uncertainties, since not all ambient air may be completely removed from the sample space. The lower resolution of the handheld spectrometer probably also contributes to the lower linear correlation. Compared to the Echelle spectrometer, which has a spectral resolution of about 25 pm, the handheld spectrometer has a resolution larger than 100 pm. The lower resolution may result in partially overlapping lines. As a result, the exact line position is not known, and line identification becomes difficult. Partial superposition can also lead to errors in the determination of peak areas and thus inaccuracies in the univariate regression, which are reflected in the lower coefficients of determination. The samples in schist show lower R^2 than those in basalt. The reason could be the material properties of both rocks. Although both rocks are fine-grained and relatively hard, schist has a lower density and is more brittle. This was also observable during measurement. The pellets with a schist matrix were significantly more frequently partially destroyed than the basalt pellets. This destruction of the pellets primarily results in a more uneven surface and a less spatial correspondence between laser focus

and surface. This may influence the intensities of the spectra, resulting in lower univariate regression curves.

The univariate regression of the field samples yields lower R^2 than the synthetic samples. The field samples differ from the synthetic samples in copper content. In addition, they are natural samples that can be more affected by matrix effects because they are exposed to weathering and contamination. This makes the elemental composition of the surrounding matrix more heterogeneous. Interfering spectral lines can occur, which in turn can lead to partial superimpositions and interference of the relevant lines. These result in correlations with lower linearity and thus lower coefficients of determination. Good correlations can only be obtained for the synthetic samples in basalt with the Echelle and handheld spectrometer. For the field samples and especially the samples in schist, the univariate regressions are not suitable as effective calibration functions due to matrix effects. For better quantification, the use of multivariate methods is necessary.

3.3. Multivariate Analysis – Partial least squares regression (PLSR)

As multivariate regression method partial least squares regression (PLSR) was applied. The regression model was validated twice. First, the stratified data set was separated into a training and a validation data set (70:30). The regression model was developed on the basis of the training data and was tested with the validation data. Second, a 10-fold cross-validation of the training data was performed during model development. The coefficient of determination R^2 and the root mean square error (RMSE) of the validation (R^2_v , RMSEV) as well as the cross-validation (R^2_{cv} , RMSECV) were used as figure of merit. Figure 3 shows an example of the correlation plot between observed and fitted response of the PLSR of Cu_2S in basalt from the Echelle spectrometer.

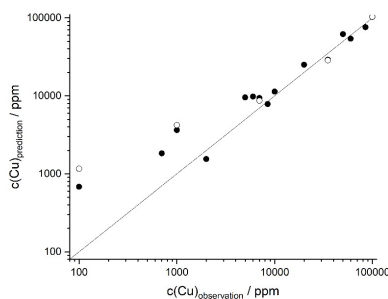


Figure 3. PLSR of CuFeS_2 data in basalt at Echelle spectrometer. The logarithms of the copper concentrations and 2 components were used for the regression model. Black dots represent calibration model ($R^2_{cv} = 0.97$, RMSECV = 5016 ppm), circles represent validation ($R^2_v = 0.95$, RMSEV = 5012 ppm).

The black dots are the result of the PLS calibration ($R^2_{cv} > 0.97$), while the circles represent the validation of the model ($R^2_v > 0.95$). Although the validation has a fairly high goodness of fit, the plot shows potential outliers that could negatively affect the correlation. To identify these potential outliers, the robust principal components analysis (ROBPCA) method of Hubert *et al.* is applied.[28]. As a result, an outlier map (Figure 4) is obtained, which is plotted between score distance and orthogonal distance.

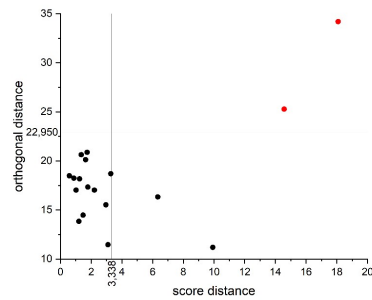


Figure 4. ROBPCA outlier map of CuFeS_2 data in basalt at Echelle spectrometer with 4 components. The robust score distance at 3.338 and the orthogonal distance at 22.950 span the quadrants of the map. Two bad leverages (red) are detected in the first quadrant. Orthogonal outliers are not included in the data set (2nd quadrant).

The first and second quadrants contain the outliers, the so-called bad leverage points and the orthogonal outliers. They are removed from the original data. The subsequent PLSR (Figure 5) gives better results in terms of linearity ($R^2_{CV} > 0.98$) of the calibration model and validation ($R^2_V > 0.96$).

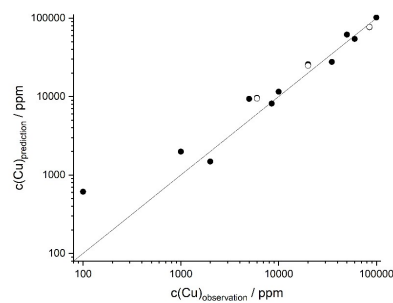


Figure 5. PLSR of CuFeS_2 data in basalt at Echelle spectrometer with 2 components after elimination of outliers, black dots represent calibration model ($R^2_{CV} = 0.98$, $\text{RMSECV} = 4880$ ppm), circles represent validation ($R^2_V = 0.96$, $\text{RMSEV} = 4586$ ppm).

This procedure was applied to all the series of samples studied on the handheld and Echelle spectrometer. The results are summarized in Table 3. In almost all cases, the elimination of outliers results in a better regression model. The PLSR results of the data measured on the handheld spectrometer show coefficients of determination of $R^2_{CV} > 0.55$. Exceptions are the sample series of the Cypriot samples ($R^2_{CV} > 0.35$) and Cu_2S in schist ($R^2_{CV} > 0.22$). In contrast to the PLSR results of the handheld spectrometer, the PLS regressions of the samples analyzed by the Echelle spectrometer have higher coefficients of determination ($R^2_{CV} > 0.85$). An exception are again the Cypriot samples with a lower $R^2_{CV} > 0.49$.

Table 3. PLSR results of the examined samples in basalt and schist at Handheld and Echelle spectrometer with 4 components after elimination of outliers by the robust principal components analysis (ROBPCA) method

sample	number of components	cross-validation		validation	
		R ² _{CV}	RMSECV	R ² _V	RMSEV
handheld spectrometer					
CuFeS ₂ in basalt	5	0.74	1.2358	0.61	1.2902
Cu ₂ S in basalt	7	0.96	0.4915	0.92	0.6313
CuFeS ₂ in schist	5	0.68	1.2229	0.61	1.3138
Cu ₂ S in schist	3	0.58	1.2960	0.60	1.2498
Cypriot mines 1+2 (basalt)	6	0.31	1.0249	0.25	1.0303
Polish mine (schist)	4	0.55	2922.8	0.45	3140.6
Echelle spectrometer					
CuFeS ₂ in basalt	4	0.90	0.7628	0.87	0.8154
Cu ₂ S in basalt	5	0.98	0.3454	0.97	0.3449
CuFeS ₂ in schist	2	0.87	0.8256	0.86	0.7759
Cu ₂ S in schist	2	0.79	0.9538	0.73	1.0026
Cypriot mines 1+2 (basalt)	2	0.52	0.8397	0.36	0.9420
Polish mine (schist)	5	0.85	1897.9	0.80	2066.4

In both cases, Echelle and handheld spectrometers, better regression models are obtained for the synthetic samples than for the field samples, except for the regression of chalcocite in schist. As already mentioned, the synthetic samples have a much smaller variation in the matrix composition than the field samples. The latter ones are inhomogeneous, due to grain size variations and are exposed to natural processes, such as chemical and physical weathering. Due to climate, ground cover, rock properties and duration of exposure, the surrounding matrices weather very differently [34]. This results in matrices that have a more complex chemical composition, which is dependent on the site of sampling. Rock powders from the same rock were added to the synthetic samples, so that hardly any differences in the matrix can be found. These differences of the rock matrices influence the regressions.

Compared to the UVR, multivariate PLS regression improves the regression models, especially for the synthetic and the field samples in schist measured by the Echelle and handheld spectrometer. For the synthetic samples in basalt, similar regression results are obtained. The UVR of the Cypriot field samples is better than the PLSR.

3.3.1. Wavelength dependence

While UVR only considers two spectral lines of the entire spectrum, namely the copper lines at 324.75 nm and 327.40 nm, PLSR includes all spectral lines in the analysis. The PLS weights in Figure 6 show how strong the individual spectral lines of the LIB spectra affect the PLS regression. In this figure, the PLS weights of the first two principal components as a function of the wavelength are displayed for the field samples from Cyprus (top) and Poland (bottom).

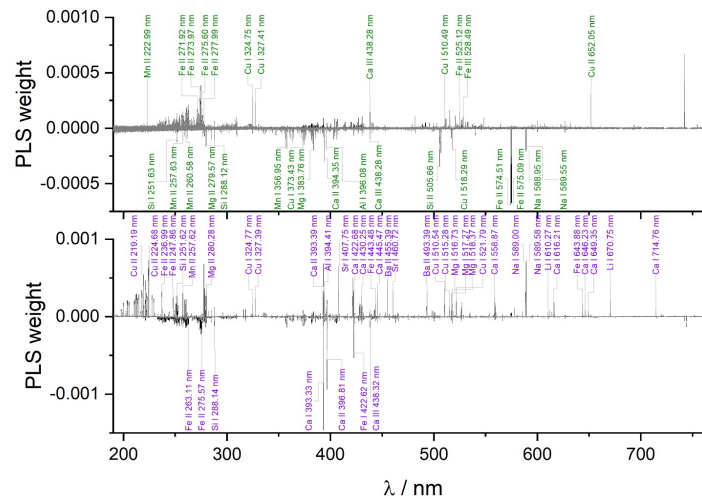


Figure 6. Weighting of PLSR as a function of wavelength for PLSR of field samples from Cyprus (top, green) and Poland (bottom, purple) with 2 components each (first component black, second component grey).

In both cases, numerous spectral lines have positive weights. This also includes the two copper lines at 324.75 nm and 327.40 nm, which were used for the univariate regression. In addition to other copper lines of the Cypriot and Polish samples, quite a few lines of the elements of the host rocks basalt and schist can be identified. In the upper spectrum of Figure 6, mainly copper lines and iron lines are found with positive weights. In particular, the copper lines originate from the copper-bearing minerals chalcopyrite and chalcocite. The iron lines may result from the mineral chalcopyrite (CuFeS_2) and the matrix basalt (Cf. Table 1). The negative weights originate from the accompanying matrix basalt. These include element lines from silicon, magnesium, manganese, aluminum, calcium and sodium. A different picture emerges when looking at the weighting of the wavelengths for the Polish samples (Figure 6 below). In addition to numerous copper lines, many lines with positive weights can be identified that are part of the schist matrix. As in the case of the Cypriot samples, the Cu lines originate from the minerals chalcopyrite and chalcocite. In contrast to the Cypriot samples, significantly more Cu lines can be found here, because the copper contents of the Polish samples are significantly higher than those of the samples from Cyprus. In addition to the copper lines, lines of Ca, Mg, as well as Ba, Na and Li have positive weights. These are part of the schist matrix. Mainly iron lines, but also Ca lines have negative weights. While calcium is part of the accompanying matrix schist, iron lines may originate from the copper mineral chalcopyrite and the matrix.

3.4. Principal component analysis (PCA)

One way to investigate the matrix influence is clustering by the principal component analysis (PCA). For this purpose, all samples examined with LIBS are characterized by PCA. The score plot shows the first two principal components that explain most of the variance. In the score plot, closely grouped data points indicate low

variance in the chemical composition of the matrix, whereas a wide distribution of data points visualizes significant differences in the composition of the matrix.

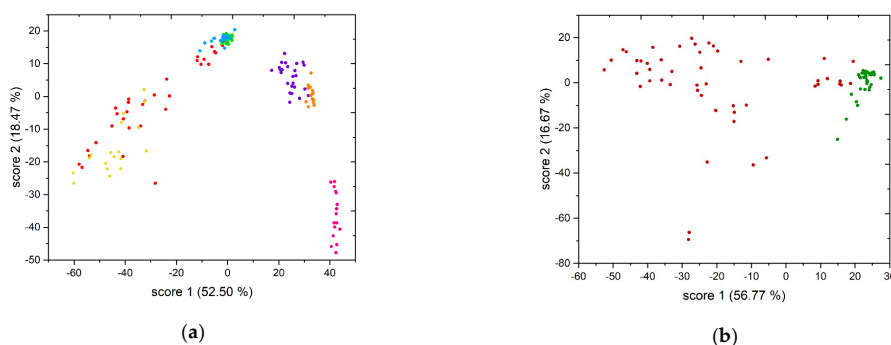


Figure 7. (a) Score plot of PCA of all samples analyzed with LIBS on the handheld spectrometer (red - Cypriot mine 1, yellow - Cypriot mine 2, green - Cu_2S in basalt, blue - CuFeS_2 in basalt, purple - Polish mine, orange - Cu_2S in schist, pink - CuFeS_2 in schist). (b) Score plot of PCA of Cyprian samples (dark red) and samples containing Cu_2S and CuFeS_2 in basalt (dark green).

Figure 7a shows the score plot of the handheld spectrometer. The first two principal components explain about 71 % of the variance of the analyzed samples. The samples containing basalt (Cyprian mine 1 (red), Cyprian mine 2 (yellow), Cu_2S (green) and CuFeS_2 (blue) in basalt) are spatially well separated from the samples containing schist (Polish mine (purple), Cu_2S (orange) and CuFeS_2 (pink) in schist). The samples form clusters which are well separated depending on the deposit (field samples) or host rock (synthetic samples) but the clusters of the two Cyprus deposits are poorly separated from each other. Synthetic samples such as the Cu_2S -bearing samples in basalt (green) and schist (orange) as well as the chalcopyrite-bearing samples in basalt (blue) are closely grouped. This indicates small differences in the chemical composition of the matrix and explains their excellent regression fits (Table 3). The samples from the Polish mine (purple) and the synthetic CuFeS_2 samples in schist (pink) represent a slightly larger distribution. The matrices of these samples have larger variances in their chemical composition. This results in slightly weaker correlations in the PLS regression. However, the samples of Cyprus mines 1 (red) and 2 (yellow) have the widest distribution in the score plot, indicating larger differences in the chemical composition of the matrices. This also results in a weak correlation of the PLS regression (Table 3). The reason for the large variations in chemical composition in basalt is probably chemical and physical weathering. Depending on the collection site, the field samples were exposed to atmospheric conditions to varying degrees, which led to the differences in composition. These differences were visible on the specimens macroscopically, especially by color differences, but also inclusions and different grain sizes. They are clearly illustrated by the score plot in Figure 7b. Here, the first two principal components from the Cypriot samples (dark red) and the corresponding synthetic samples in basalt (dark green) are shown. In addition to the two clusters that formed, the wide distribution of data points in the cluster of Cypriot field samples is striking. In contrast, the data points of the samples containing Cu_2S and CuFeS_2 in basalt have a rather narrow distribution, illustrating a high similarity of chemical composition, which in turn can be found in the PLS regression, indicated by a rather high coefficient of determination ($R^2_{\text{CV}} > 0.74$, Table 3).

The PCA of the samples analyzed by the Echelle spectrometer yields a better cluster separation of the different sources (Figure 8). Similar to the score plot of the handheld spectrometer, the basalt containing samples (Cyprian mine 1 (red), Cyprian mine 2 (yellow), Cu_2S (green) and CuFeS_2 (blue) in basalt) are spatially separated from the schist containing samples (Polish mine (purple), Cu_2S (orange) and CuFeS_2 (pink) in schist). The clusters are more clearly separated from each other than in Figure 7a, which is also reflected in the coefficients of determination ($R^2_{\text{CV}} > 0.79$) of the PLS regressions (Table 3). At the same time, the data points are more widely separated in the clusters. The regressions of the samples analyzed on the Echelle spectrometer have higher coefficients of determination than those of the samples measured on the handheld spectrometer. The score plot of the samples measured on the Echelle spectrometer mainly shows better separated clusters and more evenly distributed data points.

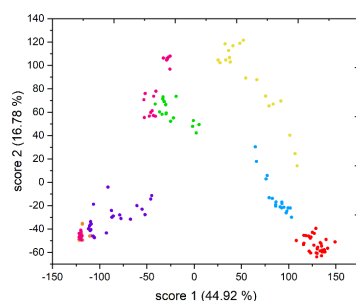


Figure 8. Score plot of PCA of all samples analyzed by LIBS on Echelle spectrometer (red - Cypriot mine 1, yellow - Cypriot mine 2, green - Cu_2S in basalt, blue - CuFeS_2 in basalt, purple - Polish mine, orange - Cu_2S in schist, pink - CuFeS_2 in schist).

3.5. Alternative regression models

The PLS regression has become a standard tool of chemometrics in chemistry and engineering [35]. In this paragraph, other regression methods are also investigated. For this purpose, the samples were examined by three additional regression methods and different kernel functions. Besides the linear methods principal component regression (PCR) and linear support vector machine (SVM) regression, nonlinear methods such as the quadratic and the cubic SVM, as well as Gaussian process regression (GPR) were examined. As kernel functions, the rational quadratic GPR, exponential GPR and matern 5/2 GPR were used. Using the field samples from Poland measured on the Echelle spectrometer as an example, the results are shown in Table 4.

Table 4. Different regression models with coefficients of determination (R^2) and root mean square error (RMSE) of the regression and validation using the field samples from Poland measured on the Echelle spectrometer after the elimination of outliers by the ROBPCA method.

model	cross-validation		validation	
	R^2_{CV}	RMSECV	R^2_{v}	RMSEV
PLS regression	0.85	1897.9	0.80	2066.4
PCR	0.73	2636.0	0.69	2943.1
linear SVM	0.71	2744.8	0.65	3161.1
quadratic SVM	0.67	2920.8	0.68	3023.5

cubic SVM	0.62	3125.1	0.74	2714.9
rational quadratic GPR	0.73	2646.3	0.59	3395.1
exponential GPR	0.76	2469.4	0.58	3420.7
matern 5/2 GPR	0.74	2564.8	0.59	3387.1

Table 4 shows that there are models which are able to describe the investigated data well, e.g. the linear regression ($R^2_{CV} = 0.73$, $R^2_V = 0.69$), nevertheless none of the shown models is able to describe the data as accurately as the PLS regression ($R^2_{CV} = 0.85$, $R^2_V = 0.80$). All other models achieve lower coefficients of determination. This is the reason for applying PLS regression as the method of choice in this work.

3.6. Validation of the PLSR calibration based on the field samples

In the exploration of copper deposits many samples have to be measured and only a few samples can be drawn for a later investigation by laboratory-based methods (e.g. ICP-OES). This limited sample volume often results from limited transport capacity and the high cost of reference analysis. The prerequisite for on-site measurements in the field is a reliable calibration model. One way of creating this is based on the standard addition method, in which various amounts of the copper mineral are added to a representative host rock. However, since the host rock often varies greatly in its chemical composition in the field, strong deviations in the LIBS signal are to be expected. This is reflected in the PCA score plot, where the clusters of synthetic and real samples are clearly separated and differ in their area. One way of adjusting the calibration function to the field conditions can be the addition of some real samples, which have been characterized (labeled) by reference analysis, to the calibration model. A way to validate the calibration is to apply the obtained calibration model to the field samples. Besides validation, this also shows if and how comprehensive the calibration can be. Additionally, an idea of the predictions on the basis on the PLSR calibration models for other samples with similar associated rocks is gained.

The PLSR calibration models of the synthetic samples of Cu_2S and $CuFeS_2$ in basalt and schist, respectively, were applied to the Cypriot and Polish field samples (all measured on the Echelle and handheld spectrometers, respectively) for this purpose.

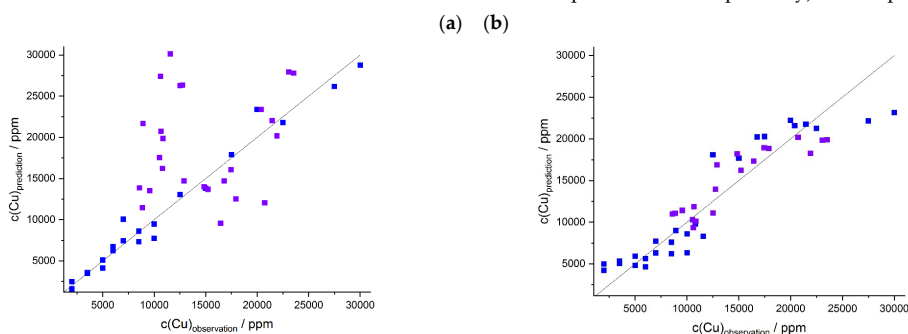


Figure 9. (a) PLSR (4 components) of the synthetic samples (Cu_2S and $CuFeS_2$ in schist; dark blue) applied to the Polish field samples (purple): $R^2_V = -1.95$, RMSEV = 8228 ppm; (b) PLSR (4 components) of the synthetic samples (Cu_2S and $CuFeS_2$ in schist) and 22 % of the field samples (dark blue) applied to the Polish field samples (purple): $R^2_V = 0.79$, RMSEV= 2180 ppm; all measured on Echelle spectrometer.

Figure 9a shows the result of the PLSR calibration of the synthetic samples of Cu_2S and CuFeS_2 in schist (dark blue) and the validation with the Polish field samples (purple). The validation only achieves a coefficient of determination of $R^2 = -1.95$. Thus, the PLSR calibration of the synthetic samples in schist is not suitable to quantitatively predict element contents of the Polish field samples. To increase the prediction accuracy, data from the field samples were successively added to the synthetic samples and the calibration and validation were performed again. The best result was obtained after adding 22 % randomly chosen field samples to the calibration model of the synthetic samples (Figure 9b). A coefficient of determination of the validation of $R^2 = 0.79$ could be obtained (Table 5).

Table 5. Validation results (R^2 , RMSE) of PLSR of the synthetic samples in schist validated with the Polish field samples

samples		spectrometer	components	validation	
training data	validation data			R^2_v	RMSEV
syn. samples in schist incl. ~22 % polish field samples	Polish field samples	Echelle	4	0.79	2180.3
syn. samples in schist incl. ~40.5 % polish field samples	Polish field samples	handheld	4	0.42	3560.4

Thus, this calibration method is suitable to quantitatively predict element contents in unknown field samples. The corresponding investigations of the field samples obtained by the handheld instrument and from the Cypriot field samples are not successful. This could be expected because the corresponding figures of merit of regression (Table 3) are poor but in principle similar relative trends are obtained.

In contrast to the Polish samples ($c(\text{Cu}) = 0.8 - 2.4 \%$), the Cypriot samples ($c(\text{Cu}) = 41 - 10000 \text{ ppm}$) possess a much lower concentration of copper. This seems to be the reason for the weaker correlation of the data. The necessary addition of the field samples to the training data shows that the matrix effects of the field samples influence the calibration. Since the complex chemical composition of the field samples cannot be represented in the synthetic samples, the addition of some field samples to the calibration data is a reasonable way.

Figure 10 shows the evolution of the R^2_v of the validation data with increasing ratio of field samples to the synthetic samples up to a ratio of about 50 %.

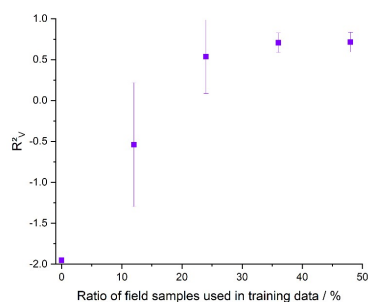


Figure 10. Evolution of the R^2_v of the validation data with increasing proportion of randomly chosen field samples in the training data for the Polish samples measured on the Echelle (purple). This process was performed five times.

If this ratio increases, the coefficients of determination also increase and converge to a plateau at almost 22 % of the field samples added to the synthetic data. In the case of the data measured on the Echelle spectrometer, the coefficient of determination reaches a value of $R^2 = 0.79$ at an admixture of 22 % and of $R^2 = 0.81$ at an admixture of 50 %. Especially with regard to the application during prospecting and exploration work, the use of a smaller field sample data set is reasonable, since a smaller part of field samples has to be taken and analyzed in the laboratory. This can save time, resources and costs and still achieve good results.

4. Summary and Conclusion

LIBS is a promising method for the detection and quantification of copper in copper-bearing samples, as this technique can rapidly determine numerous elements simultaneously in a matrix with little or no sample preparation. LIBS is compromised by its strong matrix dependence. This requires careful calibration and the use of multivariate methods. The handheld instrument used in this work is particularly suited for use in the field. In addition to the mobility offered by this handheld instrument, the simplicity of use, purging with argon, and measurements in a grid are advantageous for field studies. In contrast, the stationary LIBS system with an Echelle-based detector offers a higher spectral resolution and a larger excitation energy, which also allow a more accurate view of the samples in terms of multivariate analysis of the spectra.

In the present work, two data evaluation methods are used to quantify the copper contents in the studied ores. The first approach is a calibration using synthetic samples in conjunction with univariate evaluation. Here, the peak areas of the copper peaks at 324.75 nm and 327.40 nm were used for the analysis. However, due to strong matrix effects, the obtained univariate regressions are in the most cases not suitable as calibration functions. Exceptions are the regressions of the Cypriot field samples measured with Echelle and handheld spectrometer. In a second approach, the multivariate method of PLS regression, after an elimination of outliers using ROBPCA, is used. This resulted in better calibration models. PCA was performed to estimate the matrix effects. Applying alternative regression models to the LIBS data did not improve predictions compared to PLSR. Finally, the predictive power of the PLSR was tested. For this purpose, the regressions of the synthetic samples were applied to the field samples. It turns out to be advantageous if the training data set already contains some labeled field samples. The use of a mobile LIBS spectrometer that can be deployed in the field can save on sampling, transportation and analysis costs when prospecting and exploring

for new copper deposits. At the same time, the use of a high-resolution LIBS spectrometer allows more accurate determination and monitoring of copper grades, e.g., during the refining and production process for high-purity copper.

In the future, this work should be supplemented by field samples from additional copper deposits, especially with regard to other host rocks. In addition, the samples could be expanded to include copper concentrates of different processing stages.

Authors Contributions: P.B., T.B., U.A. and H.-G.L. made the conceptualization. Data curation, formal analysis and visualization was performed by P.B. Funding acquisition and project administration carried out T.B. and H.-G.L. The resources were provided by N.K., U.A. and S.M. Methodology was performed by P.B. and T.B. Supervision was done by T.B., U.A. and H.-G.L. P.B. and T.B. wrote the original draft. S.M., N.K., U.A. and H.-G.L. have reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support of this research work within the framework of the State of Brandenburg (ILB) in the LIBSqOre project (No. 80172489).

Data Availability Statement: The field samples from Cyprus are part of a dissertation at the Institute of Geosciences at the University of Potsdam (Koerting 2021). All reference data are available on request (beitz@uni-potsdam.de).

Acknowledgments: The authors would like to thank David Mory and Katarzyna Cieslik for supporting during the project and Friederike Körting for providing the Cyprus field samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

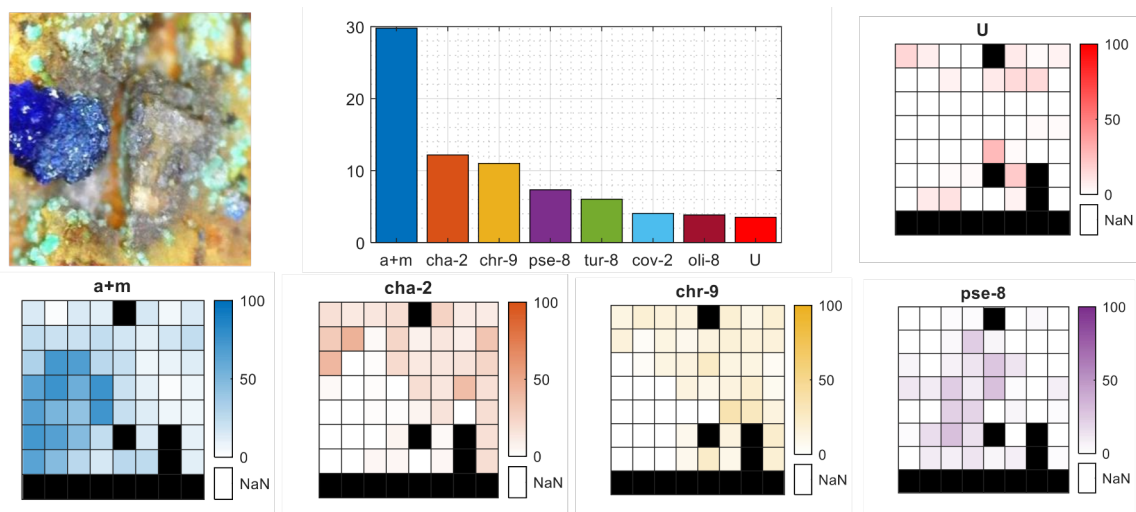
1. Umwelt Bundesamt Nichteisenmetallindustrie. Available online: <https://www.umweltbundesamt.de/themen/wirtschaft-konsum/industrieverarbeitendevon-metallen/nichteisenmetallindustrie#die-nichteisenmetallindustrie-in-deutschland> (accessed on 9 November 2021).
2. Marscheider-Weidemann, F.; Langkau, S.; Eberling, E.; Erdmann, L.; Haendel, M.; Krail, M.; Loibl, A.; Neef, C.; Neuwirth, M.; Rostek, L.; et al. Rohstoffe für Zukunftstechnologien 2021. - DERA Rohstoffinformationen 50: 366 S., Berlin. Available online: https://www.deutsche-rohstoffagentur.de/DE/Gemeinsames/Produkte/Downloads/DERA_Rohstoffinformationen/rohstoffinformationen-50.pdf?__blob=publicationFile&v=4 (accessed on 9 November 2021).
3. Neukirchen, F.; Ries, G. *Die Welt der Rohstoffe*; 2. Edition.; Springer Spektrum: Berlin, Heidelberg, 2016; ISBN 978-3-642-37738-9.
4. Gocht, W. *Wirtschaftsgeologie und Rohstoffpolitik*; 2.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1983; ISBN 978-3-540-12588-4.
5. Deutsches Kupferinstitut Berufsverband e.V. Produktionsprozesse. Available online: <https://www.kupferinstitut.de/kupferwerkstoffe/kupfer/produktionsprozesse/> (accessed on 9 November 2021).
6. Knapp, H.; Neubert, K.; Schropp, C.; Wotruba, H. Viable Applications of Sensor-Based Sorting for the Processing of Mineral Resources. *ChemBioEng Rev.* **2014**, *1*, 86–95, doi:10.1002/cben.201400011.

7. Nadolski, S.; Samuels, M.; Klein, B.; Hart, C.J.R. Evaluation of Bulk and Particle Sensor-Based Sorting Systems for the New Afton Block Caving Operation. *Miner. Eng.* **2018**, *121*, 169–179, doi:10.1016/j.mineng.2018.02.004.
8. Pořízka, P. Using Laser-Induced Breakdown Spectroscopy (LIBS) for Material Analysis, Brno University of Technology, Brno, Czech Republic, 2014. Available online: https://theses.cz/id/zekxmz/Porizka_disertace.pdf
9. Ostermann, M.; Schmid, T.; Büchele, D.; Rühlmann, M. In Echtzeit quer durch's Periodensystem. *LaborPraxis* **2016**, 1–7. Available online: <https://www.laborpraxis.vogel.de/in-echtzeit-quer-durchs-periodensystem-a-564401/> (accessed on 9 November 2021).
10. Wirtschaftsvereinigung Metalle Die NE-Metalle. Available online: <https://www.wvmetalle.de/die-ne-metalle/> (accessed on 9 November 2021).
11. Schorn, S. et al Mineralienportrait Kupfer – Eigenschaften. Available online: <https://www.mineralienatlas.de/lexikon/index.php/Mineralienportrait/Kupfer/Eigenschaften> (accessed on 9 November 2021).
12. Sievers, H. Der Einfluss von Lagerstätteneigenschaften auf eine nachhaltige Rohstoffnutzung am Beispiel Kupfer. Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen, Germany, 2005. ISBN 393253722X. Available online: <https://d-nb.info/976934922/34>.
13. Dorner, U. Rohstoffrisikobewertung – Kupfer. - DERA Rohstoffinformationen 45: 58 S.; Berlin; 2020. Available online: https://www.bgr.bund.de/DERA/DE/Aktuelles/rohstoff_kupfer.html.
14. British Geological Survey. *Copper*. Nottingham, UK, 2007. Available online: https://www2.bgs.ac.uk/mineralsuk/download/mineralProfiles/copper_profile.pdf (accessed on 9 November 2021).
15. Sicius, H.; Reise, E. *Kupfergruppe: Elemente der ersten Nebengruppe*; Essentials; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2017; ISBN 9783658172046. DOI 10.1007/978-3-658-17205-3.
16. Uhlendorff, A. Kupfergewinnung und -Produktion. Available online: <https://institut-seltene-erden.de/kupfergewinnung-und-produktion/> (accessed on 9 November 2021).
17. Angerer, G.; Mohring, A.; Marscheider-Weidemann, F.; Wietschel, M. Kupfer für Zukunftstechnologien: 52 S.; Karlsruhe, Germany, 2010; Available online: https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cce/2010/Bericht_FSEM_Cu.pdf.
18. Neukirchen, F. *Von der Kupfersteinzeit zu den Seltenen Erden*; Springer Spektrum: Berlin, Heidelberg, 2016; ISBN 9783662493465. DOI 10.1007/978-3-662-49347-2.
19. Quaiser, J. Internationales Forschungsprojekt ReAK zur Reduktion des Arsengehalts in Kupferkonzentraten 2020. Available online: <https://www.envirochemie.com/ecomaXL/files/20200527-ReAK-Reduktion-Arsengehalts-Kupferkonzentraten.pdf?download=true>. (accessed on 2 March 2022).
20. BGS Press Insights into the 'Missing Link' of How Copper Ore Deposits Form. Available online: <https://www.bgs.ac.uk/news/insights-into-the-missing-link-of-how-copper-ore-deposits-form/> (accessed on 14 September 2021).

21. Siebert, M. Neuzeitliche Minen auf Zypern. Available online: https://homersheimat.de/themen/minen-auf-zypern/minen_zypern.php (accessed on 2 March 2022).
22. Szamalek, K.; Szuflicki, M.; Malon, A.; Tymiński, M. *Mineral Resources of Poland*; Warsaw, Poland, 2017. ISBN 978-83-7863-718-9.
23. Lepore, K.H.; Fassett, C.I.; Breves, E.A.; Byrne, S.; Giguere, S.; Boucher, T.; Rhodes, J.M.; Vollinger, M.; Anderson, C.H.; Murray, R.W.; et al. Matrix Effects in Quantitative Analysis of Laser-Induced Breakdown Spectroscopy (LIBS) of Rock Powders Doped with Cr, Mn, Ni, Zn, and Co. *Appl. Spectrosc.* **2017**, *71*, 600–626, doi:10.1177/0003702816685095.
24. Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.-G.; Gebbers, R. Soil Nutrient Detection for Precision Agriculture Using Handheld Laser-Induced Breakdown Spectroscopy (LIBS) and Multivariate Regression Methods (PLSR, Lasso and GPR). *Sensors* **2020**, *20*, 418, doi:10.3390/s20020418.
25. Oszczepalski, S.; Speczik, S.; Zieliński, K.; Chmielewski, A. The Kupferschiefer Deposits and Prospects in SW Poland: Past, Present and Future. *Minerals* **2019**, *9*, doi:10.3390/min9100592.
26. Malpas, J.; Langdon, G. Petrology of the Upper Pillow Lava Suite, Troodos Ophiolite, Cyprus. *Geol. Soc. Spec. Publ.* **1984**, *13*, 155–167, doi:10.1144/GSL.SP.1984.013.01.14.
27. Koerting, F. Applications for Virtual Mine Face Geology. **2021**, doi:10.25932/publishup-49909.
28. Hubert, M.; Rousseeuw, P.J.; Vanden Branden, K. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* **2005**, *47*, 64–79, doi:10.1198/004017004000000563.
29. Schorn, S. et al Basalt. Available online: <https://www.mineralienatlas.de/lexikon/index.php/RockData?rock=Basalt> (accessed on 10 December 2021).
30. Schorn, S. et al Schiefer. Available online: <https://www.mineralienatlas.de/lexikon/index.php/RockData?rock=Schiefer> (accessed on 10 December 2021).
31. Schorn, S. et al Schiefer. Kupferschiefer. Available online: [https://www.mineralienatlas.de/lexikon/index.php/Geologisches Portrait/Schiefer#Kupferschiefer](https://www.mineralienatlas.de/lexikon/index.php/Geologisches%20Portrait/Schiefer#Kupferschiefer) (accessed on 10 December 2021).
32. Guo, G.; Niu, G.; Shi, Q.; Lin, Q.; Tian, D.; Duan, Y. Multi-Element Quantitative Analysis of Soils by Laser Induced Breakdown Spectroscopy (LIBS) Coupled with Univariate and Multivariate Regression Methods. *Anal. Methods* **2019**, *11*, 3006–3013, doi:10.1039/c9ay00890j.
33. Labutin, T.A.; Popov, A.M.; Zaytsev, S.M.; Cal'ko, I.A.; Zorov, N.B. Determination of Copper Content in Soils and Ores by Laser-Induced Breakdown Spectrometry. *Opt. Spectrosc.* **2016**, *121*, 339–342, doi:10.1134/S0030400X16080129.
34. Grotzinger, J.; Jordan, T. *Press/Siever Allgemeine Geologie*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2017; ISBN 978-3-662-48341-1. DOI 10.1007/978-3-662-48342-8.
35. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130, doi:10.1016/S0169-7439(01)00155-1.

3.5 Klassifizierung kupferhaltiger Minerale mittels laserinduzierter Breakdownspektroskopie und multivariater Datenanalyse (*Manuskript 4*)

„Classification of Copper Minerals by Handheld Laser-Induced Breakdown Spectroscopy and Nonnegative Tensor Factorisation“



Sensors **20** (18), 5152 (2020)

<https://doi.org/10.3390/s20185152>



Article

Classification of Copper Minerals by Handheld Laser-Induced Breakdown Spectroscopy and Nonnegative Tensor Factorisation

Michał Wójcik ¹, Pia Brinkmann ², Rafał Zdunek ^{1,*}, Daniel Riebe ², Toralf Beitz ², Sven Merk ³, Katarzyna Cieślak ³, David Mory ³ and Arkadiusz Antończak ¹

¹ Department of Field Theory, Electronic Circuits and Optoelectronics, Faculty of Electronics, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50370 Wrocław, Poland; michal.wojcik@pwr.edu.pl (M.W.); arkadiusz.antonczak@pwr.edu.pl (A.A.)

² Physical Chemistry, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; pbrinkma@uni-potsdam.de (P.B.); riebe@uni-potsdam.de (D.R.); beitz@uni-potsdam.de (T.B.)

³ LTB Lasertechnik Berlin GmbH, Am Studio 2c, 12489 Berlin, Germany; sven.merk@ltb-berlin.de (S.M.); katarzyna.cieslik@ltb-berlin.de (K.C.); david.mory@ltb-berlin.de (D.M.)

* Correspondence: rafal.zdunek@pwr.edu.pl

Received: 2 August 2020; Accepted: 7 September 2020; Published: 9 September 2020



Abstract: Laser-induced breakdown spectroscopy (LIBS) analysers are becoming increasingly common for material classification purposes. However, to achieve good classification accuracy, mostly noncompact units are used based on their stability and reproducibility. In addition, computational algorithms that require significant hardware resources are commonly applied. For performing measurement campaigns in hard-to-access environments, such as mining sites, there is a need for compact, portable, or even handheld devices capable of reaching high measurement accuracy. The optics and hardware of small (i.e., handheld) devices are limited by space and power consumption and require a compromise of the achievable spectral quality. As long as the size of such a device is a major constraint, the software is the primary field for improvement. In this study, we propose a novel combination of handheld LIBS with non-negative tensor factorisation to investigate its classification capabilities of copper minerals. The proposed approach is based on the extraction of source spectra for each mineral (with the use of tensor methods) and their labelling based on the percentage contribution within the dataset. These latent spectra are then used in a regression model for validation purposes. The application of such an approach leads to an increase in the classification score by approximately 5% compared to that obtained using commonly used classifiers such as support vector machines, linear discriminant analysis, and the *k*-nearest neighbours algorithm.

Keywords: LIBS; NTF; HALS; classification; copper minerals

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) [1] is a remote sensing technique used for both qualitative and quantitative analysis of various materials. The operational principle is to use a high pulse energy laser to instantaneously heat the matter to evaporate a small amount of the substrate and eject it as a plasma plume. Then, the light emitted by the plasma is dispersed and registered by a camera. After the specified time of continuous wavelength radiation, a time window exists with the quickly cooling plasma, whereas the individual spectral lines representing the elemental material composition can be recorded.

The LIBS itself (apart from intensity of elements' spectral lines) does not deliver analytically relevant information, such as classification or quantification. For these purposes, software methods

must be engaged based on the input spectra and some mathematical operations to deliver the expected measurement outcome. The algorithms used for such spectral data analysis are numerous and cover a large area of statistics and machine learning fields of study [2,3]. Among them, the most frequently used with LIBS are: classifiers (linear discriminant analysis (LDA) [4], support vector machines (SVM) [5], and k -nearest neighbours (KNN) [6]); regression models (partial least squares [7], lasso [8], and Bayesian regression [9]); clustering algorithms (k -means [10]) and artificial neural networks (ANN) [11].

There are numerous fields where LIBS analysers are used [12], such as basic scrap metal analysis [13], classification of alloys [14], or mapping of geological cores [15]. Additional uses include sophisticated applications such as adulteration detection in milk [16], discrimination of heavy-metal-contamination in seafood [17], analysis of pathological tissues [18], or precision agriculture [19]. They are even used with the most demanding space missions [20] and biohazard detection [21]. The size of the LIBS equipment can vary from small handheld devices to large workstations. The size is usually commensurate with the increase in resolution, limit of detection, repeatability, and general unit functionality. However, the device size limits its usage in many in-situ applications where a handy analyser is perfect. Such situations occur when the amount of equipment is limited by personal lifting capacity or where the measurement location is extremely tight (i.e., mines and other geological sites).

In this study, we focus on the possibility of using a compact handheld LIBS to analyse the geological samples online—on-site. The materials analysed will be copper minerals existing in a natural state on their base substrate, so no sample preparations will be made. The outcome for the measurements should be a mineral identification of the rocks exposed to laser radiation. However, this is a nontrivial task because the mineral samples are very heterogeneous with irregular geometrical shapes. This heterogeneity makes it difficult for classification learners as the labels may be incorrectly assigned or impure, leading to misclassifications between minerals. Conversely, the irregular mineral shape becomes challenging for the optics and lasing capabilities of the handheld device, and the registered spectra will differ significantly in intensity and spectral channel coverage. Our proposed approach is to use a linear regression model for the classification purpose, while the regressors originate from a blind source separation algorithm. This algorithm will be capable of distinguishing the mineral spectra of interest from the base rock or other impurities of the analysed sample. In this case, a novel combination of LIBS with a non-negative tensor factorisation (NTF) [22] method was applied.

The NTF is an unsupervised learning method for extracting mode-related non-negative latent components from a multiway array (tensor). Assuming that the measured LIBS spectra from multiple measurement points can be collected to a 3-way tensor, the aim of NTF is to extract a few artificial (latent) spectra from each mineral in a given dataset, and then use them as predictors in a regression model for classification purposes. The latent spectra can be regarded as common patterns in the observed LIBS spectra and have a multilinear relationship with the spectra across each mode of the observed tensor. Such multimodal relationships cannot be revealed with matrix factorisation models, such as principal component analysis (PCA). Moreover, due to nonnegativity constraints, NTF yields spectra that have a physical sense and easy interpretation, whereas PCA provides only some orthogonal components (with negative entries) that could not be used as predictors in a regression model.

The final intent of this work is to create an analytical method that could be embedded within handheld or other mobile LIBS devices or systems, which can be used as a support and verification tool for geologists while prospecting for copper deposits.

2. Materials

2.1. Overview

This work is based on the analysis of 62 different copper minerals whose copper content varies significantly. Those minerals are: **azurite**, **malachite**, **brochantite**, **copiapite**, **devilline**, **fornacite**, **langite**, **nakauriite**, **natrochalcite**, **osarizawaite**, **posnjakite**, **vauquelinite**, **arthurite**, **chenevixite**,

clinoclase, conichalcite, cornetite (**co0**), cornubite, descloizite, duftite, libethenite, mottramite, olivenite, parnauite, pseudomalachite, richelsdorffite, tsumebite, turquoise, tyrolite, ajoite, allophane, chrysocolla, creaseyite, diopside, halloysite, plancheite, vesuvianite, algodonite, antimonpearceite, bornite, bourmonite, chalcopyrite, colusite, covellite, digenite, enargite, freibergite, germanite, gladiolite, idaite, jaskolskiite, krupkaite, seligmannite, stannite, stromeyerite, tetrahedrite, umangite, boleite, kinoite, cuprite, delafossite, and tenorite.

To shed light on the given classification problem and understand the general similarities of the minerals, they were all grouped in accordance with the Strunz Classes [23] given in Table 1. The first three characters of each mineral name (indicated with bold font in the above list) together with a Strunz Class number constitute short names used in the remainder of this article (with one exception for cornetite being **co0–8**). The exact Strunz Class assignment for each mineral is given in Table S1 in the supplementary data.

Table 1. List of mineral classes of analysed samples (based on original Strunz Class [23]).

Mineral Class Name	Class Number C
Sulfides	2
Halides	3
Oxides and Hydroxides	4
Carbonates and Nitrates	5
Sulfates	7
Phosphates	8
Silicates	9

The material suppliers were chosen from among different countries and/or geographical regions to differentiate the population of rocks with specific minerals and make the database more versatile. The origins of the samples include six continents, 29 countries, and 65 different regions as follows:

Australia (Kambalda), Austria (Tirol, Steiermark), Belarus (Rhodopien), Chile (Calama, Taltal, Atacama, Chuquicamata), Czechia (Severocesky, Horni Slavkov, Morava, Sredocesky), Democratic Republic Congo (Katanga), France (Corsica), Germany (Aachen, Harz, Hessen, Westerwald, Sauerland, Schwarzwald, Osthessen, Bad Ems, Saxony, Thuringia, Mansfeld), Greece (Laurion), Hungary (Rudabanya), Italy (Udine, Neapel), Japan (Aichi), Kazakhstan (Dzhezkazgan, Majkojyn), Morocco (Agadir, Bou Azzer, Bou Skour), Mexico (Durango), Namibia (Otavi, Kaokoveld), Peru (Huanzala), Poland (Polkowice), Portugal (Estremoz), Russia (Jakutien, Ural), Slovakia (Piesky, Michalovce, Lubjetova), Spain, Switzerland (Grisons, Wallis), Sweden (Gruvasen, Vena), Tajikistan (Mushiston Deposit), UK (Bristol, Cornwall), Ukraine (Nagolny Krjzh), USA (Nevada, Arizona, Montana, Utah, New Mexico, Missouri, California, Michigan, Montana), and Zambia (Kitwe).

The mineral shapes and sizes attached to the rocks were extremely diverse, from perfect crystals spread on the surface of the base rock to thin layers, often combined with the base rock, with unknown percentage distributions of both. Figure 1 presents examples of such mineral shapes. Examples are for chalcopyrite, azurite, and malachite, where each pair of photos (Figure 1a–f, respectively) shows two cases: (1) a well-built and clear crystal structure without impurities attached to the base rock, and (2) evenly distributed mineral over the rock surface with an unknown percentage mix with the base rock. Moreover, the attached crystal sizes differed from large, as in the case of chalcopyrite (Figure 1a), to very small as in the case of azurite (Figure 1c).

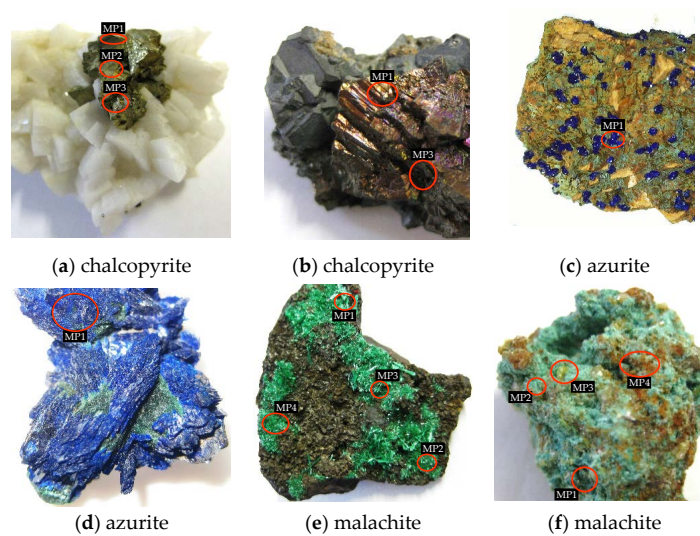


Figure 1. Examples of mineral distribution on the rock samples.

A total of 127 rock samples were taken into consideration. The statistics for the number of rock samples per mineral are presented in Figure S2 in the supplementary data.

2.2. Element Composition

As the chemical bonds are destroyed during the plasma creation, the analysed spectral signal is primarily dependent on the elements that existed in the material before the laser action. All 62 minerals are described in Table S1 in the supplementary data by their empirical element composition taken from [24], where a total of 27 elements were found. Then, based on the elemental composition, an investigation of the similarities between them was conducted with the use of PCA to find theoretically undistinguishable minerals by LIBS. In this way, a new N -dimensional ($N = 27$) space was created with new orthogonal variables called principal components (PCs) that describe the dataset sorted from the highest to the lowest variance [25]. Even with such PCA transformation, it was impossible to find good representative PCs to determine how the minerals are distinguishable from each other in a lower dimensional space (i.e., 3D representation).

In that case, N -dimensional measures are used to find mineral pairs that are most like each other. For this purpose, the Matlab[®] pdist function was used with four selected distance measures: Euclidean $d_{st}^2 = (x_s - x_t)^T(x_s - x_t)$, City block $d_{st} = \sum_j |x_{sj} - x_{tj}|$, Chebyshev, $d_{st} = \max_j \{|x_{sj} - x_{tj}|\}$,

and cosine $d_{st} = 1 - \frac{x_s^T x_t}{\|x_s\| \|x_t\|}$ [26]. The final computation was an average of those four measures as they gave similar yet different orders of hard to distinguish mineral pairs.

Figure 2a presents the least distinguishable 21 pairs of minerals. The worst cases were the **mal-5** and **azu-5** groups as well as a group consisting of **lan-7**, **pos-7**, and **bro-7**. These two barely reach 2% of the maximum distance between minerals in the PCA space. Figure 2b presents the worst-case groups of minerals extracted from Figure 2a with the other minerals distributed within the space of the first two PCs. Those groups are the already mentioned **lan-7/pos-7/bro-7** and **mal-5/azu-5** as well as **co0-8/lib-8/pse-8**, **cli-8/cor-8/oli-8**, **dio-9/chr-9/ajo-9**, **hal-9/all-9**, **gla-2/kru-2**, and the last one, **tsu-8/vau-7**, which is actually the first pair that shows close similarity above the Strunz Class division; hence, there is no need to search for another similar groups.

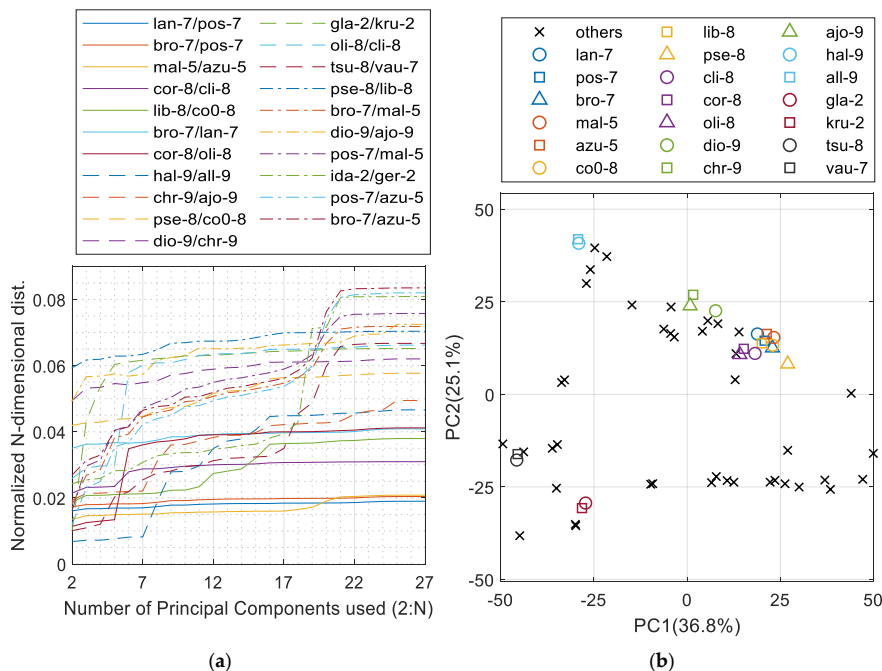


Figure 2. PCA analysis of the mineral dataset based on element composition: (a) the N-dimensional distance measure between the closest pairs of minerals; (b) 2D PCs plot indicating the most similar groups of minerals extracted.

Further analysis of the elemental composition values (Figure 3) clearly indicated that malachite and azurite as well as the group of langite, posnjakite, and brochantite cannot be distinguished from each other. As a result, new mineral labels were created: **a+m** for the **mal-5/azu-5** and **blp** for the other three. These minerals were combined and labelled together for further classification. The full heatmap with all 27 elements and 62 minerals is shown in Figure S3 in the supplementary data.

Cu	52.05	54.05	56.2	57.48	55.31	56.63	53.16	60.56	50.09	46.61	44.91	40.31	33.86	32.37			9	9.26	5.83	3.96	
H	2.06	1.71	1.34	0.91	0.58	0.9	0.42	0.19	0.79	0.89	0.36	1.28	1.92	0.82	1.56	2.24	0.14	0.15			
S	6.57	6.82	7.09															4.67	17.65	17.99	
O	39.32	37.42	35.37	36.18	37.14	33.27	33.46	27.44	29.43	30.51	28.27	40.6	45.06	40.74	55.78	57.57	20.4	20.99			
C				5.43	6.97																
As								19.69	21.98	26.47											
Pb																	58.7	60.41	19.01	12.92	
Al												2.05	2.06	20.9	23.97						
Si												17.82	17.1	20.74	21.76	16.22					
P					9.2	12.96	11.81											4.39	4.52		
Cr																		7.37			
Bi																				57.51	65.13
Na														1.02							
K														2.24							
	lan-7	pos-7	bro-7	mal-5	azu-5	co0-8	lib-8	pse-8	cli-8	cor-8	oli-8	dio-9	chr-9	ajo-9	hal-9	all-9	vau-7	tsu-8	kru-2	gla-2	

Figure 3. Heatmap (with values) representing elemental composition among the most similar minerals in accordance with the mineral's empirical formula [24]. The sorting from left to right was performed based on the outcome of Figure 2.

3. Experimental

The measurements were performed using a handheld LIBS device (Z-300, SciAps). The DPSS Nd:YAG laser emits radiation at a wavelength of 1064 nm, with a repetition rate up to 50 Hz, a pulse energy of 5–6 mJ, and a pulse duration of 1–2 ns. The spectrometer covers a range of 190–950 nm. Immediately before and during the measurement, the measuring region was purged with Ar gas to remove ambient air and enhance the LIBS signals. Each specific measurement point (MP) on the rock sample consisted of 64 single shots in an 8 × 8 grid covering an area of roughly 2–4 mm². It is worth noting that the device did not save spectra below a specific intensity threshold. Therefore, in some cases, fewer than 64 spectra were recorded per MP.

The mineral crystals on the surface were uniform, and the area covered by the 8 × 8 grid often extended beyond the crystal of interest. For that reason, all MPs had to be manually labelled. An example of such labelling is shown in Figure 4 and Table 2.

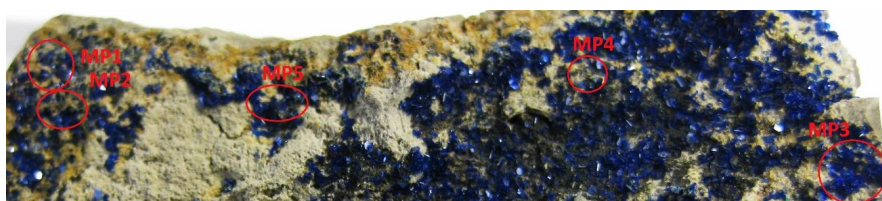


Figure 4. Image of azurite sample measuring points.

Table 2. Example description table of the sample shown in Figure 4.

MP	Mineral	Reference	Coverage
1	azurite	0	80
2	azurite	0	20
3	azurite	1	90
4	azurite	0	30
5	azurite	0	60

When an MP seems to cover more than 90% of the interesting mineral, it is flagged as a *possible reference*, and can then be used for both training and validation. If not flagged with logic 1 in the Reference column, then the MP remains as *validation only*. In the Coverage column of the description table, a percentage value of the mineral of interest within the 8 × 8 grid is provided. These values are rough assessments (limited to a difference of +/-5%) made just after the LIBS measurement, based on the pattern of laser spots on the sample and descriptions of the minerals of interest made by geologists. This will assist with further automated validation of MPs (because if they reach that percentage value, it means that classification succeeded). Even if the coverage description may in some cases be slightly inaccurate, this does not favour our proposed NTF method as the bias remains the same for all classification algorithms used.

During further classification, the MP is a single entity and the 8 × 8 grid spots will not be separated or analysed exclusively. In total, 458 MPs were recorded, of which 311 were flagged as *possible reference* and 147 as *validation only*. The full statistics of MPs per mineral are given in Figure S1 in the supplementary data.

4. Method

4.1. Latent Spectrum Extraction

The latent spectrum extraction in its principal form is used to select some underlying (latent) spectra from observed LIBS spectra that are considered as mixtures of multiple latent components generated by multiple spectral sources. The latent spectra are more frequent and have a common pattern across all observed LIBS spectra. In an ideal case, after the extraction, one latent spectrum should resemble an artificial spectrum of the desired source. In practise, we obtain few latent components that approximately represent the true spectra of the analysed minerals.

In the investigated case, each object is probed with several to a dozen measuring points. Thus, each MP contains a maximum of 64 spectra. Let $\mathbf{y}_{i_1, i_2}^{(m)} \in \mathbb{R}_+^{I_3^{(m)}}$ be the i_1 -th spectrum of the m -th object, measured in the i_2 -th MP. Each MP is assumed to provide $I_1^{(m)}$ spectra, and the number of MPs is equal to $I_2^{(m)}$, where $i_2 = 1, \dots, I_2^{(m)}$. The space of non-negative numbers is expressed by \mathbb{R}_+ . The spectra in each MP are indexed according to a lexicographical order, that is, $i_1 = 1, \dots, I_1^{(m)}$. The number $I_1^{(m)}$ is usually lower than 64 because some shots that correspond to the spectra of a low variance (below a threshold) must be neglected. The spectral resolution is determined by the number of samples in each spectrum, that is, the number $I_3^{(m)}$. Because the spectral resolution is the same for each observed spectrum, then $\forall m : I_3 = I_3^{(m)}$. We analyse M objects, where $m = 1, \dots, M$, assuming that each registered spectrum $\mathbf{y}_{i_1, i_2}^{(m)}$ can be regarded as a superposition of latent spectra that could be pure spectra of analysed minerals (endmembers) or other unwanted or perturbing spectra. The latent spectra for the m -th object can be collected into the matrix $\mathbf{U}^{(m,3)} = [\mathbf{u}_1^{(m,3)}, \dots, \mathbf{u}_{J_m}^{(m,3)}] \in \mathbb{R}_+^{I_3 \times J_m}$, where J_m is the number of latent spectra in the m -th object. Considering the above, the spectrum $\mathbf{y}_{i_1, i_2}^{(m)}$ can be expressed by the following superposition rule:

$$\mathbf{y}_{i_1, i_2}^{(m)} = \xi_{i_1, i_2, 1} \mathbf{u}_1^{(m,3)} + \dots + \xi_{i_1, i_2, J_m} \mathbf{u}_{J_m}^{(m,3)} = \sum_{j_m=1}^{J_m} \xi_{i_1, i_2, j_m} \mathbf{u}_{j_m}^{(m,3)}, \quad (1)$$

where the coefficient $\xi_{i_1, i_2, j_m} \geq 0$ determines the contribution of the j_m -th latent spectrum to the i_1 -th observed spectrum of the i_2 -th MP in the m -th object. The coefficient ξ_{i_1, i_2, j_m} can then be factorised as $\xi_{i_1, i_2, j_m} = u_{i_1, j_m}^{(m,1)} u_{i_2, j_m}^{(m,2)}$, where $u_{i_1, j_m}^{(m,1)} \geq 0$ represents the contribution from the i_1 -th shot and $u_{i_2, j_m}^{(m,2)} \geq 0$ refers to the contribution from the i_2 -th MP. Let $\mathbf{u}_{j_m}^{(m,1)} = [u_{i_1, j_m}^{(m,1)}] \in \mathbb{R}_+^{I_1^{(m)}}$ be a vector of coefficients $u_{i_1, j_m}^{(m,1)}$ for $i_1 = 1, \dots, I_1^{(m)}$, and $\mathbf{u}_{j_m}^{(m,2)} = [u_{i_2, j_m}^{(m,2)}] \in \mathbb{R}_+^{I_2^{(m)}}$ for $i_2 = 1, \dots, I_2^{(m)}$. Sweeping over the indices i_1 and i_2 , let $\mathcal{Y}^{(m)} = [\mathbf{y}_{i_1, i_2}^{(m)}] \in \mathbb{R}_+^{I_1^{(m)} \times I_2^{(m)} \times I_3}$ be a 3-way array (3-modal tensor) created from a set of spectra $\{\mathbf{y}_{i_1, i_2}^{(m)}\}$ for the m -th object. It is thus easy to notice that model (1) takes the form

$$\mathcal{Y}^{(m)} = \sum_{j_m=1}^{J_m} \mathbf{u}_{j_m}^{(m,1)} \circ \mathbf{u}_{j_m}^{(m,2)} \circ \mathbf{u}_{j_m}^{(m,3)}, \quad (2)$$

where the symbol \circ denotes the outer product. Model (2) can be equivalently expressed in the form

$$\mathcal{Y}^{(m)} = \mathcal{J}_m \times_1 \mathbf{U}^{(m,1)} \times_2 \mathbf{U}^{(m,2)} \times_3 \mathbf{U}^{(m,3)}, \quad (3)$$

where $\mathbf{U}^{(m,1)} = [\mathbf{u}_1^{(m,1)}, \dots, \mathbf{u}_{j_m}^{(m,1)}] \in \mathbb{R}_+^{I_1^{(m)} \times j_m}$, $\mathbf{U}^{(m,2)} = [\mathbf{u}_1^{(m,2)}, \dots, \mathbf{u}_{j_m}^{(m,2)}] \in \mathbb{R}_+^{I_2^{(m)} \times j_m}$, $\mathcal{J}_m \in \mathbb{R}_+^{j_m \times j_m \times j_m}$ is a superdiagonal identity tensor, and the symbol \times_n stands for the tensor-matrix product across the n -th mode.

Note that all factor matrices $\{\mathbf{U}^{(m,1)}, \mathbf{U}^{(m,2)}, \mathbf{U}^{(m,3)}\}$ contain only nonnegative numbers, and hence, model (3) can be regarded as the standard non-negative tensor factorisation (NTF) [22], which is a particular case of the CANDECOMP/PARAFAC (CP) decomposition [27,28].

Factor $\mathbf{U}^{(m,3)}$ contains j_m latent spectra, and $\mathbf{U}^{(m,1)}$ and $\mathbf{U}^{(m,2)}$ represent the contribution coefficients (concentrations) of the latent spectra to observations across the first and second modes of the tensor $\mathcal{Y}^{(m)}$, respectively. The latent spectra can thus be obtained by performing the NTF of $\mathcal{Y}^{(m)}$, given the assumed number j_m .

There are numerous computational strategies for NTF, and nearly all of them are based on an alternating optimisation scheme with unfolding imposed on each mode. Model (3) expressed in the unfolded version takes the form

$$\mathbf{Y}_{(n)}^{(m)} = \mathbf{U}^{(m,n)} (\odot_{p \neq n} \mathbf{U}^{(m,p)})^T, \tag{4}$$

where $\mathbf{Y}_{(n)}^{(m)} \in \mathbb{R}_+^{I_n^{(m)} \times \prod_{p \neq n} I_p^{(m)}}$ is a matrix obtained by the unfolding tensor $\mathcal{Y}^{(m)}$ along its n -th mode; where $n = 1, 2, 3$, and the symbol \odot stands for the Khatri-Rao product. Note that the system in (4) is considerably overdetermined because $\prod_{p \neq n} I_p^{(m)} \gg j_n$. To alleviate the problem of scaling ambiguity in the NTF, the columns in matrices $\mathbf{U}^{(m,2)}$ and $\mathbf{U}^{(m,3)}$ are normalised to the unit l_1 -norm. The system of linear equations in (4) can be solved with numerous linear solvers subject to nonnegativity constraints. In our study, we used the hierarchical alternating least-squares (HALS) algorithm proposed in [29], and then computationally improved in [30]. It belongs to a family of block coordinate descent update algorithms with monotonic convergence and computational complexity of $\mathcal{O}(N j_m \prod_{n=1}^N I_n^{(m)})$. The graphical representation of the NTF model is presented in Figure S4 of the supplementary data.

4.2. Regression Model Using Latent Spectra

To estimate the percentage rate of minerals in a newly measured MP, the latent spectra are extracted from the known MPs labelled with the so-called strong reference. We assume that j_m latent spectra are extracted from the m -th labelled object using the NTF. The latent spectra after being postprocessed can be regarded as regressors for predicting the percentage of minerals in such an unknown MP. Any unknown spectrum $\tilde{\mathbf{y}} \in \mathbb{R}_+^{I_3}$ is assumed to be approximated by a linear regression model,

$$\tilde{\mathbf{y}} \cong \sum_{m=1}^M \sum_{j_m=1}^{j_m} \alpha_{j_m}^{(m)} \mathbf{u}_{j_m}^{(m,3)}, \text{ where } \sum_{m=1}^M \sum_{j_m=1}^{j_m} \alpha_{j_m}^{(m)} = 1. \tag{5}$$

Coefficient $\alpha_{j_m}^{(m)}$ represents the contribution of the j_m -th latent spectrum from the m -th object to the unknown spectrum $\tilde{\mathbf{y}}$. Let $\mathbf{U} = [\mathbf{U}^{(1,3)}, \dots, \mathbf{U}^{(M,3)}] \in \mathbb{R}_+^{I_3 \times \sum_{m=1}^M j_m}$ and $\boldsymbol{\alpha} = [\alpha_1^{(1)}, \dots, \alpha_{j_1}^{(1)}, \alpha_1^{(2)}, \dots, \alpha_{j_M}^{(M)}] \in \mathbb{R}_+^{\sum_{m=1}^M j_m}$. Coefficient $\alpha_{j_m}^{(m)}$ can be estimated from model (5) by solving the following regularised constrained least-squares problem:

$$\boldsymbol{\alpha}_* = \operatorname{argmin}_{\boldsymbol{\alpha}} \frac{1}{2} \|\tilde{\mathbf{y}} - \mathbf{U}\boldsymbol{\alpha}\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\alpha}\|_2^2, \text{ s.t. } \boldsymbol{\alpha} \geq 0, \text{ and } \|\boldsymbol{\alpha}\|_1 = 1, \tag{6}$$

where $\lambda \geq 0$ is a regularisation parameter that controls the overfitting. In this study, problem (6) was solved using the interior-point least-squares algorithms for regularised box-constrained problems,

implemented in the function `lsqlin` in Matlab® 2016a. Note that coefficients in vector α_* can also be regarded as unknown percentages of expected minerals in the analysed MP.

4.3. Determining the Number of Latent Spectra per Mineral

The last problem to be solved is how to determine the correct (sufficient) J_m number of latent spectra for each m -th object. The problem would be trivial if only pure mineral samples were analysed, since only a single latent spectrum may represent the desired mineral. However, in our measurements, impurities in the spectra resulting from base rocks and other minerals will occur. Moreover, the NTF is sensitive to the difference in light intensity distribution between spectrometer channels, which is not constant because the light is propagated inside the device and thus can cause additional perturbations of the desired m -th object spectra. In that case, the assumption was made that we will search for J_m latent spectra in which most of them represent the scattered spectra of the desired mineral. The less contributing examples after the so-called self-regression will be labelled as unknown spectra (U).

In such a case, we developed an iterative process in which we start from $J_m = 2$ ($J_m = 1$ will be similar to a weighted average of the data) and increment the number until reaching a break loop condition. The loop breaking condition is based on two parameters: level L and ratio R , which will be set up prior to this procedure.

The L -condition is superior and relates to the cumulative contribution given by the sum of coefficients $\left\{ \alpha_{j_m}^{(m)} \right\}$ of the first K latent spectra ($j_m = 1, \dots, K$), sorted in descending order of their contributions. Note that number K cannot be equal to J_m , which means that at least one latent spectrum should be classified as undefined U . If the cumulative contribution of K latent spectra in a given iterative step is equal to or greater than the L value, then the first K spectra are assigned as the desired mineral, and the last $J_m - K$ spectra are labelled as U . Otherwise, J_m is incremented and again the L -condition is checked.

If the L -condition is satisfied, then the second R -condition should be checked. This condition requires the minimum ratio between the last (K -th) spectrum assigned as a mineral to the first ($J_m - K + 1$) spectrum assigned as U . This relation is very important, as for a higher value of J_m , the distribution of latent spectra contributions might end up equal. Therefore, it may be difficult to assess the U spectra, as the difference between the last labelled as mineral and first labelled as U may be only a few percentage points. If the R -condition is not met, J_m is incremented, and the L -condition is checked.

After meeting these two (L and R) conditions, we achieve the solution for which we have the desired level of contributions and the ratio of the contributions of the last mineral spectrum to the first U spectrum large enough to assume the U spectra are really the unwanted signals. An example of such a loop operation with given L/R conditions is shown in Figure S5 of the supplementary data.

Obviously, for some combination of R and L in a given m -th object, such a pair of conditions can never be met. To avoid an infinite loop, parameter $J_m^{(max)}$ is introduced, which is the maximum J_m number to be incremented in the loop. If the conditions are not met after $J_m = J_m^{(max)}$, then a given pair of L/R conditions is excluded from consideration for all m objects.

Figure 5 presents an example of such latent spectra extraction for malachite with R set to 1.8 and L to 0.96. In this case, three components were extracted as scattered in the m -th object and one as the U spectrum.

For the final validation process of a new unknown MP, the scattered K latent spectra of each m -th object become different regressors. Finally, the predicted percentage contribution for the m -th object in the new MP becomes the sum of the single K contributions of the m -th object scattered latent spectra. The classifying label is then decided on the mineral that has the top percentage contribution.

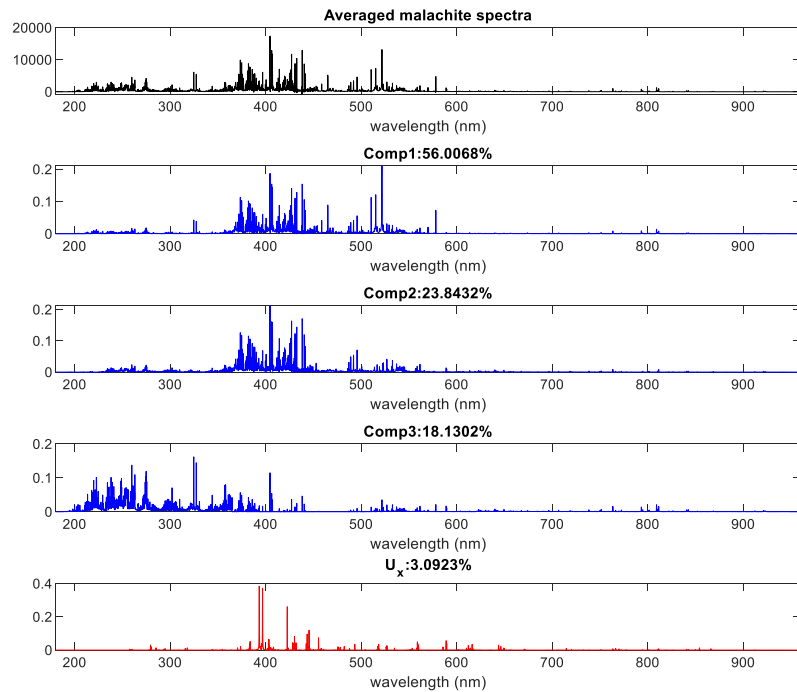


Figure 5. Example of latent spectra extraction. The averaged spectrum originates from all malachite MPs.

5. Results

5.1. Setup

The number of MPs flagged as *validation only* differs among minerals, some do not even have such or have only one. To make the proportion of training and validation data equal among the minerals, some of the MPs flagged as *possible reference* were also moved to be validation data. This new division can be observed in Figure S6 of the supplementary data.

The parameters that were set for our algorithm were a maximum number of latent spectra ($J_m^{(max)} = 20$), ratios R (from 1.1 to 2.0), and levels L (from 0.80 to 0.98). The following sections present 10×10 heatmaps with classification measures from which we can select the best parameters for the analysis of the confusion matrix.

Because we deal with an imbalanced dataset, there is a risk of the results being overwhelmed by the outcome of the larger mineral classes. Thus, the metrics of precision and recall are introduced together with their bounding metric called F -measure [31]:

$$Precision = \frac{tp}{tp + fp}, \quad Recall = \frac{tp}{tp + fn}, \quad F = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

where tp is the true positive rate, fp is the false positive rate, and fn is the false negative rate.

The proposed NTF-based classification also introduces the U class (Section 4.3), which should be included in some way in the metrics in (7). For that purpose, we propose to differentiate two cases: U_{in} and U_{ex} .

The U_{in} case assumes that we include MPs labelled as U in the classification score calculation. However, the metrics in (7) require, among others, true positive rates, which in the case of U class, will never exist (U is not a true class). This will result in precision and recall being zero all the time and independently from the number of MPs labelled as U , the F -measure (if calculated) for U class will always be zero and, therefore, lower the total F -measure for all the classes together. The only way to introduce the U class into the F -measure, therefore, is assuming that if any MP is going to be assigned as U class, it will appear as a false negative value for the original true class (i.e., azurite labelled as U).

The U_{ex} case totally excludes MPs labelled as U from the classification score calculation. This decision was made on the assumption that U labels give us the information that the measurement outcome is uncertain and the user should repeat the measurement on such a sample for more confidence (this is not yet an error at that point but restrains us from introducing false positive rates into another class).

Again, as U is not a true class, the false positive rates will never exist for it, so, actually, the only difference between U_{in} and U_{ex} scores will be held by the recall part of the equation in (7) (the precision will stay the same for both).

The NaN values within the heat maps are related to the L/R pairs with which the algorithm could not find a latent extraction solution for at least one mineral. For improved clarity, each heatmap has the top three L/R solutions listed in its title.

The result of the proposed algorithm is contrasted with the performance of the Matlab[®] built-in classification algorithms: SVM (templateSVM: KernelFunction—linear; BoxConstraint—1; standardised input), LDA (fitcdiscr: DiscrimType—linear; standardised input), and KNN (fitknn: Distance—Euclidean; NumNeighbors—1; DistanceWeight—equal, standardised input) [2].

As the confusion matrices in the case of our 59 classes were very large, they were added as supplementary data (Figures S8–S11).

5.2. Analysis of Training Data

The first portion of the results is focussed on finding the ideal combination of parameters for our proposed method. To do this, the model was trained and validated using the same training data (Figure S6—supplementary data). This is reasonable, as while the proposed algorithm extracts the latent spectra, it does it within a single mineral class and is unaware of the existence of other classes. Because of this, it is unlikely for the model with so many classes to reach a 100% score even when trained and validated with the same data (which is different in the case of the SVM that reached 100% under the same conditions). However, this gives us an opportunity to revise the model on the basis of training data and determine the best combination of R and L values for the given dataset.

From Figure 6, we see that the F -measure score increased with increasing L value. Moreover, it is clear that we reached the parameterisation boundary from each side of the 10×10 matrices, as it was impossible to deliver results for L equal or greater than 0.98 and R equal to 1.8 (or greater). If the L values were taken from the range [0.8, 0.82], we had few possible solutions and much lower scores. Finally, it was futile to use R values smaller than 1.1 because 1.0 would designate equality.

The best results for both $F_{U_{in}}$ (Figure 6a) and $F_{U_{ex}}$ (Figure 6b) are a combination of $R = 1.5$ and $L = 0.96$, and those parameters will be selected as a final solution with the validation dataset.

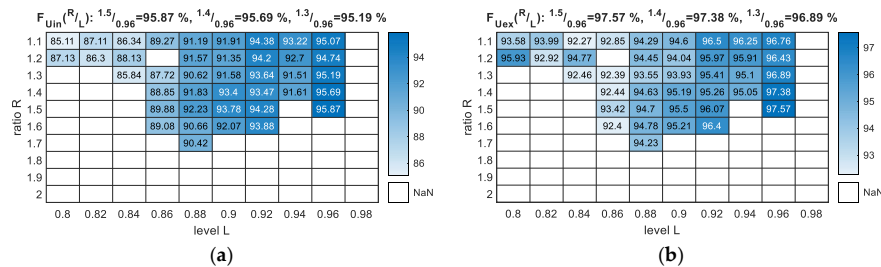


Figure 6. *F*-measure output for the parameterisation of the NTF method with *R* and *L* variables (training data vs. training data): (a) F_{Uin} —scores with *U* class included as an error; (b) F_{Uex} —scores with *U* class excluded from measure.

5.3. Analysis of Validation Data

For this section, analyses were taken with separate training and validation data, in accordance with Figure S6 in the supplementary data. Although we already selected parameters in the previous section, we decided to perform the same parameterisation of *R* and *L*, resulting in 10 × 10 matrices to verify that the selection was accurate.

Similar to Figure 6, the results in Figure 7 indicate (with few exceptions) that the *F*-measure score increased with an increase in *L*. The boundary parameters also remained the same as the learned model, as in the previous section.

Figure 7 presents the *F*-measure results of the parameterisation, but apart from the scores for the *Uin* and *Uex* cases, it also presents the results in comparison to the best selected classifier (Figure 7c,d), which, in our case, was SVM scoring 67.22%. It is important to note here that this was the highest value that was possible to reach for an SVM trying different kernel functions and their parameterisation. In fact, the basic linear SVM scored the best among all SVM variants that was investigated. The additional results for NTF covering the *F*-measure’s partial scores—precision and recall—are presented in Figure S7 in the supplementary data.

Although the best results in the case of Figure 7 were not for *R* = 1.5 and *L* = 0.96, these results were selected as final because they could be foreseen (Section 5.2) and did not differ much from the other high scores, especially in the case of the highest results such as F_{Uex} .

Table 3 presents the discussed measures obtained for the NTF-based method and three standard classifiers for the analysed validation data. The best built-in solution was SVM, which reached an *F*-measure of approximately 17.5% better than that associated with KNN and LDA. The proposed algorithm reached 1.87% (*Uin*) and 5.02% (*Uex*) of the *F*-measure score with respect to the second-best SVM, followed by increases in every other case where R_{Uex} had the highest gain of 6.92%.

Table 3. Classification measures of the validation dataset with the use of analysed methods. Brackets indicate the NTF accuracy gain in comparison to the best SVM classifier.

Measure	SVM	KNN	LDA	NTF _{Uin}	NTF _{Uex}
Precision (%)	71.76	51.77	52.01	74.74 (+2.98)	74.74 (+2.98)
Recall (%)	68.48	54.97	54.16	70.13 (+1.65)	75.40 (+6.92)
F-Measure (%)	67.22	49.48	49.77	69.09 (+1.87)	72.24 (+5.02)

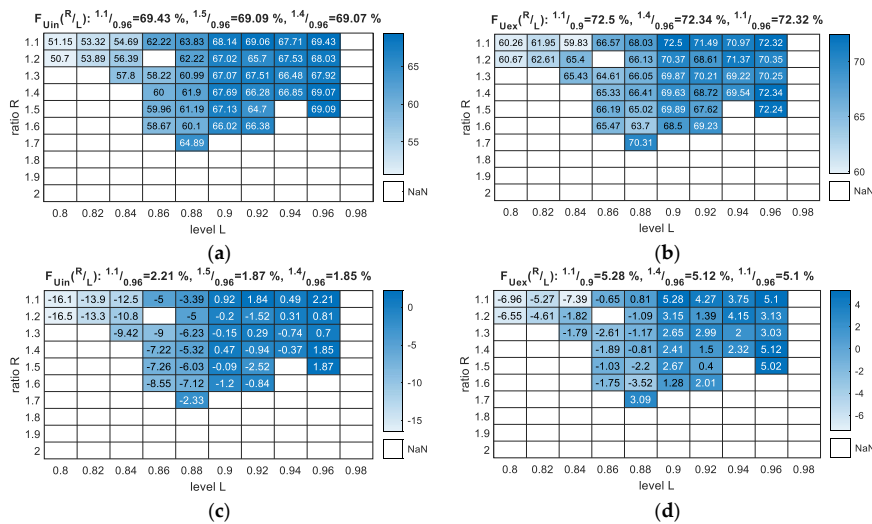


Figure 7. Classification F -measure output for the parameterisation of the NTF method with R and L variables (validation dataset): (a) $F_{U_{in}}$ —scores with U class included as an error; (b) $F_{U_{ex}}$ —scores with U class excluded from measure; (c) $F_{U_{in}}$ —scores with U class included as an error in accordance with the best SVM result from Table 3; (d) $F_{U_{ex}}$ —scores with U class excluded from measure in accordance with the best SVM result from Table 3.

The confusion matrices for the built-in classifiers present a trend to seek a host for the more likely hard-to-assess samples. In the case of the best out of three SVM (Figure S8—supplementary data), 23 mineral classes had a 100% score, and the classes that caused the most errors were **blp** (19 false positives) and **a+m** (12 false positives). For LDA (Figure S9—supplementary data), only 11 mineral classes had a 100% score, and the most confusing was **all-5** (16 false positives). The KNN (Figure S10—supplementary data) performed similarly to LDA for the final F -measure score, while 13 mineral classes were perfectly assessed, and the classes that caused the most errors were **a+m** (19 false positives) and **ajo-5** (13 false positives). Using the NTF method (Figure S11—supplementary data), we managed to classify the top F -measure score and the top 25 mineral classes without error. The U class perfectly took the top host position for the hardest to assess samples (17 false positives), followed by **cha-2** (14 false positives) and **cup-4** (8 false positives), which is very reasonable.

The CPU time and disk space usage of the above methods are compared in Table S12 in the supplementary data. It is clear from this table that the proposed NTF-based method requires less disk space and needs half the time for data validation compared to the competing SVM.

5.4. Example of Mineral Contribution for Selected MPs

The regression method proposed, apart from classifying the MPs, also gives the mineral percentage contribution followed by their geometrical distribution within the 8×8 shooting grid. Figure 8 presents the results for MP1 from the selected chalcopyrite sample. The bar plot presents the single contributions of the regressors (J_m latent spectra) summed for each mineral. In this MP case, the top count goes to chalcopyrite (almost 80%), leaving all the other minerals far behind, so the given classification label is correct. The geometrical distributions of the four main minerals (m -th objects) and the unknown class were equal within the 8×8 grid. Both can be confirmed with the MP photo, as the surface looks like an equal mineral distribution without any sign of the base rock.

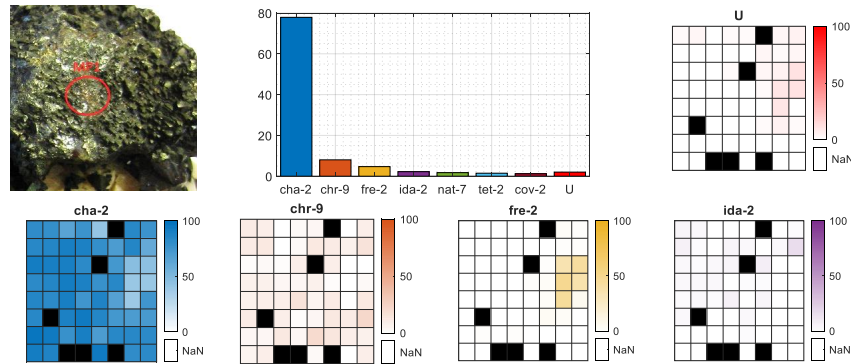


Figure 8. Mineral contribution for selected MP of chalcopyrite sample. Bar chart presents the total percentage rate of minerals per MP, the image indicates the place of measurement on the sample surface and the heatmaps present the geometrical distribution of the four main minerals and unknown class within 8×8 shooting grid. Black fields are missing spectra.

However, as mentioned in Section 2, there were also samples that did not have a consistent mineral distribution on the surface, and the crystal size was even smaller than the 8×8 shooting grid. An example of such a situation is the selected azurite sample presented in Figure 9. From the photo, we observe that the azurite crystal actually covers less than one-quarter of the visible, burned in laser pattern. This fact is clearly visible in the mineral distribution heatmaps of **a+m**, **cha-2**, and **chr-9**, where in the case of the proper azurite (**a+m**) class in the bottom-left quarter, we observe a higher contribution of that mineral, while in the same area on the **cha-2** and **chr-9** heatmaps, there is almost zero contribution. The total sum of **a+m** contribution on the bar plot is almost 30%, which perfectly covers one-quarter of the crystal that fits within the 8×8 shooting grid plus some average error of the latent spectra contribution.

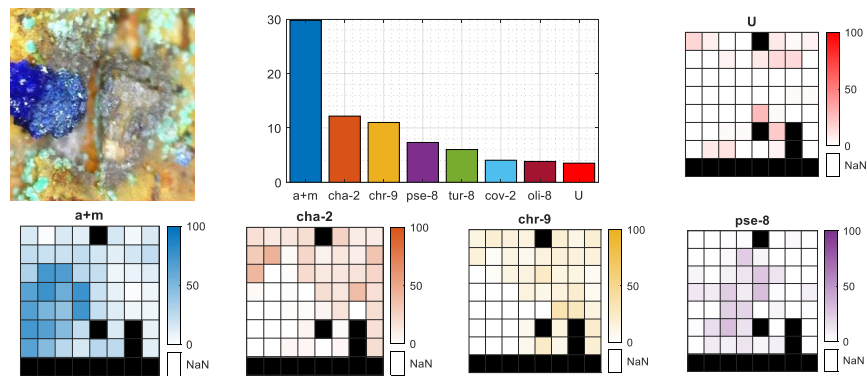


Figure 9. Mineral contribution for selected MP of azurite sample. Bar chart presents the total percentage rate of minerals per MP, the image indicates the place of measurement on the sample surface and the heatmaps present the geometrical distribution of the four main minerals and unknown class within 8×8 shooting grid. Black fields are missing spectra.

In both cases, as the classification labels were correctly set, the *U* class did not play an important role in the final percentage rate within those MPs. Both MPs are missing some spectra.

6. Conclusions

In this research, we hypothesised that it is possible to extract artificial (latent) spectra for each of the investigated minerals and use them as predictors in a linear regression model. For this purpose, NTF was used. Because of the heterogeneity of the mineral samples and weak reproducibility of the spectra acquired with the use of a handheld LIBS device, the procedure for proper latent spectra selection was proposed. In such a procedure, a parameterisation of ratio *R* and level *L* variables is required. The results show that these variables can be limited, with high accuracy, to one selection by performing validation with the use of the same data as for the model training.

The NTF-based classification performed well, reaching higher *F*-measure scores of around 1.9% (when the Unknown class was included in the measures) and around 5.0% (when the Unknown class was excluded from the measures), both in accordance with the best SVM classifier. The standard methods seem to find a host class for the hardest-to-assess samples, so using the method that already contains an Unknown class inside its model was even more reasonable here.

In addition to the output labelling required for classification purposes, a percentage contribution of the minerals within measuring points is given. Such additional information on the MPs allows the creation of 2D mapping of the shooting area 8×8 grids with a smooth distribution of the mineral classes among them.

The final regression model can be stored using low disk space, and the regression function is not as memory- and CPU-intensive (compared to the commonly used classifiers SVM, LDA, and KNN). Such a model may even be implemented in mobiles and other handheld LIBS devices and thus increase their functionality as fast, on-site mineral analysers.

Future research will be devoted to applying the method to minerals other than those containing copper as their primary element and verifying its universal usability. The NTF was confirmed as a method for extracting artificial spectra for the mineral classes and was successfully used as a regressor in a linear model.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-8220/20/18/5152/s1>, Table S1: detailed list of 62 minerals used in research; Figure S1: the number of measurement points (MPs) taken per mineral together with their partitions into *possible reference* and *validation only* ones; Figure S2: the number of different (exclusive) rock samples on which MPs were taken—listed per mineral; Figure S3: heatmap representing elements composition among all 62 minerals in accordance with the mineral's empirical formula [24]; Figure S4: NTF model; Figure S5: example of J_m number of latent spectra decision for one *m*-th object; Figure S6: dataset partitioning of the training and validation data using some of the *possible reference* MPs as validation data to evenly match the dataset among mineral classes; Figure S7: classification precision and recall output for the parameterisation of the NTF method with *R* and *L* variables (validation dataset), Figure S8: confusion matrix of SVM classifier for validation dataset; Figure S9: confusion matrix of KNN classifier for validation dataset; Figure S10: confusion matrix of LDA classifier for validation dataset; Figure S11: confusion matrix of NTF based classifier for validation dataset; Table S12: CPU time and disk space usage of the compared methods.

Author Contributions: Conceptualization, M.W., R.Z., A.A., D.R., and T.B.; methodology, R.Z. and M.W.; software, M.W. and R.Z.; validation, M.W., S.M., and K.C.; formal analysis, R.Z.; investigation, P.B. and M.W.; resources, P.B., D.R., and T.B.; data curation, P.B. and M.W.; writing—original draft preparation, M.W. and R.Z.; writing—review and editing, D.R., S.M., T.B., and A.A.; visualization, M.W.; supervision, T.B., A.A., and D.M.; project administration, K.C., D.R., and T.B.; funding acquisition, P.B., D.R., T.B., S.M., K.C., and D.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful for financial support from the German federal state of Brandenburg and the European Regional Development Fund (ERDF 2014-2020) as well as for project management by the Investitionsbank des Landes Brandenburg (ILB) and the economic development agency Brandenburg (WFBB) in the LIBSqORE project (grant number 80172489).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cremers, D.A.; Multari, R.A.; Knight, A.K. Laser-Induced Breakdown Spectroscopy. In *Encyclopedia of Analytical Chemistry*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2016; pp. 1–28. ISBN 9780470027318.
2. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2009; ISBN 9780387848587.
3. Alpaydin, E. *Introduction to Machine Learning*; Adaptive Computation and Machine Learning Series; MIT Press: Cambridge, MA, USA, 2014; ISBN 9780262325752.
4. Pontes, M.J.C.; Cortez, J.; Galvão, R.K.H.; Pasquini, C.; Araújo, M.C.U.; Coelho, R.M.; Chiba, M.K.; de Abreu, M.F.; Madari, B.E. Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain. *Anal. Chim. Acta* **2009**, *642*, 12–18. [[CrossRef](#)] [[PubMed](#)]
5. Zhang, T.; Wu, S.; Dong, J.; Wei, J.; Wang, K.; Tang, H.; Yang, X.; Li, H. Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (LIBS) coupled with support vector machine (SVM) and partial least square (PLS) methods. *J. Anal. At. Spectrom.* **2015**, *30*, 368–374. [[CrossRef](#)]
6. Li, X.; Yang, S.; Fan, R.; Yu, X.; Chen, D. Discrimination of soft tissues using laser-induced breakdown spectroscopy in combination with k nearest neighbors (kNN) and support vector machine (SVM) classifiers. *Opt. Laser Technol.* **2018**, *102*, 233–239. [[CrossRef](#)]
7. Feng, J.; Wang, Z.; West, L.; Li, Z.; Ni, W. A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy. *Anal. Bioanal. Chem.* **2011**, *400*, 3261–3271. [[CrossRef](#)] [[PubMed](#)]
8. Brickley, R.S.; Brown, D.J.; Turk, P.J.; Clegg, S. Comparing vis-NIRS, LIBS, and Combined vis-NIRS-LIBS for Intact Soil Core Soil Carbon Measurement. *Soil Sci. Soc. Am. J.* **2018**, *82*, 1482–1496. [[CrossRef](#)]
9. Menking-Hoggatt, K.; Arroyo, L.; Curran, J.; Trejos, T. Novel LIBS method for micro-spatial chemical analysis of inorganic gunshot residues. *J. Chemom.* **2019**, e3208. [[CrossRef](#)]
10. Wójcik, M.R.; Zdunek, R.; Antończak, A.J. Unsupervised verification of laser-induced breakdown spectroscopy dataset clustering. *Spectrochim. Acta Part B At. Spectrosc.* **2016**, *126*, 84–92. [[CrossRef](#)]
11. Ramil, A.; López, A.J.; Yáñez, A. Application of artificial neural networks for the rapid classification of archaeological ceramics by means of laser induced breakdown spectroscopy (LIBS). *Appl. Phys. A Mater. Sci. Process.* **2008**, *92*, 197–202. [[CrossRef](#)]
12. Hahn, D.W.; Omenetto, N. Laser-Induced Breakdown Spectroscopy (LIBS), Part II: Review of Instrumental and Methodological Approaches to Material Analysis and Applications to Different Fields. *Appl. Spectrosc.* **2012**, *66*, 347–419. [[CrossRef](#)]
13. Merk, S.; Scholz, C.; Florek, S.; Mory, D. Increased identification rate of scrap metal using Laser Induced Breakdown Spectroscopy Echelle spectra. *Spectrochim. Acta—Part B At. Spectrosc.* **2015**, *112*, 10–15. [[CrossRef](#)]
14. Herrera, K.K.; Tognoni, E.; Gornushkin, I.B.; Omenetto, N.; Smith, B.W.; Winefordner, J.D. Comparative study of two standard-free approaches in laser-induced breakdown spectroscopy as applied to the quantitative analysis of aluminum alloy standards under vacuum conditions. *J. Anal. At. Spectrom.* **2009**, *24*, 426–438. [[CrossRef](#)]
15. Kuhn, K.; Meima, J.A.; Rammelmair, D.; Ohlendorf, C. Chemical mapping of mine waste drill cores with laser-induced breakdown spectroscopy (LIBS) and energy dispersive X-ray fluorescence (EDXRF) for mineral resource exploration. *J. Geochem. Explor.* **2015**, *161*, 72–84. [[CrossRef](#)]
16. Moncayo, S.; Manzo, S.; Rosales, J.D.; Anzano, J.; Caceres, J.O. Qualitative and quantitative analysis of milk for the detection of adulteration by Laser Induced Breakdown Spectroscopy (LIBS). *Food Chem.* **2017**, *232*, 322–328. [[CrossRef](#)] [[PubMed](#)]
17. Ji, G.; Ye, P.; Shi, Y.; Yuan, L.; Chen, X.; Yuan, M.; Zhu, D.; Chen, X.; Hu, X.; Jiang, J. Laser-induced breakdown spectroscopy for rapid discrimination of heavy-metal-contaminated seafood *Tegillarca granosa*. *Sensors* **2017**, *17*, 2655. [[CrossRef](#)]
18. Hamzaoui, S.; Khleifia, R.; Jaïdane, N.; Ben Lakhdar, Z. Quantitative analysis of pathological nails using laser-induced breakdown spectroscopy (LIBS) technique. *Lasers Med. Sci.* **2011**, *26*, 79–83. [[CrossRef](#)] [[PubMed](#)]
19. Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.G.; Gebbers, R. Soil nutrient detection for precision agriculture using handheld laser-induced breakdown spectroscopy (LIBS) and multivariate regression methods (PLSR, lasso and GPR). *Sensors* **2020**, *20*, 418. [[CrossRef](#)]

20. Cremers, D.A. Space Applications of LIBS. In *Laser-Induced Breakdown Spectroscopy: Theory and Applications*; Musazzi, S., Perini, U., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 257–291. ISBN 978-3-642-45085-3.
21. DeLucia, F.C.; Samuels, A.C.; Harmon, R.S.; Walters, R.A.; McNesby, K.L.; LaPointe, A.; Winkel, R.J.; Miziolek, A.W. Laser-induced breakdown spectroscopy (LIBS): A promising versatile chemical sensor technology for hazardous material detection. *IEEE Sens. J.* **2005**, *5*, 681–689. [[CrossRef](#)]
22. Shashua, A.; Hazan, T. Non-negative tensor factorization with applications to statistics and computer vision. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 792–799.
23. Nickel, H.S.E. *Strunz Mineralogical Tables*, 9th ed.; Schweizerbart Science Publishers: Stuttgart, Germany, 2001; ISBN 9783510651887.
24. Mineralogy Database. Available online: <http://webmineral.com/> (accessed on 1 July 2020).
25. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
26. MathWorks Pdist—Pairwise Distance between Pairs of Observations. Available online: <https://www.mathworks.com/help/stats/pdist.html> (accessed on 27 July 2020).
27. Carroll, J.D.; Chang, J.J. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika* **1970**, *35*, 283–319. [[CrossRef](#)]
28. Harshman, R.A. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multimodal factor analysis. *UCLA Work Pap. Phon.* **1970**, *16*, 1–84.
29. Cichocki, A.; Zdunek, R.; Amari, S.I. Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. *Lect. Notes Comput. Sci.* **2007**, *4666 LNCS*, 169–176. [[CrossRef](#)]
30. Phan, A.-H.; Cichocki, A. Multi-Way Nonnegative Tensor Factorization Using Fast Hierarchical Alternating Least Squares Algorithm (HALS). In Proceedings of the 2008 International Symposium on Nonlinear Theory and Its Applications, Budapest, Hungary, 7–10 September 2008; pp. 41–44.
31. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

4. Diskussion

Ziel dieser Arbeit ist die Entwicklung von Kalibrierungs- und Quantifizierungs- sowie Klassifizierungsstrategien, um chemisch-analytische Voraussetzungen für eine nachhaltige Lagerstätten erkundung und Landwirtschaft zu schaffen. Für eine effizientere und somit ressourcenschonendere Nutzung natürlicher Ressourcen, wie Boden oder mineralischer Rohstoffe, müssen die erforderlichen Parameter charakterisiert werden. Dies sollte mittels einer modernen, instrumentellen Analysetechnik, die die relevanten Analyten identifizieren und zuverlässig detektieren kann, realisiert werden. Dafür wurde die laserinduzierte Breakdownspektroskopie an geologischen Proben in Kombination mit uni- und multivariater Datenanalyse eingesetzt. Die laserspektroskopischen Untersuchungen erfolgten an Agrarböden und SEE-haltigen Böden sowie an kupferhaltigen mineralischen Rohstoffen.

4.1 Kalibrierung

Ein nachhaltiger und ressourcenschonender Umgang bei der Nutzung natürlicher Ressourcen sind die obersten Ziele, um einerseits die Umwelt durch eine Rohstoffentnahme oder -nutzung möglichst gering zu belasten und andererseits möglichst langfristig explorierte Lagerstätten zu nutzen. Bei der Prospektion und Exploration neuer Lagerstätten sowie der gezielten Ausbeutung bestehender Lagerstätten bedeuten vor allem die zahlreichen geophysikalischen und -chemischen Analysemethoden zur Untersuchung der Materie starke Eingriffe in die Natur. Insbesondere durch Bohrungen und Probennahmen, die vorrangig Aufschluss über die Rohstoffgehalte und damit die Bauwürdigkeit der Lagerstätte liefern sollen, wird die Umwelt beeinträchtigt. Eine Analysetechnik, die ohne oberflächliche Probennahme vor Ort eine Aussage über die Rohstoffgehalte liefern kann, ist die ideale Lösung, um Eingriffe in die Umwelt zu reduzieren oder gänzlich zu vermeiden. Eine solche Untersuchungsmethode kann die LIBS sein. Für eine Quantifizierung, also eine Gehaltsbestimmung der elementaren Zusammensetzung, wird eine Kalibrierung benötigt. Eine Kalibrierung spiegelt die Beziehung zwischen den Merkmalen einer Testprobe und einer realen Probe wieder.⁹⁴ Um diese Beziehung bestmöglich wiederzugeben, muss die Kalibrierung einige Anforderungen erfüllen. Im Allgemeinen muss die Kalibrierung eine hohe Präzision und Linearität aufweisen. Für die Anwendung bei geologischen Proben sollte die Kalibrierung zusätzlich einen großen Elementgehaltbereich, zum Teil mehrere Größenordnungen, abdecken und im besten Fall auf unterschiedliche Lagerstätten anwendbar sein. In der LIBS wird die Kalibrierung elementbezogen durchgeführt. Das heißt, dass die zu untersuchenden Elemente spektrale Linien im detektierten LIB-Spektrum aufweisen müssen. Diese sollten nach Möglichkeit keine Überlappungen zu anderen Linien aufweisen und sehr geringe oder keine Selbstabsorptionseffekten zeigen.⁹⁵ Problematisch bei der Kalibrierung mittels LIBS ist zudem das mögliche Auftreten von Matrixeffekten, die die Kalibrierung beeinflussen können. Eine ideale Lösung könnte CF-LIBS sein, jedoch bleibt die analytische Genauigkeit dieser kalibrierungsfreien Methode hinter der Genauigkeit der kalibrierungsbasierten Verfahren zurück.⁹⁶ Ursächlich hierfür sind unter anderem Annahmen die getroffen werden, wie etwa, dass die Plasmazusammensetzung exakt der Probenstöchiometrie entspricht.⁹⁷ Die CF-LIBS ist bis heute Gegenstand der Forschung, da sie bisher, vor allem bei komplexen Matrices, keine exakten Ergebnisse liefert.⁹⁵

Ziel der in diesem Abschnitt vorgestellten Untersuchungen war die Entwicklung einer Kalibrierungsstrategie für LIBS an geologischen Feldproben. Verglichen wurden uni- und multivariate

Methoden im Hinblick auf eine stabile Kalibrierung der spektroskopischen Methode unter der Berücksichtigung von Matrixeffekten. Dafür wurden spektroskopische Untersuchungen mittels LIBS an Agrarböden durchgeführt. Es wurden Proben von zwei unterschiedlichen, landwirtschaftlich genutzten Feldern bezüglich der Elementgehalte von Calcium, Magnesium und Eisen analysiert. Eine getestete Kalibrierungsmethode beruhte auf einer univariaten Regression. Univariante Methoden können als Charakterisierungsmethoden benutzt werden, um Probeneigenschaften aus Spektren abzuleiten. Für die Bestimmung dieser Eigenschaften des untersuchten Probenmaterials müssen nach Haaland und Thomas drei Voraussetzungen erfüllt sein: Die Substanzeigenschaften sind von der chemischen Struktur abhängig, eine Änderung der chemischen Struktur führt zu einer Veränderung der Substanzeigenschaften, die sich in einer Änderung der Spektren manifestiert und schließlich sind diese Eigenschaften linear mit den Spektrenintensitäten verknüpft.⁹⁸ Die für die univariate Regression ausgewählten spektralen Linien des LIB-Spektrums sollen zusätzlich die oben beschriebenen Charakteristika aufweisen. Der erste getestete univariate Kalibrierungsansatz basiert auf einer Standardaddition. Hierbei wurde ausschließlich eine Bodenprobe für die Kalibrierung genutzt und das entwickelte Modell auf die LIBS-Daten beider untersuchten Felder angewendet. Im Gegensatz zu dieser Methode beruhte die univariate Kalibrierung im zweiten Ansatz auf der Referenzanalytik aller Proben eines der beprobten Felder. Mittels der LIBS-Daten des zweiten Feldes erfolgte die Validierung der Vorhersage. Als dritte Strategie wurde ein multivariater Ansatz gewählt. Dieser basierte auf der partiellen Regression der kleinsten Quadrate (PLSR). Für das Training des Modells wurden die LIBS-Daten der Agrarböden des ersten Feldes genutzt. Validiert wurden die Ergebnisse zum einem mittels einer 20-fachen Kreuzvalidierung mit den LIBS-Daten des ersten Feldes und zum anderen mit den LIBS-Daten des zweiten Feldes.

Ein typisches LIB-Spektrum der untersuchten Bodenproben ist beispielhaft links in Abbildung 8 gezeigt. Das Spektrum weist im detektierten Spektralbereich zahlreiche Linien der im Boden enthaltenen Elemente auf. Bei einer univariate Kalibrierung können jedoch nur einige Linien berücksichtigt werden. Deshalb muss eine Auswahl der detektierten Linien getroffen werden. Diese müssen die oben aufgeführten Anforderungen erfüllen, damit eine stabile Kalibrierung erzeugt

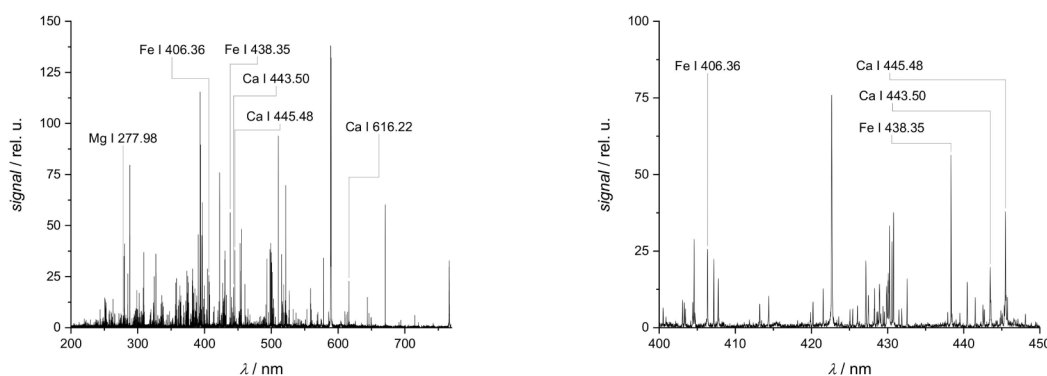


Abbildung 8: Links: Typisches LIB-Spektrum einer untersuchten Bodenprobe im untersuchten Wellenlängenbereich 200 nm bis 770 nm. Markiert sind die Linien der Elemente Calcium, Eisen und Magnesium, die zur Evaluierung der Kalibrierungsstrategie herangezogen wurden. Rechts: Ausschnitt des LIB-Spektrums im Wellenlängenbereich von 400 nm bis 450 nm.

werden kann. In Abbildung 8 rechts ist erkennbar, dass die markierten Linien frei von Überlappungen sind. Als für eine Kalibrierung geeignet, wurden drei Spektrallinien des Elements Calcium (443,50 nm, 445,48 nm, 616,22 nm) sowie zwei Linien des Elements Eisen (406,36 nm, 438,35 nm) und eine Linie

des Elements Magnesium (277,98 nm) ausgewählt. Anhand dieser Linien erfolgte die Evaluierung der drei Kalibrierungsstrategien.

Für Calcium haben sich alle drei getesteten Kalibrierungsstrategien im Hinblick auf die Anwendbarkeit auf das zweite Feld als erfolgreich erwiesen. Neben der multivariaten Methode und der univariaten Referenzmethode zeigt auch die Kalibrierung basierend auf der Standardaddition gute Ergebnisse für die Anwendung auf verschiedenen Feldern. Im Gegensatz dazu ist die Standardadditionsmethode für die Elemente Eisen und Magnesium bei der Anwendbarkeit auf das zweite Feld limitiert. Ursache hierfür sind die Matrixeffekte, die bei dieser Methode nicht ausreichend berücksichtigt werden. Dies führt bereits bei der Anwendung der Kalibrierung auf dem ersten Feld zu Abweichungen der vorhergesagten Gehalte von den Referenzwerten. Auch bei der Anwendung auf das zweite Feld sind diese Abweichungen vorhanden. Somit ist diese Kalibrierungsstrategie lediglich für eine grobe Abschätzung der Elementgehalte geeignet. Der Ansatz der univariaten Referenzmethode und der multivariaten Analyse mittels PLSR erzielen dahingegen auch für die Elemente Eisen und Magnesium gute Ergebnisse. In beiden Fällen werden größere Datenmengen in die Kalibrierung einbezogen, was die bessere Berücksichtigung der Matrixeffekte und damit eine verbesserte Vorhersage der Elementgehalte zur Folge hat. Die PLSR liefert für alle untersuchten Elemente ähnliche Resultate. Allerdings weichen die Vorhersagewerte mit zunehmender Konzentration stärker von den Referenzwerten ab. Insgesamt werden mittels der PLSR die geeignetsten Kalibrationskurven erhalten. Im Vergleich zur univariaten Referenzmethoden leisten sie aber keine substanzielle Verbesserung der Vorhersage. Die Ergebnisse, die durch die univariate Referenzmethode erzielt werden, sind insbesondere für Calcium sehr gut geeignet. Für Eisen und Magnesium werden zwar schlechtere Resultate erhalten, sie sind aber für eine Kalibrierung dennoch gut geeignet.

Die Evaluierung der drei Kalibrierungsstrategien hat gezeigt, dass die Matrixeffekte die Kalibrierung erheblich beeinflussen. Um die Effekte bestmöglich zu berücksichtigen und damit die negativen Einflüsse auf die Kalibrierung zu minimieren, ist es von Vorteil, wenn die Proben für die Kalibrierung aus dem Pool der Feldproben stammen. Diese müssen jedoch mittels einer Laboranalytik referenziert werden. Dadurch ist die Probennahme für die Erstellung des lokal abgestimmten LIBS-Kalibriermodells, der dreidimensionalen Bestimmung der Elementgehalte mit LIBS, sowie die Laboranalytik für die Erkundung von Lagerstätten weiter relevant. Die LIBS-Messungen vor Ort können jedoch genutzt werden, um eine Auswahl zu treffen, welche Proben im Labor analysiert werden müssen. Dadurch wird die Erzeugung unnötig großer Probenmengen vermieden. Decken diese Proben einen breiten Bereich der Elementgehalte ab, kann auf deren Grundlage eine stabile Kalibrierung erzeugt werden. Diese ist dann wiederum für die Vorhersage der mittels der LIBS gemessenen Gehalte notwendig.

Der Einsatz der LIBS im Feld zum Aufbau einer Kalibrierung, die zur Bestimmung von weiteren Gehalten im Feld geeignet ist, erlaubt eine gezieltere Probennahme und -analyse und trägt somit zu einer umweltschonenden Ressourcennutzung bei. Durch die gezieltere Probennahme würden weniger Bohrungen vorgenommen, diese schneller vor Ort mittels LIBS analysiert werden und somit deutlich weniger Rohstoffe aus der Umwelt entnommen und damit die Natur an der potentiellen Lagerstätte geschont. Gleichzeitig müssen im Labor weniger Proben analysiert werden, was auch einen verminderten Einsatz von Analysechemikalien und -geräten zur Folge hat. Dies führt zu reduzierten Personal- und Laborkosten sowie zu einer zeitlich Ersparnis. Durch dieses Vorgehen können schneller

Entscheidungen über eine potentielle Lagerstätte getroffen werden, wodurch der Rohstoffbedarf rascher gedeckt werden kann.

4.2 Quantifizierung

Die Bestimmung von Erzgehalten ist neben einer stabilen Kalibrierung ein weiterer Schwerpunkt in der Prospektion und Exploration von Lagerstätten mineralischer Rohstoffe. Die ermittelten Rohstoffgehalte erlauben eine Einschätzung der Bauwürdigkeit der potentiellen Lagerstätte sowie der Rentabilität der neuen Mine. Auch die gezielte Ausbeutung bestehender Lagerstätten sollte unter den Aspekten der Nachhaltigkeit und der Umweltschonung betrachtet werden. In bestehenden Lagerstätten könnten dadurch gewinnbringende Erzgänge identifiziert werden. Bereiche der Lagerstätte, die kaum oder keine ergiebigen Erzvorkommen enthalten, würden nicht abgebaut. Somit entfallen schwere Eingriffe in die Natur. Vor allem der hohe Flächenbedarf, bedingt durch den Abbau im Tagebau, aber auch durch Reststoffhalden und die Infrastruktur, könnte so auf ein Notwendiges reduziert werden.¹²

Eine geeignete Quantifizierung könnte zudem zur Qualitätskontrolle der Produktion der mineralischen Rohstoffe nach dem Abbau in der Lagerstätte beitragen. Durch eine Überwachung der Erz- bzw. Konzentrataufbereitung könnten Abweichungen in der Qualität zügig entdeckt und behoben werden. Eine teure und zeitintensive Laboranalytik könnte hier durch den vor-Ort-Einsatz der LIBS in Kombination mit multivariaten Methoden durch eine schnelle Echtzeitanalyse ersetzt werden. Die Umwelt würde geschont, da zum einen keine Proben entnommen und zu einem Labor transportiert werden müssten und zum anderen eine material- und energieintensive Laboranalytik entfallen würde.

Die LIB-Spektren von SEE- und kupferhaltigen Mineralen wurden mit dem Ziel verschiedene Quantifizierungsstrategien für mineralische Rohstoffe zu evaluieren, untersucht. In beiden Fällen wurde bei der Analyse der LIBS-Daten als Kalibrierungsstrategie die univariate Regression basierend auf einer Standardaddition durchgeführt. Die Quantifizierung der SEE-Gehalte in den untersuchten Mineralen erfolgte mittels der multivariaten Intervall-PLS-Regression. Die iPLSR wurde intern durch eine Kreuzvalidierung sowie extern mittels Testdaten des LIBS-Datensatzes validiert. Die quantitative Analyse der LIBS-Daten an kupferhaltigen Proben wurde anhand einer PLS-Regression durchgeführt, nach dem Ausreißer mittels der robusten Hauptkomponentenanalyse (ROBPCA) ermittelt wurden. Die PLSR wurde intern durch eine 10-fache Kreuzvalidierung validiert. Eine entsprechende externe Validierung erfolgte klassisch durch die Anwendung des entwickelten Modells auf einen bisher nicht in das Modell einbezogenen Testdatensatz der LIBS-Daten. Außerdem wurde die PLSR an den Spektren der synthetischen Proben durchgeführt und diese mittels der Feldprobenspektren validiert.

Wie schon das LIB-Spektrum des Bodens (Abbildung 8) zeigt, sind LIB-Spektren typischerweise linienreich. Zusätzlich zu den relevanten Elementlinien der mineralischen Rohstoffe, treten Linien der umgebenden Begleitgesteine auf. Diese begleitenden Gesteine können die sogenannten Matrixeffekte bedingen, die wiederum einen negativen Einfluss auf die Spektren haben. Die Untersuchungen zur Kalibrierungsstrategie haben bereits verdeutlicht, dass sich die Matrixeffekte erschwerend auf die Auswertung der Spektren auswirken. Dies ist auch bei der Quantifizierung der Fall. Der Einsatz multivariater Methoden kann diese Einflüsse reduzieren. Eine multivariate Methode die in der vorliegenden Arbeit zur Anwendung kam, ist die partielle Regression der kleinsten Quadrate (PLSR). Das Potential dieser multivariaten Methode für die Quantifizierung wurde anhand der LIBS an

Kupfererzen untersucht. Die Kupfererze der Feldproben stammten aus zwei unterschiedlichen Lagerstätten in Zypern und Polen.

Im Gegensatz zur univariaten Regression gehen bei der PLSR nicht nur einzelne Elementlinien, sondern das gesamte Spektrum in die multivariate Analyse ein. Die Vielzahl an Informationen, die durch das Einbeziehen aller Merkmale in die Analyse, erfasst werden, verbessern die Präzision des Modells. Zudem tragen sie dazu bei, dass das Modell gegenüber äußeren Einflüssen robuster wird. Gleichzeitig kann der Einsatz aller Variablen eines Spektrums zu einem sogenannten overfitting, also einer Überanpassung, führen. Eine Überanpassung kann auftreten, wenn im Modell mehr Terme als benötigt oder ein zu komplexer Ansatz verwendet werden. Daraus können zwei Fälle resultieren: das genutzte Modell ist deutlich flexibler als es für die zu untersuchenden Daten sein muss oder das Modell enthält zusätzlich zu den relevanten Variablen auch nicht benötigte Informationen.⁹⁹ Es resultieren Modelle die beispielsweise schlechtere Vorhersagen oder sogar Fehlprognosen liefern. Durch die Einbeziehung nicht benötigter Variablen kann die Merkmalsauswahl des Modells negativ beeinflusst werden. Ein weiterer möglicher nachteiliger Effekt des overfittings ist die schlechtere Übertragbarkeit des Modells.⁹⁹ Trotzdem ist die PLSR eine sehr robuste Methode, die häufig zur Datenanalyse eingesetzt wird.⁸⁷ Um die Gefahr einer Überanpassung zu reduzieren und die Aussagefähigkeit der PLSR zu erhöhen, kann die Eliminierung von Ausreißern hilfreich sein. Eine Möglichkeit Ausreißer zu ermitteln ist die robuste Hauptkomponentenanalyse (ROBPCA). Durch die ROBPCA können in den zu untersuchenden Daten Spektren ermittelt werden, die durch große Abweichungen von der Gesamtheit der Spektren als Ausreißer klassifiziert werden. Diese können aus dem der Analyse zugrundeliegenden Datensatz entfernt werden und tragen nicht zum Regressionsmodell bei. Die ROBPCA führte in nahezu allen Fällen zu einer Verbesserung der PLSR-Ergebnisse im Bezug auf die Linearität der kreuzvalidierten Daten. Insbesondere bei den am Echelle-Spektrometer gemessenen LIB-Spektren konnten so PLSR-Modelle erhalten werden, die sich durch eine hohe Linearität auszeichnen. Dies gilt sowohl für die kreuzvalidierten als auch für die mit Testdaten validierten LIBS-Daten. Die LIBS-Daten, die mittels des Handheld-Spektrometers gewonnen wurden, zeigen ähnliche Ergebnisse. Allerdings ist die Linearität geringer als bei den am Echelle-Spektrometer gemessenen Daten. Eine Ausnahme bildete in beiden Fällen der Datensatz der Feldproben aus Zypern für die nur ungenügende Regressionen erzielt werden konnten. Ursächlich hierfür könnten die Effekte der umgebenden Matrices der Proben sein. Der Großteil der zyprischen Proben stammt aus einer stillgelegten Mine. Die Feldproben waren stärker der Verwitterung und lokalen Einflüssen ausgesetzt, als die Feldproben aus aktiven Minen in Zypern und Polen. Für die so durchgeführte PLSR müssen die Kupfergehalte der Feldproben bekannt sein, die zur Entwicklung des multivariaten Modells herangezogen werden. Damit ist, im Rahmen der Regressionsmodellentwicklung und -validierung, eine Probennahme und -analytik verbunden. Um zu prüfen, ob auf die Laboranalytik von Feldproben gänzlich verzichtet werden könnte, wurde eine weitere Möglichkeit der Modellvalidierung untersucht. Dafür wurde das Modell der PLSR anhand synthetischer Proben entwickelt und mittels der Feldproben validiert. Hier zeigte sich, dass lineare Regressionen mit höheren Bestimmungskoeffizienten erreicht werden können, wenn dem Trainingsdatensatz zur Entwicklung des Modells bereits einige referenzierte Feldproben zugesetzt werden. Ab etwa einem Drittel Feldproben im Trainingsdatensatz, kann ein stabiles Regressionsmodell mit einer guten Linearität entwickelt werden. Dies deckt sich mit den Ergebnissen der Untersuchungen zur Kalibrierungsstrategie. Es bedeutet zwar, wie im Fall der Kalibrierung, dass nicht vollständig auf die Probenentnahme und -analytik verzichtet werden kann, sie kann jedoch auf

ein Mindestmaß an relevanten Proben reduziert werden und trägt so zu einer Schonung der Umwelt und der Ressourcen bei.

Bei der PLSR als multivariates Quantifizierungsmodell hat sich gezeigt, dass die Matrixeffekt weiterhin negativ auf die Spektren und damit auf die Quantifizierung wirken. Eine weitere multivariate Analyseverfahren mit der einerseits der Einfluss dieser Effekte weiter reduziert und andererseits das Risiko der Überanpassung minimiert werden kann, ist die Intervall-PLSR (*i*PLSR). Hierbei wird durch eine Variablenselektion das Gesamtspektrum in gleichgroße spektrale Intervalle geteilt. Das resultierende Teilspektrum, das in die PLSR-Analyse einbezogen wird, enthält die relevanten Informationen, wie z.B. Elementlinien. Durch dieses Vorgehen wird die Analyse auf wichtige spektrale Bereiche fokussiert und kann beispielsweise Interferenzen aus anderen Bereichen eliminieren. Gleichzeitig sinkt durch die geringere Anzahl der Variablen, die für Entwicklung des Regressionsmodells zur Verfügung stehen, das Potential für eine Überanpassung. In der vorliegenden Arbeit erfolgte die Evaluierung der *i*PLSR-Methode anhand SEE-haltiger Böden und Gesteine. Es wurden synthetische und Feldproben untersucht. Für beide Probensätze konnten Regressionen mit einer hohen Linearität erhalten werden, wobei die Feldproben eine etwas geringere Linearität der Regression im Bezug auf den Bestimmungskoeffizienten aufwiesen. Ursächlich hierfür ist die umgebende Matrix der SEE-Minerale, die durch Umwelteinflüsse heterogener ist, als es bei den synthetischen Proben der Fall ist. Abbildung 9 zeigt die Zusammenhänge zwischen der Anzahl der Intervalle und der Intervallbreite, sowie der Anzahl der Komponenten für die Feldproben und synthetischen Proben in Gestein und Boden.

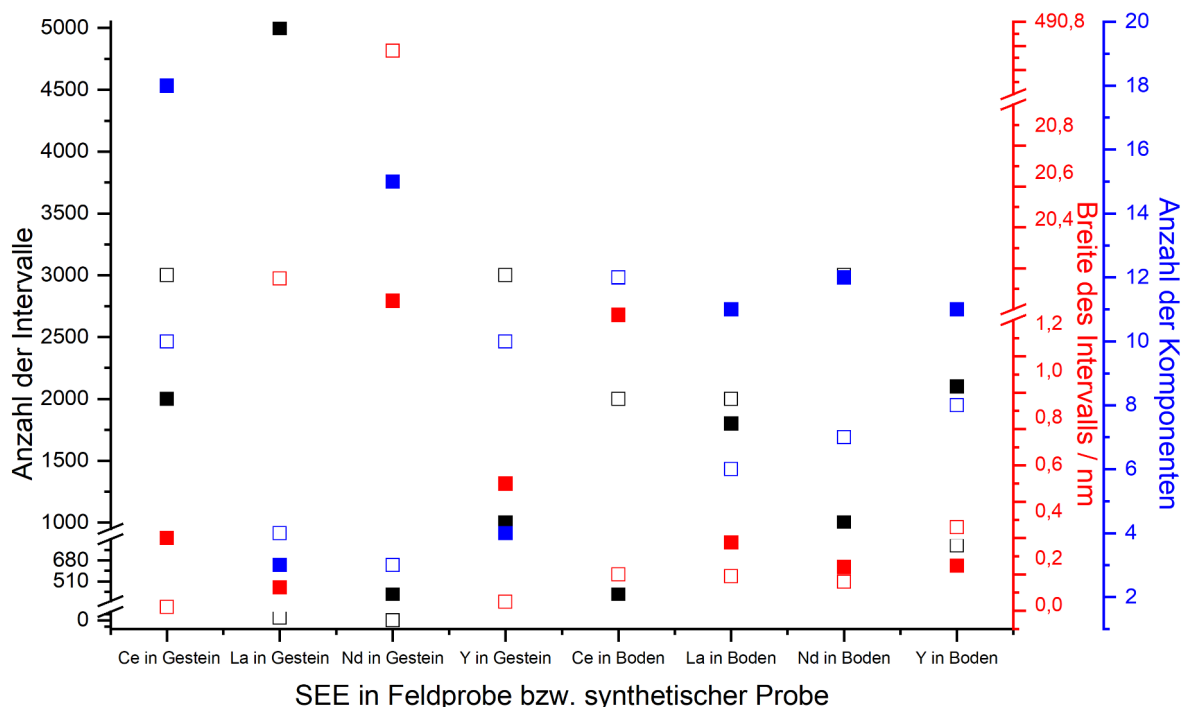


Abbildung 9: Überblick der *i*PLSR-Ergebnisse der SEE-haltigen Gesteine und Böden in synthetischen (offene Quadrate) bzw. Feldproben (gefüllte Quadrate) in Abhängigkeit von der Anzahl der Intervalle (schwarz), der Intervallbreite (rot) und der Anzahl der Komponenten (blau).

Für die synthetischen Proben (offene Quadrate) zeigt sich im Vergleich zu den Feldproben (gefüllte Quadrate), dass die Anzahl der Intervalle stärker schwankt. Bei den Feldproben rangiert die

Intervallanzahl bis auf eine Ausnahme in ähnlichen Größenbereichen. Dies führt auch zu nahezu gleichbleibenden Intervallbreiten, die alle schmal sind. Die Breiten der Intervalle bei den synthetischen Proben sind hingegen deutlich verschiedener. Es treten sehr schmale, aber auch sehr breite Intervalle auf. Die spektrale Breite der Teilspektren kann jedoch problematisch sein. Werden die Bereiche sehr schmal, liegt nahezu eine automatisierte UVR vor, in die nur wenige Merkmale des Spektrums eingehen, häufig nur die zu analysierenden Elementlinien. Dies ist vor allem bei den Feldproben der Fall. Werden die Bereiche dahingegen sehr breit gewählt, wie sie bei den synthetischen Proben auftreten, liegt eine klassische PLSR, in die zahlreiche Merkmale des Spektrums eingehen, vor. Hier besteht wiederum die Gefahr der Überanpassung. Zu hohe Komponentenanzahlen können ebenfalls zu einer Überanpassung, aber auch zu Fehlern der Vorhersage führen. Vor allem die Feldproben weisen zum Teil sehr hohe Komponentenanzahlen auf. Aber auch für die synthetischen Proben werden hohe Anzahlen genutzt. Dennoch zeigt sich das Potential des Regressionsmodells für die Quantifizierung mineralischer Rohstoffe im Feld.

Die zur Quantifizierung von mineralischen Rohstoffen genutzten multivariaten Modelle PLSR und iPLSR resultieren in Regressionsmodellen mit einer hohen Linearität. Es ist mit beiden Analysemethoden möglich die mit LIBS erhalten spektroskopischen Daten zu beschreiben und die Gehalte der mineralische Rohstoffe zu quantifizieren. Aufgrund der Möglichkeit einer Überanpassung bei beiden Modellen, ist ein vorsichtiger Umgang, vor allem im Bezug auf die Qualität, Art und Menge der zugrunde liegenden LIBS-Daten erforderlich. Der Einsatz der multivariaten Methoden zur Quantifizierung der Rohstoffgehalte ermöglicht einen schonenderen Umgang mit den natürlichen Ressourcen. Einerseits kann für eine Quantifizierung mittels PLSR die Probennahme und -analytik auf ein notwendiges Mindestmaß reduziert werden und andererseits ermöglicht die Mobilität der LIBS-Technik Untersuchungen vor Ort. Diese führen durch die computergestützte Sofortanalyse mittels multivariater Methoden unmittelbar zu belastbaren Ergebnisse, die schnelle Entscheidungen über potentielle Lagerstätten, neue Erzgänge in bestehenden Lagerstätten oder die Qualität in der Zwischen- und Endprodukte in der Rohstoffproduktion und Verarbeitung erlauben. Dabei kann auf eine aufwändige Probennahme und teure Laboranalytik verzichtet werden. Dies resultiert in einer Schonung der Ressourcen Umwelt, Personal, Zeit, Chemikalien, Energie und Kosten. Zudem bietet die Quantifizierung vor Ort die Möglichkeit den Abbau der Rohstofflagerstätten und eine langfristige Nachnutzung detailliert zu planen.

4.3 Klassifizierung

Für die Bewertung des Potentials einer bestehenden oder zu erkundenden Rohstofflagerstätte muss eine engmaschige und großflächige Analyse der vorliegenden Gesteine, Minerale und Böden erfolgen. Es werden zahlreiche lokale Messpunkte benötigt, um ein möglichst genaues Abbild der Elementverteilung vor Ort zu erhalten. In der Regel werden diese Messpunkte im Hinblick auf mehrere Komponenten analytisch untersucht. Dies führt zu Datensätzen von enormer Größe und Komplexität. Um die Gesamtheit der in der Lagerstätte erhobenen Daten, also die Beziehungen jeder Probe zu jeder anderen im Datensatz, bewerten zu können, reichen quantitative Methoden nicht aus. Besser geeignet sind unüberwachten und überwachten Methoden der explorativen Datenanalyse, die Korrelationen in multivariaten Datensätzen identifizieren und die untersuchten Proben aufgrund ihrer spezifischen Eigenschaften klassifizieren können. Zudem können diese Methoden Informationen über Trends in den Daten liefern. Die ermittelten Beziehungen innerhalb eines Datensatzes können dann für eine anschließende Quantifizierung genutzt werden. Zu den Techniken der unüberwachten

explorativen Datenanalyse gehört unter anderem die Hauptkomponentenanalyse, welche Strukturen und Regelmäßigkeiten in multivariaten Datensätzen erkennen kann. Sie ist allerdings nicht geeignet, um die Zugehörigkeit einer unbekannt Probe zu einer identifizierten Klasse zu bestimmen. Dafür sind die überwachten Methoden der Mustererkennung entwickelt worden. Mit ihrer Hilfe können unbekannte Proben definierten Gruppen zugeordnet werden. Diese Zuordnung erlaubt es spezifische Eigenschaften der Proben abzuleiten.¹⁰⁰ Ein Beispiel für die Techniken der Mustererkennung ist die nicht-negative Tensorfaktorisierung (NTF).

In den vorgestellten Arbeiten erfolgte die Klassifizierung von Böden und Mineralen anhand der Hauptkomponentenanalyse, mit dem Ziel die Matrices in der Probenumgebung hinsichtlich ihrer Herkunft und Zusammensetzung zu charakterisieren. Zusätzlich wurde eine Klassifikation von Kupfermineralen mittels nicht-negative Tensorfaktorisierung durchgeführt. Die Basis dieses Ansatzes bildete die Extraktion von Quellspektren für jedes spektroskopisch untersuchte Mineral unter Verwendung von NTF. Die Kennzeichnung dieser Quellspektren beruhte auf ihrem prozentualen Anteil innerhalb des Datensatzes. In einem Regressionsmodell wurden diese latenten Spektren für die Validierung genutzt.

Die vorhergehenden Untersuchungen zur Kalibrierung und Quantifizierung bei LIB-Spektren haben bereits gezeigt, dass die Probenumgebung einen Einfluss auf die Datenanalyse hat. Eine Möglichkeit diese Einflüsse der Matrices zu untersuchen ist die Hauptkomponentenanalyse, mit der sich Zusammenhänge zwischen den multivariaten Daten in Form von Ähnlichkeiten und Unterschieden beschreiben lassen.¹⁰⁰ Jedes LIB-Spektrum wird hierbei als ein Punkt, im einfachsten Fall, in einem zweidimensionalen Raum dargestellt. Können Muster erkannt und die Spektren den Mustern zugeordnet werden, gruppieren sich die Punkte in Clustern. Die Ausbildung von Clustern zeigt Unterschiede zwischen den Punkten verschiedener Cluster an und erlaubt Rückschlüsse auf die Herkunft und chemische Zusammensetzung der Proben. Die Verteilung der Datenpunkte innerhalb eines Clusters zeigt an, ob die Proben eines Cluster homogen oder heterogen sind und erlaubt damit auch Aussagen über die chemische Zusammensetzung. Je enger die Punkte verteilt sind, desto homogener und chemisch ähnlicher sind die Proben in ihrer Zusammensetzung. Eine weite Verteilung entspricht größeren Unterschieden in der Zusammensetzung. Dies korreliert mit heterogeneren Proben. Die Klassifizierung der untersuchten Agrarböden zeigte die Ausbildung zweier Cluster, in die die Datenpunkte der zwei untersuchten Felder getrennt werden konnten. Die engere Verteilung der Punkte des ersten untersuchten Feldes zeigte, dass diese Proben eine homogenere Zusammensetzung aufwiesen, als die des zweiten Feldes. Ähnliche Ergebnisse konnten bei der Hauptkomponentenanalyse der kupferhaltigen Feldproben aus den Kupferlagerstätten in Zypern und Polen erhalten werden. Die Clusterbildung resultierte auch hier aus den verschiedenen Herkunftsorten und zusätzlich aus den lokal vorliegenden Begleitgesteinen der Proben. Dies führte zu einer zusätzlichen räumlichen Trennung der Cluster in der PCA-Auftragung, die Proben mit dem Gestein Basalt enthielten und von den die in Schiefer vorlagen. Innerhalb der Cluster der Feldproben lag eine unterschiedlich starke Verteilung vor. Diese resultierte vorrangig aus den Umwelteinflüssen, denen die Proben an ihren Herkunftsorten ausgesetzt waren. Die verschiedenen atmosphärischen Bedingungen der Lokalitäten haben unterschiedliche chemische und physikalische Verwitterungen hervorgerufen. Dies zeigt auch die PCA der synthetischen Proben, zu deren Herstellung eine Matrix für alle Proben verwendet wurde. Folglich sind die Datenpunkte der synthetischen Proben enger gruppiert, was auf eine homogenere Zusammensetzung schließen lässt. Die PCA an Böden, wie im Fall der SEE, ist etwas komplizierter, da dieses Probenmaterial neben dem physikalisch verwittertem

Gestein auch aus organischem Material aus der Biosphäre und atmosphärischem Staub besteht. Gleichzeitig hat sich gezeigt, dass die Spezifikationen der verwendeten LIBS-Instrumente wie beispielsweise die spektrale Auflösung, einen Einfluss auf die Verteilung der Datenpunkte in der PCA haben. Eine höhere spektrale Auflösung wie beim Echelle-basierten LIBS-Gerät führte im Vergleich zum Handheld-LIBS-Gerät zu einer deutlichen Separierung der Cluster untereinander und zu weniger überlappenden Datenpunkten innerhalb eines Clusters. Dadurch konnte die Aussagekraft erhöht werden.

Im Gegensatz zur PCA erfolgt mit der nicht-negativen Tensorfaktorisierung eine Identifikation kupferhaltiger Minerale. Herausfordernd ist die Heterogenität der Proben, die auf der unterschiedlichen chemischen Zusammensetzung und der unregelmäßigen Geometrie der Mineralkristalle beruht. Sie erschwert die Klassifizierung, da sie zu Fehlklassifizierungen aufgrund von Unreinheiten oder fehlerhaften Bezeichnungen führen kann. Mit der Kombination aus LIBS und NTF konnten die untersuchten Kupferminerale erfolgreich identifiziert werden. Die NTF-basierte Klassifizierung konnte zudem bessere Ergebnisse erzielen, als typische Klassifizierungsmethoden, wie beispielsweise die lineare Diskriminanzanalyse. Die fotografische Dokumentation der Messpunkte am Mineral erlaubt zusätzlich die Erstellung einer 2D-Karte der geometrischen Verteilung der Minerale im Untersuchungsbereich der jeweiligen Gesteinsproben.

Der Einsatz von Klassifizierungsmethoden, wie PCA und NTF, ermöglicht einerseits die Charakterisierung von Einflüssen der umgebenden Matrices auf die LIB-Spektren und andererseits die Identifikation und Einteilung von Kupfermineralen. Die Methoden sind im Vergleich zu anderen Klassifizierungsmethoden weniger rechenintensiv und benötigen weniger Speicherplatz. Sie sind dadurch für den Einsatz in mobilen Geräten geeignet und erlauben eine schnellere Analytik vor Ort, wodurch wiederum die Probennahme und -analytik reduziert werden kann. Die Mineralidentifikation vor Ort ermöglicht außerdem eine genaue Abbauplanung der Rohstoffe samt Erzaufbereitung und damit einen gezielteren Abbau der Rohstoffe.

5. Zusammenfassung und Ausblick

Zukunftstechnologien und Megatrends in der Energie- und Verkehrswende bedingen eine stetige Nachfrage an Funktionsmetallen wie Kupfer und SEE. Durch diesen Bedarf ist die Bauwürdigkeit von Lagerstätten mineralischer Rohstoffe gestiegen, sodass auch bis dato unattraktive Depots lukrativ werden. Um der Nachfrage gerecht zu werden, kann es sich trotz geringer Erzgehalte und toxischer Begleitelemente lohnen, scheinbar erschöpfte Lagerstätten weiter auszubeuten oder neue Lagerstätten zu erschließen. Der moderne Bergbau muss sich dabei neuen Herausforderungen stellen, insbesondere im Umweltschutz und der sozialen Verantwortung. Das Ziel der vorliegenden Arbeit war die Entwicklung von Strategien zur ressourcenschonenden Prospektion und Exploration neuer Lagerstätten, zur Identifikation neuer Erzgänge in bestehenden Lagerstätten sowie zur Prozessanalytik in der Produktion. Die analytische Aufgabe sollte mittels der laserinduzierten Breakdownspektroskopie in Kombination mit uni- und multivariaten Methoden gelöst werden. Mit der Absicht Kalibrierungs-, Quantifizierungs- und Klassifizierungsstrategien für die LIBS zu evaluieren, wurden spektroskopische Untersuchungen an Böden sowie an kupfer- und SEE-haltigen Mineralen durchgeführt und die erhobenen spektralen Daten mit multivariaten Analysemethoden ausgewertet.

Für die Evaluierung der Kalibrierungsstrategie standen neben synthetischen Proben, die im Labor erzeugt wurden, mittels Laboranalytik referenzierte Daten von Feldproben zur Verfügung, mit denen die Messergebnisse geprüft werden konnten. Die LIBS in Kombination mit der univariaten Referenzmethode zeigte die besten Resultate für die Beispielanalyten Calcium, Eisen und Magnesium. Stammt ein Teil der Proben für die Kalibrierung aus dem Pool der Feldproben, können die Beeinträchtigungen in der LIBS durch die Matrixeffekte reduziert werden.

Mit der PLSR und der *i*PLSR wurden für die Entwicklung einer Quantifizierungsstrategie zwei multivariate Modelle getestet. Beide Methoden zeigen hohe Potentiale die analytische Aufgabe zu lösen. Auch hier hat sich gezeigt, dass die störenden Matrixeffekte durch die Einbeziehung von einigen referenzierten Feldproben minimiert werden können. Dennoch besteht bei beiden Modellen die Gefahr der Überanpassung. Dadurch ist bei beiden Methoden ein sensibler Umgang mit den LIBS-Daten im Bezug auf die Spektrenqualität, -art und -menge zu berücksichtigen.

Für die Klassifizierung wurden zwei Modelle der explorativen Datenanalyse genutzt. Mit der Hauptkomponentenanalyse kam eine Methode zum Einsatz, mit der die Matrixeffekte der Feldproben charakterisiert werden können. Sie ist gut geeignet Strukturen und Ähnlichkeiten in der Datengesamtheit zu erkennen. Dadurch können Gesteine, Minerale und Böden unterschieden, lokale Unterschiede zwischen verschiedenen Minen und innerhalb einer Mine identifiziert und Heterogenitäten der chemischen Zusammensetzung detektiert werden. Das zweite Modell ist die nicht-negative Tensorfaktorisierung. Dies ist eine Methode der Mustererkennung, die die Einteilung der Proben in Klassen bestimmter Eigenschaften erlaubt. Mit ihrer Hilfe können Eigenschaften unbekannter Proben identifiziert werden.

In allen drei untersuchten analytischen Fragestellungen zeigt sich, dass die Matrixeffekte der Proben für die LIBS und die Auswertung herausfordernd sind. Die Evaluierung hat gezeigt, dass es für die Methoden vielversprechender ist, wenn ein Teil der Referenzproben aus dem Pool der Feldproben stammt. Dadurch können die Matrixeffekte in der Modellentwicklung berücksichtigt und die negativen Auswirkungen reduziert werden. Dieser Ansatz bedingt allerdings weiterhin eine

Probennahme und Laboranalytik im Rahmen der Regressionsmodellentwicklung und -validierung. Im Gegensatz zum bisher üblichen Vorgehen bei der Prospektion und Exploration kann auf die Probennahme nahe der Oberfläche weitestgehend verzichtet beziehungsweise auf wenige Proben reduziert werden, die zudem gezielt ermittelt werden können, um so eine größtmögliche Bandbreite an Elementgehalten in der Lagerstätte abzudecken. Die räumliche Elementverteilung innerhalb der Lagerstätte ist weiterhin über Bohrproben bestimmbar. Hier kann die LIBS jedoch zu einer deutlichen Beschleunigung des Analyseverfahrens beitragen, da auf weitere Laboranalytik verzichtet werden kann. Diese Vorgehensweise ist möglich, da die LIBS in Kombination mit der multivariaten Analyse soweit miniaturisiert werden kann, dass die Messungen und Analyse vor Ort durchführbar sind. Insbesondere für die Feldmessungen an bestehenden und neuen Lagerstätten ist die Nutzung des tragbaren Handheld-LIBS-Geräts geeignet. Das Handheld-Gerät bietet zwar eine geringe spektrale Auflösung, allerdings wird aufgrund dessen auch weniger Rechen- und Speicherkapazität und damit Zeit benötigt. Das Echelle-basierte LIBS-Instrument liefert dahingegen hoch aufgelöste Spektren und damit detailliertere Informationen. Mehr Variablen im Spektrum bedeuten aber auch einen höheren Rechenaufwand, den Einsatz von leistungsfähigerer Hardware, mehr Speicherkapazität und höhere Versorgungskosten. Das Instrument ist daher besser für einen stationären Einsatz zum Beispiel bei der Prozessanalytik in der Produktion und Verarbeitung der mineralischen Rohstoffe geeignet. Mit beiden Systemen konnten die zu untersuchenden Analyten sicher detektiert und die erhaltenen Spektren mittels der multivariaten Methoden ausgewertet werden. Vor allem der Einsatz der multivariaten Analyse erlaubte eine schnelle Datenverarbeitung.

Die Kombination der LIBS mit der multivariaten Datenanalyse kann in der Analyse geologischer Proben ein effizientes Werkzeug sein. Durch die Beschleunigung der zeit- und personalaufwändigen Schritte der Prospektion und Exploration können oberflächennahe, potentielle Erzgängen und Lagerstätten schneller und ressourcenschonender identifiziert werden. Gleichzeitig können der Umfang und der Aufwand der Probennahme und -analyse reduziert und dadurch der Bergbauprozess nachhaltiger und effizienter gestaltet werden, um die Rohstoffverknappung auszubremsen. Zudem werden bei den Feldmessungen vor Ort nur geringe Mengen an zu analysierendem Gestein, Mineral oder Boden verbraucht bzw. beschädigt. Die erhaltenen Ergebnisse erlauben schließlich ein gutes Abbild der vorliegenden oberflächennahen Elementgehaltverteilung in der Lagerstätte. Damit können gezielt Bohrungen durchgeführt und die mittels LIBS analysierten Bohrkerne für eine dreidimensionale Kartierung der Lagerstätte genutzt werden. Dies ermöglicht eine detaillierte und zügige Planung für einen gezielten Abbau der mineralischen Rohstoffe. Dadurch können die Ressourcen und damit die Umwelt geschont werden, da beispielsweise der Flächenverbrauch durch Reststoffhalden oder die Infrastruktur minimiert werden kann.

Weiterführende Arbeiten sollten auf eine noch umfangreichere Datenbasis gestellt werden, um die Modelle weiterzuentwickeln und die Vorhersagekraft der Methoden zu erhöhen. Dafür sollten neben anderen mineralischen Rohstoffen auch verschiedene Böden, Gesteine und Agrarböden sowie unterschiedliche Lagerstätten und Agrarflächen berücksichtigt werden. Bei einer global umfassenden Datenlage könnte ein entsprechendes Regressionsmodell überall auf der Welt ohne vorherige Probennahme durch den Benutzer eines mobilen LIBS-Gerätes sofort zur quantitativen Elementbestimmung eingesetzt werden. Zusätzlich sollte an der Übertragbarkeit evaluierter Modelle auf andere geologische Systeme geforscht werden. Für eine sichere Anwendung der Methode im Feld sollte der Fokus auf die Nutzung des Handheld-Instruments gelegt werden. Neben Handheld-Geräten ist auch der Einsatz von ferngesteuerten oder stand-off-Instrumenten denkbar, die die Zugänglichkeit

im Feld erhöhen könnten. Außerdem könnte das Spektrum der evaluierten Modelle durch weitere chemometrische Methoden vergrößert werden, um die Genauigkeit der Vorhersagen insbesondere in der Quantifizierung aber auch in der Klassifizierung zu erhöhen. Idealerweise kann ein Modell entwickelt werden, das lediglich auf synthetischen Proben beruht und somit die Referenzierung von Feldproben im Labor überflüssig macht. Die LIBS-Ergebnisse der synthetischen Proben, die den Auswertungen zugrunde liegen, könnten in einer Datenbank auf den Endgeräten für den Einsatz im Feld zur Verfügung stehen.

Literaturverzeichnis

- 1 Umweltbundesamt. Ressourcenschonung in der Umweltpolitik.
<https://www.umweltbundesamt.de/themen/abfall-ressourcen/ressourcenschonung-in-der-umweltpolitik> (Zugriff am 02.02.2022).
- 2 Lutter, S.; Giljum, S.; Gözet, B.; Wieland, H.; Manstein, C. *Die Nutzung Natürlicher Ressourcen*; Dessau-Roßlau, 2018.
- 3 Umweltbundesamt. Ressourcennutzung und ihre Folgen.
<https://www.umweltbundesamt.de/themen/abfall-ressourcen/ressourcennutzung-ihre-folgen> (Zugriff am 02.02.2022).
- 4 Umweltbundesamt. Rohstoffe als Ressource.
<https://www.umweltbundesamt.de/daten/ressourcen-abfall/rohstoffe-als-ressource> (Zugriff am 02.02.2022).
- 5 Chen, T.; Zhang, T.; Li, H. Applications of Laser-Induced Breakdown Spectroscopy (LIBS) Combined with Machine Learning in Geochemical and Environmental Resources Exploration. *TrAC Trends Anal. Chem.* **2020**, *133*, 116113. <https://doi.org/10.1016/j.trac.2020.116113>.
- 6 Dorner, U. *45 DERA Rohstoffinformationen. Rohstoffrisikobewertung - Kupfer*; 2020.
- 7 Erdmann, M. *Seltene Erden. Informationen zur Nachhaltigkeit*; Hannover, 2021 .
- 8 Schorn, S. et al. Mineralienportrait Kupfer - Eigenschaften.
<https://www.mineralienatlas.de/lexikon/index.php/Mineralienportrait/Kupfer/Eigenschaften> (Zugriff am 09.11.2021).
- 9 Wirtschaftsvereinigung Metalle. Die NE-Metalle. <https://www.wvmetalle.de/die-ne-metalle/> (Zugriff am 09.11.2021).
- 10 Bertau, M.; Müller, A.; Fröhlich, P.; Katzberg, M. *Industrielle Anorganische Chemie*, 4. Auflage.; Wiley: Weinheim, Deutschland, 2013; 260-264. <https://doi.org/10.1002/9783527649563>.
- 11 Steinacker, S.; Rumpold, R.; Antrekowitsch, J. Behandlung Und Verwertung Der Reststoffe Aus Der Kupfer-Metallurgie. In *Mineralische Nebenprodukte und Abfälle 2. Aschen, Schlacken, Stäube und Baurestmassen*; Thomé-Kozmiensky, K. J., Ed.; Thomé-Kozmiensky Verlag GmbH: Neuruppin-Nietwerder, Deutschland, 2015; 545-557.
- 12 Gilsbach, L. *Kupfer. Informationen zur Nachhaltigkeit*; Hannover, 2020.
- 13 Angerer, G.; Mohring, A.; Marscheider-Weidemann, F.; Wietschel, M. *Kupfer für Zukunftstechnologien*; Karlsruhe, 2010.
- 14 Walters, A.; Lusty, P.; Hill, A. *Rare Earth Elements*; Nottingham, Großbritannien, 2011.
- 15 Marscheider-Weidemann, F.; Langkau, S.; Hummen, T.; Erdmann, L.; Tercero Espinoza, L.; Angerer, G.; Marwede, M.; Benecke, S. *28 DERA Rohstoffinformationen. Rohstoffe für Zukunftstechnologien 2016*; Berlin, Deutschland, 2016.

- 16 Chen, T.; Zhang, T.; Li, H. Applications of Laser-Induced Breakdown Spectroscopy (LIBS) Combined with Machine Learning in Geochemical and Environmental Resources Exploration. *TrAC Trends Anal. Chem.* **2020**, *133*, 116113. <https://doi.org/10.1016/j.trac.2020.116113>.
- 17 Neukirchen, F.; Ries, G. *Die Welt der Rohstoffe*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; 13ff., 56f., 206ff., 230ff. <https://doi.org/10.1007/978-3-642-37739-6>.
- 18 Gocht, W. *Wirtschaftsgeologie und Rohstoffpolitik*, 2.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1983; 11ff. <https://doi.org/10.1007/978-3-642-48070-6>.
- 19 Oyedotun, T. D. T. X-Ray Fluorescence (XRF) in the Investigation of the Composition of Earth Materials: A Review and an Overview. *Geol. Ecol. Landscapes* **2018**, *2* (2), 148–154. <https://doi.org/10.1080/24749508.2018.1452459>.
- 20 Lemière, B.; Harmon, R. S. XRF and LIBS for Field Geology. In *Portable Spectroscopy and Spectrometry*; Crocombe, R., Leary, P., Kammrath, B., Eds.; Wiley, 2021; 455-497. <https://doi.org/10.1002/9781119636489>.
- 21 Harmon, R. S.; Lawley, C. J. M.; Watts, J.; Harraden, C. L.; Somers, A. M.; Hark, R. R. Laser-Induced Breakdown Spectroscopy-An Emerging Analytical Tool for Mineral Exploration. *Minerals* **2019**, *9* (12), 1–45. <https://doi.org/10.3390/min9120718>.
- 22 Pořízka, P. Using Laser-Induced Breakdown Spectroscopy (LIBS) for Material Analysis, Dissertation, Brno University of Technology, Brno, Tschechien, 2014.
- 23 Grotzinger, J.; Jordan, T. *Press/Siever Allgemeine Geologie*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2017; 54, 71ff. <https://doi.org/10.1007/978-3-662-48342-8>.
- 24 Sebastian, U. *Gesteinskunde. Ein Leitfaden für Einsteiger und Anwender*, 4. Auflage.; Springer Spektrum, 2018; 6ff. <https://doi.org/https://doi.org/10.1007/978-3-662-55323-7>.
- 25 Stosch, H.-G.; Hollerbach, R.; Eckhardt, J.-D.; Kleinschrodt, R. *Übungen zur Mineral- und Gesteinsbestimmung für Studierende der Geologie und der Mineralogie*. **2008**.
- 26 Strunz, H.; Nickel, E. H. *Strunz Mineralogical Tables*, 9. Auflage.; E. Schweizerbart'sche Verlagsbuchhandlung: Stuttgart, 2001.
- 27 Rösler, H. J. *Lehrbuch Der Mineralogie*, 5. Auflage.; Deutscher Verlag für Grundstoffindustrie: Leipzig, 1991.
- 28 Neukirchen, F. *Von der Kupfersteinzeit zu den Seltenen Erden*; Springer Spektrum: Berlin, Heidelberg, 2016; 21f. <https://doi.org/10.1007/978-3-662-49347-2>.
- 29 Sicius, H. *Kupfergruppe: Elemente der ersten Nebengruppe; essentials*; Springer Fachmedien Wiesbaden: Wiesbaden, 2017; 9ff. <https://doi.org/10.1007/978-3-658-17205-3>.
- 30 British Geological Survey. *Mineral profile: Copper*. MineralsUK 2007.
- 31 Uhlendorff, A. Kupfergewinnung und -produktion. <https://institut-seltene-erden.de/kupfergewinnung-und-produktion/> (Zugriff am 09.11.2021).

- 32 Spektrum Akademischer Verlag. Bauwürdigkeit. <https://www.spektrum.de/lexikon/geowissenschaften/bauwuerdigkeit/1490> (Zugriff am 27.04.2022).
- 33 Geological Survey Department. Troodos Ophiolite. http://www.moa.gov.cy/moa/gsd/gsd.nsf/page47_en/page47_en?OpenDocument (Zugriff am 02.03.2022).
- 34 Siebert, M. Neuzzeitliche Minen auf Zypern. https://homersheimat.de/themen/minen-auf-zypern/minen_zypern.php (Zugriff am 02.03.2022).
- 35 Szamałek, K.; Szuflicki, M.; Malon, A.; Tymiąski, M. *Mineral Resources of Poland*; Warschau, 2017; 68.
- 36 Reinhardt, K. Seltene Erden. *Chemie unserer Zeit* **1984**, 18 (1), 24–34. <https://doi.org/10.1002/ciuz.19840180104>.
- 37 Dushyantha, N.; Batapola, N.; Ilankoon, I. M. S. K.; Rohitha, S.; Premasiri, R.; Abeyasinghe, B.; Ratnayake, N.; Dissanayake, K. The Story of Rare Earth Elements (REEs): Occurrences, Global Distribution, Genesis, Geology, Mineralogy and Global Production. *Ore Geol. Rev.* **2020**, 122 (January), 103521. <https://doi.org/10.1016/j.oregeorev.2020.103521>.
- 38 Balaram, V. Rare Earth Elements: A Review of Applications, Occurrence, Exploration, Analysis, Recycling, and Environmental Impact. *Geosci. Front.* **2019**, 10 (4), 1285–1303. <https://doi.org/10.1016/j.gsf.2018.12.005>.
- 39 Reiners, C. S. Was ist das Seltene an den Seltenen Erden? *Chemie Unserer Zeit* **2001**, 35 (2), 110–115. [https://doi.org/10.1002/1521-3781\(200104\)35:2<110::AID-CIUZ110<3.0.CO;2-T](https://doi.org/10.1002/1521-3781(200104)35:2<110::AID-CIUZ110<3.0.CO;2-T).
- 40 Ramos, S. J.; Dinali, G. S.; Oliveira, C.; Martins, G. C.; Moreira, C. G.; Siqueira, J. O.; Guilherme, L. R. G. Rare Earth Elements in the Soil Environment. *Curr. Pollut. Reports* **2016**, 2 (1), 28–50. <https://doi.org/10.1007/s40726-016-0026-4>.
- 41 Cremers, D. A.; Radziemski, L. J. *History and Fundamentals of LIBS*; 2006; Vol. 9780521852, 1–39. <https://doi.org/10.1017/CBO9780511541261.002>.
- 42 Telle, H. H.; Ureña, Á. G. *Laser Spectroscopy and Laser Imaging*; CRC Press, 2018; 405–426. <https://doi.org/10.1201/9781315156989>.
- 43 Höhse, M. Kombination von Raman- und Laserinduzierter Plasmaspektroskopie (LIBS) für die Mikroanalyse. Dissertation, Humboldt-Universität zu Berlin, Berlin, Deutschland, 2011.
- 44 Moncunill, X. C. Laser-Induced Breakdown Spectroscopy (LIBS): A Potential Quality Tool for Infant Formula Manufacture. Dissertation, Technological University Dublin, Dublin, Irland, 2019, 21–42. <https://doi.org/10.21427/dcm0-fn22>.
- 45 NASA Science Mars Exploration Program. Mars Curiosity Rover. <https://mars.nasa.gov/msl/spacecraft/instruments/chemcam/> (Zugriff am 28.01.2022).
- 46 NASA Science Mars Exploration Program. SuperCam. <https://mars.nasa.gov/mars2020/spacecraft/instruments/supercam/> (Zugriff am 28.01.2022).

- 47 Zhu, X.; Xu, T.; Lin, Q.; Duan, Y. Technical Development of Raman Spectroscopy: From Instrumental to Advanced Combined Technologies. *Appl. Spectrosc. Rev.* **2014**, *49* (1), 64–82. <https://doi.org/10.1080/05704928.2013.798801>.
- 48 Ankerhold, G.; Kohns, P.; Beresko, C. Berührungsfreie Elementanalyse mit Lasern. Ein Laserspektroskopisches Verfahren mit einem breiten Anwendungsspektrum. *GIT Labor-Fachzeitschrift* **2012**, *9*, 654–656.
- 49 SciAps. User Guide SciAps Z-200 and Z-300 Handheld LIBS Analyzers. SciAps.
- 50 Rai, V. N.; Thakur, S. N. *Physics and Dynamics of Plasma in Laser-Induced Breakdown Spectroscopy*, 2nd ed.; Elsevier B.V., 2020. <https://doi.org/10.1016/b978-0-12-818829-3.00004-6>.
- 51 Thakur, S. N.; Singh, J. P. Fundamentals of Laser Induced Breakdown Spectroscopy. *Laser-Induced Break. Spectrosc.* **2007**, 3–21. <https://doi.org/10.1016/B978-044451734-0.50004-1>.
- 52 Bohling, C.; John, A.; Cordts, L.; Gmbh, S. Sensoren auf Basis der Laserinduzierten Breakdown-Spektroskopie (LIBS) für industrielle Anwendungen. www.secopta.de.
- 53 Gornushkin, I. B. Plasma Fundamentals and Diagnostics. Vortrag, 2018.
- 54 Beresko, C. Laserinduzierte Plasmaspektroskopie (LIBS) als bildgebendes Verfahren zur ortsaufgelösten topographischen Elementanalyse. Dissertation, Universität Koblenz-Landau, Mainz, Deutschland, 2016, 9-23.
- 55 Kramida, A.; Ralchenko, Y.; Reader, J.; Team, N. A. NIST Atomic Spectra Database. National Institute of Standards and Technology: Gaithersburg, MD 2021. <https://doi.org/https://doi.org/10.18434/T4W30F>.
- 56 Winefordner, J. D.; Gornushkin, I. B.; Correll, T.; Gibb, E.; Smith, B. W.; Omenetto, N. Comparing Several Atomic Spectrometric Methods to the Super Stars: Special Emphasis on Laser Induced Breakdown Spectrometry, LIBS, a Future Super Star. *J. Anal. At. Spectrom.* **2004**, *19* (9), 1061–1083. <https://doi.org/10.1039/b400355c>.
- 57 Popov, A. M.; Zaytsev, S. M.; Seliverstova, I. V.; Zakuskin, A. S.; Labutin, T. A. Matrix Effects on Laser-Induced Plasma Parameters for Soils and Ores. *Spectrochim. Acta - Part B At. Spectrosc.* **2018**, *148*, 205–210. <https://doi.org/10.1016/j.sab.2018.07.005>.
- 58 Segnini, A.; Xavier, A. A. P.; Otaviani-Junior, P. L.; Ferreira, E. C.; Watanabe, A. M.; Sperança, M. A.; Nicolodelli, G.; Villas-Boas, P. R.; Oliveira, P. P. A.; Milori, D. M. B. P. Physical and Chemical Matrix Effects in Soil Carbon Quantification Using Laser-Induced Breakdown Spectroscopy. *Am. J. Anal. Chem.* **2014**, *05* (11), 722–729. <https://doi.org/10.4236/ajac.2014.511080>.
- 59 Takahashi, T.; Thornton, B. Quantitative Methods for Compensation of Matrix Effects and Self-Absorption in LIBS Signals of Solids. *Spectrochim. Acta - Part B At. Spectrosc.* **2017**, *138*, 31–42. <https://doi.org/10.1016/j.sab.2017.09.010>.
- 60 Merk, S. Plasmadiagnostik und Quantifizierung für die Laserinduzierte Plasmaspektroskopie. Dissertation, Humboldt-Universität zu Berlin, Berlin, Deutschland, 2012.

- 61 Gornushkin, I. B.; Shabanov, S. V.; Merk, S.; Tognoni, E.; Panne, U. Effects of Non-Uniformity of Laser Induced Plasma on Plasma Temperature and Concentrations Determined by the Boltzmann Plot Method: Implications from Plasma Modeling. *J. Anal. At. Spectrom.* **2010**, *25* (10), 1643. <https://doi.org/10.1039/c0ja00016g>.
- 62 Safi, A.; Tavassoli, S. H.; Cristoforetti, G.; Tognoni, E.; Campanella, B.; Legnaioli, S.; Pagnotta, S.; Poggialini, F.; Palleschi, V. Exploiting Self-Absorption for Plasma Characterization in Laser-Induced Breakdown Spectroscopy Experiments: A Comparison of Two Recent Approaches. *Anal. Chem.* **2019**, *91* (13), 8595–8601. <https://doi.org/10.1021/acs.analchem.9b01885>.
- 63 Griem, H. *Spectral Line Broadening by Plasmas*; Academic Press: New York, 1974.
- 64 *Laser-Induced Breakdown Spectroscopy*; Musazzi, S., Perini, U., Eds.; Springer Series in Optical Sciences; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; Vol. 182 . <https://doi.org/10.1007/978-3-642-45085-3>.
- 65 SPECTRO Analytical Instruments GmbH. Welche Spektrometer- Optiktechnologie bietet die bessere Leistung – Echelle oder ORCA? SPECTRO Analytical Instruments GmbH 2017.
- 66 Eberhard Karls Universität Tübingen. Institut für Astronomie & Astrophysik. Das Echelle-Spektrometer. <https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/physik/institute/astronomie-und-astrophysik/astronomie-hea/forschung/abgeschlossene-projekte/orfeus/echelle-spektrometer/> (Zugriff am 15.02.2022).
- 67 Spektrum. Echellegitter. Spektrum Akademischer Verlag, Heidelberg. <https://www.spektrum.de/lexikon/optik/echellegitter/741> (Zugriff am 15.02.2022).
- 68 Maier. Gitterspektrometer - Gittermonochromator. FH München: München. http://dodo.fb06.fh-muenchen.de/maier/PhysikPH4T/Blaetter/40_Beugung2_BA.pdf
- 69 Becker-Ross, H.; Florek, S. V. Echelle Spectrometers and Charge-Coupled Devices. *Spectrochim. Acta - Part B At. Spectrosc.* **1997**, *52* (9–10), 1367–1375. [https://doi.org/10.1016/s0584-8547\(97\)00024-4](https://doi.org/10.1016/s0584-8547(97)00024-4).
- 70 Böttcher, F.; Eberlein, J.; Ermer, J.; Könemann, A.; Künzle, J.; Limberger, M. *Kernprozesse im Pharmalabor. GMP-Konforme Analytik von der Probenahme bis zur Dokumentation*, 1. Auflage.; Thomas Peither, M. & P. A., Ed.; Maas & Peither AG – GMP-Verlag: Schopfheim (Germany), 2019; 3f.
- 71 Noll, R.; Fricke-Begemann, C.; Connemann, S.; Meinhardt, C.; Sturm, V. LIBS Analyses for Industrial Applications-an Overview of Developments from 2014 to 2018. *J. Anal. At. Spectrom.* **2018**, *33* (6), 945–956. <https://doi.org/10.1039/c8ja00076j>.
- 72 Gros, S. Multivariate Korrektur des Temperatureinflusses in der NIR-spektroskopischen Materialfeuchtebestimmung. Dissertation, Georg-August-Universität zu Göttingen, Göttingen, Deutschland **2009**, 49–50.
- 73 Handl, A.; Kuhlenkasper, T. *Multivariate Analysemethoden. Theorie und Praxis mit R*, 3. Auflage; Springer Spektrum, 2017; 15ff., 25ff. <https://doi.org/10.1007/978-3-662-54754-0>.

- 74 Sandilands, D. Univariate Analysis. In *Encyclopedia of Quality of Life and Well-Being Research*; Springer Netherlands: Dordrecht, 2014; 6815ff. https://doi.org/10.1007/978-94-007-0753-5_3108.
- 75 Schneider, A.; Hommel, G.; Blettner, M. Lineare Regressionsanalyse - Teil 14 Der Serie Zur Bewertung Wissenschaftlicher Publikationen. *Dtsch. Arztebl.* **2010**, *107* (44), 776–782. <https://doi.org/10.3238/arztebl.2010.0776>.
- 76 The Mathworks, I. Coefficient of Determination (R-Squared). <https://de.mathworks.com/help/stats/coefficient-of-determination-r-squared.html> (Zugriff am 21.02.2022).
- 77 The Mathworks, I. Linear Regression. https://de.mathworks.com/help/matlab/data_analysis/linear-regression.html (Zugriff am 21.02.2022).
- 78 Wehrens, R. *Chemometrics with R*; Springer-Verlag Berlin Heidelberg: Heidelberg, Dordrecht, London, New York, 2011; 152f. <https://doi.org/10.1007/978-3-642-17841-2>.
- 79 Unnikrishnan, V. K.; Choudhari, K. S.; Kulkarni, S. D.; Nayak, R.; Kartha, V. B.; Santhosh, C. Analytical Predictive Capabilities of Laser Induced Breakdown Spectroscopy (LIBS) with Principal Component Analysis (PCA) for Plastic Classification. *RSC Adv.* **2013**, *3* (48), 25872. <https://doi.org/10.1039/c3ra44946g>.
- 80 Pořízka, P.; Demidov, A.; Kaiser, J.; Keivanian, J.; Gornushkin, I.; Panne, U.; Riedel, J. Laser-Induced Breakdown Spectroscopy for in Situ Qualitative and Quantitative Analysis of Mineral Ores. *Spectrochim. Acta - Part B At. Spectrosc.* **2014**, *101*, 155–163. <https://doi.org/10.1016/j.sab.2014.08.027>.
- 81 Pearson, K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philos. Mag.* **1901**, *6* (2), 559–572.
- 82 Swarbrick, B. *Multivariate Data Analysis For Dummies*, CAMO Softw.; John Wiley & Sons, Ltd: Chichester, West Sussex, England, 2012; 13ff.
- 83 Frisvad, J. C. Principal Component Analysis. Presentation, Technical University of Denmark, Lyngby, Dänemark.
- 84 Schönbrodt, S. *Maschinelle Lernmethoden für Klassifizierungsprobleme*; BestMasters; Springer Fachmedien Wiesbaden: Wiesbaden, 2019; 5ff. <https://doi.org/10.1007/978-3-658-25137-6>.
- 85 Pořízka, P.; Klus, J.; Képeš, E.; Prochazka, D.; Hahn, D. W.; Kaiser, J. On the Utilization of Principal Component Analysis in Laser-Induced Breakdown Spectroscopy Data Analysis, a Review. *Spectrochim. Acta - Part B At. Spectrosc.* **2018**, *148*, 65–82. <https://doi.org/10.1016/j.sab.2018.05.030>.
- 86 Gauglitz, G.; Dieterle, F.; Reichert, M. Multivariate Datenanalyse – Hauptkomponentenanalyse. http://www.chemgapedia.de/vsengine/vlu/vsc/de/ch/13/vlu/daten/multivariate_datenanalyse_sensor/hauptkomponentenanalyse.vlu.html (Zugriff am 18.02.2022).

- 87 Hubert, M.; Rousseeuw, P. J.; Vanden Branden, K. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics* **2005**, *47* (1), 64–79. <https://doi.org/10.1198/004017004000000563>.
- 88 Goueguel, C. L. Multivariate Outlier Detection in High-Dimensional Spectral Data. 2020. <https://towardsdatascience.com/multivariate-outlier-detection-in-high-dimensional-spectral-data-45878fd0ccb8>.
- 89 Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- 90 James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, 2013; 237f. <https://doi.org/10.1007/978-1-4614-7138-7>.
- 91 Lohninger, H. PLS - Partial Least Squares Regression. http://www.statistics4u.com/fundstat_eng/dd_pls.html (Zugriff am 17.02.2022).
- 92 Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. Interval Partial Least-Squares Regression (i PLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54* (3), 413–419. <https://doi.org/10.1366/0003702001949500>.
- 93 Andersen, C. M.; Bro, R. Variable Selection in Regression-a Tutorial. *J. Chemom.* **2010**, *24* (11–12), 728–737. <https://doi.org/10.1002/cem.1360>.
- 94 MATLAB. Version R2022a (9.12.0.1884302). The MathWorks, Inc.: Natick, Massachusetts, USA, 2022.
- 95 OriginPro. Version 2022 (9.9.0.225) OriginLab Corporation: Northampton, MA, USA, 2022.
- 96 Danzer, K.; Currie, L. A. Guidelines for Calibration in Analytical Chemistry. Part I. Fundamentals and Single Component Calibration (IUPAC Recommendations 1998). *Pure Appl. Chem.* **1998**, *70* (4), 993–1014. <https://doi.org/10.1351/pac199870040993>.
- 97 Riebe, D.; Erler, A.; Brinkmann, P.; Beitz, T.; Löhmannsröben, H.-G.; Gebbers, R. Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture. *Sensors* **2019**, *19* (23), 5244. <https://doi.org/10.3390/s19235244>.
- 98 Zhang, N.; Ou, T.; Wang, M.; Lin, Z.; Lv, C.; Qin, Y.; Li, J.; Yang, H.; Zhao, N.; Zhang, Q. A Brief Review of Calibration-Free Laser-Induced Breakdown Spectroscopy. *Front. Phys.* **2022**, *10*. <https://doi.org/10.3389/fphy.2022.887171>.
- 99 Wu, S.; Zhang, T.; Tang, H.; Wang, K.; Yang, X.; Li, H. Quantitative Analysis of Nonmetal Elements in Steel Using Laser-Induced Breakdown Spectroscopy Combined with Random Forest. *Anal. Methods* **2015**, *7* (6), 2425–2432. <https://doi.org/10.1039/C4AY02601B>.
- 100 Haaland, D. M.; Thomas, E. V. Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* **1988**, *60* (11), 1193–1202. <https://doi.org/10.1021/ac00162a020>.

-
- 101 Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (1), 1–12.
<https://doi.org/10.1021/ci0342472>.
- 102 Lavine, B. K. Clustering and Classification of Analytical Data. In *Encyclopedia of Analytical Chemistry*; John Wiley & Sons, Ltd: Chichester, UK, 2000.
<https://doi.org/10.1002/9780470027318.a5204>.

Abkürzungsverzeichnis

CCD	ladungsgekoppeltes Bauelement (engl.: charge-coupled device)
CF-LIBS	kalibrationsfreie LIBS (engl.: calibration-free LIBS)
CV	Kreuzvalidierung (engl.: cross-validation)
ICCD	verstärktes ladungsgekoppeltes Bauelement (engl.: intensified charge-coupled device)
<i>i</i> PLSR	Intervall-PLSR (engl.: interval partial least squares regression)
LIBS	laserinduzierte Breakdownspektroskopie (engl.: laser-induced breakdown spectroscopy)
LTE	lokales thermodynamisches Gleichgewicht (engl.: local thermodynamic equilibrium)
MCD	minimale Kovarianzdeterminante (engl.: minimum covariance determinant)
Nd:YAG	Neodym-dotierter Yttrium-Aluminium-Granat
NIST	Nationales Institut für Normen und Technologien (engl.: national institute of standards and technology)
NTF	nicht-negative Tensorfaktorisierung (engl.: nonnegative tensor factorisation)
OD	orthogonaler Abstand (engl.: orthogonal distance)
PC	Hauptkomponente (engl.: principal component)
PCA	Hauptkomponentenanalyse (engl.: principal component analysis)
PLSR	partielle Regression der kleinsten Quadrate (engl.: partial least squares regression)
PP	Projektionsnachverfolgung (engl.: projection pursuit)
RMSE	Wurzel der mittleren Fehlerquadratsumme (engl.: root mean square error)
ROBPCA	robuste Hauptkomponentenanalyse (engl.: robust principal component analysis)
SD	Abstand der Scores (engl.: score distance)
SEE	Seltene Erdelemente
SSC	stratiforme sedimentgebundene Kupferlagerstätte (engl.: stratiform sediment-hosted copper deposits)
UV	ultraviolett
UVA	univariate Analyse
UVR	univariate Regression
VMS	vulkanogene massive Sulfidlagerstätte (engl.: volcanic massive sulfide deposit)

Abbildungsverzeichnis

Abbildung 1: Schematischer Aufbau des Messsystems der LIB-Spektroskopie mit 1 Lasersystem, 2 Lichtleitungs- und Sammeloptiken und 3 optischer Detektoreinheit, nach SciAps46.....	17
Abbildung 2: Zeitlicher Verlauf der Plasmabildung in der LIB-Spektroskopie unter Einwirkung eines Laserstrahls auf eine Oberfläche eines Feststoffs, nach Bohling. ⁴⁹	18
Abbildung 3: Zeitliche Entwicklung des Plasmas in Abhängigkeit von der Signalintensität in der LIB-Spektroskopie, nach Telle ³⁹	19
Abbildung 4: Schematische Darstellung des Prinzips der Hauptkomponentenanalyse (PCA) im zweidimensionalen Raum mit den Hauptkomponenten PC1 und PC2 (links) und im Score-Plot (rechts), nach Gauglitz <i>et al.</i> ⁸³	26
Abbildung 5: Originalmatrix X der Hauptkomponentenanalyse mit N Proben und M Variablen zerlegt in die NxK Score-Matrix S, KxM Loadingsmatrix L^T und die NxM Fehlermatrix F, nach Frisvad. ⁸⁰	26
Abbildung 6: Ausreißerkarte der robusten Hauptkomponentenanalyse (ROBPCA). Die Datenpunkte werden im Diagramm in Abhängigkeit vom orthogonalen Abstand (orthogonal distance) und dem Score-Abstand (score distance) aufgetragen. Durch die cutoffs SD und OD resultieren 4 Quadranten, in denen die Punkte klassifiziert werden: I bad leverage-Punkte, II orthogonale Ausreißer, III reguläre Beobachtungen und IV good leverage-Punkte, nach Hubert <i>et al.</i> ⁸⁴	27
Abbildung 7: Prinzip der PLSR. Schematisch dargestellt sind die Matrices der unabhängigen X-Daten und der abhängigen Y-Daten. Beide Matrices sind aus einer Score-Matrix S bzw. U, Loadingsmatrix L^T bzw. M^T und einer Fehlermatrix F bzw. G aufgebaut. Durch eine lineare Regression der Score-Matrices S und U resultiert eine neue Matrix B, die den Zusammenhang beider Matrices wiedergibt, nach Lohninger. ⁸⁸	28
Abbildung 8: Links: Typisches LIB-Spektrum einer untersuchten Bodenprobe im untersuchten Wellenlängenbereich 200 nm bis 770 nm. Markiert sind die Linien der Elemente Calcium, Eisen und Magnesium, die zur Evaluierung der Kalibrierungsstrategie herangezogen wurden. Rechts: Ausschnitt des LIB-Spektrums im Wellenlängenbereich von 400 nm bis 450 nm.....	107
Abbildung 9: Überblick der <i>i</i> PLSR-Ergebnisse der SEE-haltigen Gesteine und Böden in synthetischen (offene Quadrate) bzw. Feldproben (gefüllte Quadrate) in Abhängigkeit von der Anzahl der Intervalle (schwarz), der Intervallbreite (rot) und der Anzahl der Komponenten (blau).....	111

Tabellenverzeichnis

Tabelle 1: Einteilung der Minerale in Mineralklassen nach der Mineralsystematik nach Strunz.23, 24. 13	
Tabelle 2: Einteilung der Lanthanide und Yttrium in leichte und schwere SEE.....	15

Anhang

S.1 Ergänzende Informationen zu den Publikationen

Im Folgenden sind die zusätzlichen Informationen zu den veröffentlichten Publikationen Manuskript 2 (S1.1) und Manuskript 4 (S1.2) aufgeführt.

S1.1 Detektion von Seltenen Erdelementen in Mineralen und Böden mittels laserinduzierter Breakdownspektroskopie und Interval-PLSR (*Manuskript 2*)

„Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS“

Minerals **11** (12), 1379 (2021)

<https://doi.org/10.3390/min11121379>



Supplementary Material

Detection of Rare Earth Elements in Minerals and Soils by Laser-Induced Breakdown Spectroscopy (LIBS) Using Interval PLS

Nina Rethfeldt ¹, Pia Brinkmann ¹, Daniel Riebe ¹, Toralf Beitz ^{1,*}, Nicole Köllner ², Uwe Altenberger ² and Hans-Gerd Löhmannsröben ¹

¹ University of Potsdam, Physical Chemistry, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; rethfeldt@uni-potsdam.de (N.R.); pbrinkma@uni-potsdam.de (P.B.); daniel.riebe@alumni.uni-potsdam.de (D.R.); beitz@uni-potsdam.de (T.B.); loeh@chem.uni-potsdam.de (H.G.L.)

² University of Potsdam, Structural Petrology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany; nicolek@uni-potsdam.de (N.K.); uwe@geo.uni-potsdam.de (U.A.)

* Correspondence: beitz@uni-potsdam.de

In this supplementary material, the concentrations of the synthetic samples in rock and soil are summarized in table form (Table S1), and the reference data of the field samples are given (Table S2).

Table S1. Concentrations of the REE synthetic samples in rock and soil. Each pellet contains only one REE in one concentration.

In Rock				In Soil			
c(Ce)/ppm	c(La)/ppm	c(Nd)/ppm	c(Y)/ppm	c(Ce)/ppm	c(La)/ppm	c(Nd)/ppm	c(Y)/ppm
11	10	10	10	11	10	12	15
45	25	25	25	29	30	28	24
112	60	60	39	41	39	41	41
268	97	99	59	61	62	64	67
749	155	160	89	102	100	104	106
1230	299	301	98	158	157	157	160
1342	397	400	175	239	241	242	235
1310	493	505	288	400	402	398	406
1406	603	597	390	669	670	672	672
1509	699	696	487	1099	1096	1094	996
1611	996	988	1044	1799	1789	1790	1802
2927	1750	1843	2904	2902	2875	2882	2912
8207	4963	5023	4992	4898	4838	4844	4880
13085	8269	8288	8081	7990	7987	7964	8028
23203	15629	23274	22011	12799	13286	13281	13097
39572		39958					

Table S2. Reference data of field samples and loess-containing soil matrix, all in wt% except the REE (in ppm).

Rock Sample	SiO ₂	TiO ₂	ZrO ₂	Cr ₂ O ₃	Al ₂ O ₃	Fe ₂ O ₃	MnO	MgO	CaO	Na ₂ O	K ₂ O	P ₂ O ₅	Ce	La	Nd	Y
01	75.93	0.17	0.01	0.0003	11.06	5.33	0.01	2.31	0.95	2.01	0.77	0.27	47.34	20.04	35.10	407
02	74.91	0.20	0.01	0.0003	11.00	7.16	0.02	3.63	0.54	0.13	0.98	0.47	116.13	64.72	82.22	215
03	51.66	0.20	0.01	0.0007	7.73	11.75	0.04	5.62	10.08	0.93	1.06	7.84	2373.5	926.90	1441.8	1355
04	38.71	0.62	0.002	0.0011	9.88	22.40	0.07	8.65	7.80	1.55	0.47	6.36	3066.6	1221.5	1854.7	2473
05	20.80	0.11	-	0.0040	4.00	62.19	0.02	1.29	3.48	0.90	1.29	2.42	4758.9	2067.0	2075.7	1131
06	77.43	0.17	-	>0.001	11.17	1.51	>0.001	0.74	0.46	7.35	0.61	0.07	20.78	9.31	12.00	40
07	67.82	0.10	-	>0.001	3.75	1.24	0.06	1.14	11.95	2.87	0.13	8.28	2679.1	985.18	1918.5	3014
08	74.69	0.17	-	>0.001	9.47	7.22	0.02	2.87	1.13	1.22	1.16	0.66	738.64	314.26	426.59	450
09	59.58	0.41	-	0.0050	7.86	14.02	>0.001	3.78	1.49	1.37	2.12	3.15	7115.7	3219.4	3486.7	15086
10	27.90	0.49	-	0.0150	7.00	15.30	>0.001	6.40	10.16	6.60	3.36	10.36	33201	14182	16971.8	5137
11	83.98	0.12	-	0.0020	5.57	4.59	0.01	0.75	0.11	0.75	3.40	0.07	240.24	116.84	109.31	36
12	68.51	0.57	-	0.0110	10.88	7.52	0.03	3.12	0.11	0.94	6.52	0.05	63.27	29.94	29.10	22
13	62.30	0.66	0.02	0.0003	17.64	3.98	72.00	1.24	2.32	3.97	5.75	0.21	1217.0	518.51	681.71	1309
14	54.30	0.08	0.89	0.0003	17.76	6.31	0.34	0.13	0.79	11.05	4.31	0.01	2.60	1.44	0.86	13
15	57.13	0.03	0.61	0.0003	19.69	4.98	0.06	0.07	0.48	11.23	3.87	0.01	15.88	7.35	8.77	39
16	50.46	0.04	4.42	-	11.81	1.83	0.17	0.70	9.76	5.14	8.84	0.01	1184.4	468.00	629.79	1094
17	40.99	0.01	1.56	-	17.15	0.52	0.15	1.05	14.97	9.88	3.24	0.01	1546.8	724.63	715.27	1400

18	54.79	0.02	0.19	-	20.51	3.64	0.04	0.08	0.36	15.80	4.16	0.00	52.03	20.55	29.04	94
19	54.02	0.05	1.42	-	16.93	5.11	0.20	0.10	0.74	15.86	3.25	0.02	903.63	383.85	463.73	1176
20	48.81	0.10	2.36	-	16.47	6.17	0.32	0.12	1.25	16.12	4.31	0.01	1538.4	645.22	853.19	2446
21	53.09	0.14	1.60	-	13.61	9.14	0.26	0.14	1.00	16.35	2.58	0.01	1212.5	511.47	677.65	1883
22	47.76	0.06	2.33	-	20.55	4.35	0.18	0.14	1.94	15.23	2.59	0.01	575.51	256.07	319.99	1031
23	60.93	0.07	0.79	-	14.11	6.24	0.12	0.15	0.71	16.05	0.18	0.02	536.56	229.83	305.33	1007
24	5.91	0.27	-	-	3.07	5.81	1.48	10.83	22.27	0.05	0.07	0.03	952.32	510.95	422.25	36
25	8.17	0.26	-	-	4.43	16.13	1.61	7.87	21.80	0.09	0.09	0.05	1070.9	568.99	481.48	40
26	3.14	2.59	-	-	1.62	2.85	1.06	11.07	25.61	0.01	0.28	0.48	698.99	363.26	285.71	61
27	2.96	0.56	-	-	1.14	5.73	0.94	11.39	26.13	0.21	0.10	2.03	808.38	406.25	332.59	97
28	14.00	1.59	-	-	4.40	4.58	1.44	15.00	19.82	0.06	0.02	0.42	1834.4	959.11	684.69	93
29	16.86	1.98	-	-	5.24	10.79	1.61	13.96	16.05	0.15	0.02	0.53	2555.2	1314.2	955.42	123
soil sample	SiO ₂	TiO ₂	ZrO ₂	Cr ₂ O ₃	Al ₂ O ₃	Fe ₂ O ₃	MnO	MgO	CaO	Na ₂ O	K ₂ O	P ₂ O ₅	Ce	La	Nd	Y
01	71.97	0.45	0.02	0.0018	9.08	1.91	0.03	0.76	1.37	1.36	2.61	0.16	45.22	21.91	21.35	25
02	41.39	0.29	0.01	0.0014	5.33	2.83	0.10	0.75	0.76	0.32	0.77	0.54	661.55	296.95	355.68	535
03	49.98	0.35	0.01	0.0021	5.54	9.14	0.34	1.97	2.91	0.42	0.61	3.6	4259.3	1769.8	2434.2	2123
04	13.29	0.14	0.00	0.0008	2.54	1.22	0.05	0.51	0.99	5.00	0.49	0.53	112.59	47.99	65.00	108
05	65.84	0.49	0.02	0.0016	8.72	2.77	0.04	0.79	1.26	1.08	2.32	0.17	204.74	89.94	103.14	114
06	24.03	0.29	4.00	0.0013	4.85	5.95	0.24	1.99	2.29	5.00	1.05	1.01	1144.3	525.97	544.66	966

4 of 4

Minerals 2021, 11, 1379

07	43.40	0.29	1.28	0.0018	11.05	5.40	0.54	0.60	0.96	2.37	2.06	0.55	116.63	47.48	44.22	159
08	73.30	0.41	0.03	0.0014	10.71	2.87	0.09	0.54	1.30	2.26	2.87	0.18	61.24	28.75	26.10	33
09	68.47	0.49	0.02	0.0016	10.07	3.17	0.16	0.69	1.39	1.62	2.61	0.22	59.61	23.00	18.21	18
10	72.54	0.46	0.02	0.0013	10.81	3.14	0.14	0.58	1.23	2.15	2.93	0.17	61.10	22.71	18.31	20
11	3.52	0.05	0.0004	0.0003	0.44	0.20	0.01	0.25	1.00	5.00	0.15	0.11	42.03	22.35	24.09	40
12	3.43	0.06	0.0004	0.0003	1.01	0.45	0.01	0.26	1.53	5.00	0.17	0.15	54.14	19.93	13.40	41
13	68.91	0.49	0.03	0.0009	11.06	3.50	0.20	0.68	1.19	2.15	2.78	0.12	68.74	27.64	24.46	41
14	14.64	1.00	-	-	4.83	35.54	3.79	2.31	4.64	1.35	0.33	0.62	9104.5	4916.6	3596.7	203
15	9.99	1.86	-	-	4.28	30.08	1.44	3.65	11.62	0.23	0.59	7.71	1073.3	568.53	479.70	329
16	21.95	2.09	-	-	6.97	15.94	2.03	6.07	1.04	0.37	0.28	0.38	3132.0	1521.1	1126.1	187
soil matrix	76.15	0.93	-	-	9.77	3.20	0.12	0.71	0.87	1.13	2.43	0.18	-	-	-	-

S1.2 Klassifizierung kupferhaltiger Minerale mittels laserinduzierter Breakdownspektroskopie und multivariater Datenanalyse (*Manuskript 4*)

„Classification of Copper Minerals by Handheld Laser-Induced Breakdown Spectroscopy and Nonnegative Tensor Factorisation“

Sensors **20** (18), 5152 (2020)

<https://doi.org/10.3390/s20185152>



Supplementary data

Table S1. (part 1) Detailed list of 62 minerals used in research. C refers to Class (Table 1). Bold font in names refers to the short names used in figures.

Mineral name	C	Empirical formula	Mineral name	C	Empirical formula
algodonite	2	Cu ₆ As	azurite	5	Cu ₃ (CO ₃) ₂ (OH) ₂
antimonpearceite	2	Ag ₇ Cu ₈ Sb _{1.5} As _{0.5} S ₁₁	malachite	5	Cu ₂ (CO ₃)(OH) ₂
bornite	2	Cu ₅ Fe ²⁺ S ₄	brochantite	7	Cu ₄ (SO ₄)(OH) ₆
bournonite	2	PbCu ₂ Sb ₂ S ₃	copiapite	7	Fe ²⁺ Fe ³⁺ ₄ (SO ₄) ₆ (OH) ₄ •20(H ₂ O)
chalcocopyrite	2	Cu ₂ Fe ²⁺ S ₂	devilline	7	CaCu ₄ (OH) ₆ (SO ₄) ₂ •3(H ₂ O)
colusite	2	Cu _{12.5} V ⁵⁺ As _{1.5} Sb _{0.5} Sn _{0.6} Ge _{0.3} S ₁₆	formacite	7	Pb ₇ Cu ₄ (AsO ₄)(CrO ₄)(OH)
covellite	2	CuS	langite	7	Cu ₄ (SO ₄)(OH) ₆ •2(H ₂ O)
digenite	2	Cu ₉ S ₅	nakauriite	7	Mn ²⁺ Ni ₃ Cu ₄ (SO ₄)(CO ₃)(OH) ₆ •48(H ₂ O)
enargite	2	Cu ₃ As ₄ S ₄	natrochalcite	7	NaCu ₂ (SO ₄) ₂ (OH) ₄ •(H ₂ O)
freibergite	2	Ag ₇ Cu ₈ Fe ²⁺ _{1.5} Sb _{0.5} As _{0.5} S ₁₃	osarizawaite	7	PbCu ₄ Al ₄ (SO ₄) ₂ (OH) ₆
germanite	2	Cu ₂₄ Fe ²⁺ ₄ Ge ₄ S ₃₂	posnjakite	7	Cu ₄ (SO ₄)(OH) ₆ •(H ₂ O)
gladite	2	PbCu ₂ Bi ₂ S ₉	vaquelinite	7	Pb ₇ Cu ₄ (CrO ₄)(PO ₄)(OH)
idaite	2	Cu ₅ Fe ²⁺ S ₆	arthurite	8	Cu ₂ Fe ³⁺ ₂ (AsO ₄) _{0.7} (PO ₄) _{0.2} (SO ₄) _{0.1} O _{1.5} (OH) _{0.5} •4(H ₂ O)
jaskolskiite	2	Pb _{2.2} Cu _{0.2} Sb _{1.2} Bi _{0.6} S ₅	chenevixite	8	Cu ₂ Fe ²⁺ ₂ (AsO ₄) ₂ (OH) ₄ •(H ₂ O)
krupkaite	2	PbCu ₂ Bi ₂ S ₆	clinoclase	8	Cu ₃ (AsO ₄)(OH) ₅
seligmannite	2	PbCu ₂ As ₂ S ₃	comichalcite	8	CaCu ₄ (AsO ₄)(OH)
stannite	2	Cu ₂ Fe ²⁺ Sn ₂ S ₄	cornetite (co0)	8	Cu ₃ (PO ₄)(OH) ₃
stromeyerite	2	CuAgS	cornubite	8	Cu ₃ (AsO ₄) ₂ (OH) ₄ •(H ₂ O)
tetrahedrite	2	Cu ₉ Fe ²⁺ ₃ Sb ₃ S ₁₃	descloizite	8	PbZn ₄ (VO ₄)(OH)
umangite	2	Cu ₅ Se ₂	dufite	8	PbCu ₄ (AsO ₄)(OH)
boleite	3	KAg ⁹⁺ Cu ₂₄ Pb ₂₆ Cl ₆₂ (OH) ₄₈	libethenite	8	Cu ₂ (PO ₄)(OH)
kinoite	3	Ca ₂ Cu ₂ Si ₂ O ₁₀ •2(H ₂ O)	mottramite	8	PbCu ₄ (VO ₄)(OH)
cuprite	4	Cu ₂ O	olivenite	8	Cu ₂ (AsO ₄)(OH)
delafossite	4	CuFe ³⁺ O ₂	parmauite	8	Cu ₉ (AsO ₄) ₂ (SO ₄)(OH) ₁₀ •7(H ₂ O)
tenorite	4	CuO	pseudomalachite	8	Cu ₃ (PO ₄) ₂ (OH)
delafossite	4	CuFe ³⁺ O ₂	turquoise	8	Cu ₃ Al ₄ (PO ₄) ₂ (OH) ₈ •4(H ₂ O)

Table S1. (part 2) Detailed list of 62 minerals used in research. C refers to Class (Table 1). Bold font in names refers to the short names used in figures.

Mineral name	C	Empirical formula	Mineral name	C	Empirical formula
richelsdorfite	8	$\text{Ca}_2\text{Cu}_3\text{Sb}(\text{AsO}_4)_2\text{Cl}(\text{OH})_4 \cdot 6(\text{H}_2\text{O})$	creaseyite	9	$\text{Pb}_2\text{Cu}_2\text{Fe}^{2+}_{1.75}\text{Al}_{0.25}\text{Si}_5\text{O}_{17} \cdot 6(\text{H}_2\text{O})$
tsumebite	8	$\text{Pb}_2\text{Cu}(\text{PO}_4)(\text{SO}_4)(\text{OH})$	diopside	9	$\text{CaSi}_2\text{O}_6(\text{OH})_2$
tyrolite	8	$\text{CaCu}_3(\text{AsO}_4)_2(\text{CO}_3)(\text{OH})_4 \cdot 6(\text{H}_2\text{O})$	halloysite	9	$\text{Al}_2\text{Si}_2\text{O}_7(\text{OH})_4$
ajonite	9	$\text{K}_{2.25}\text{Na}_{1.75}\text{Cu}_{20}\text{Al}_2\text{Si}_{12}\text{O}_{76}(\text{OH})_{16} \cdot 8(\text{H}_2\text{O})$	planchette	9	$\text{Cu}_6\text{Si}_6\text{O}_{22}(\text{OH})_4 \cdot (\text{H}_2\text{O})$
allopmane	9	$(\text{Al}_2\text{O}_3)(\text{SiO}_2)_{1.3} \cdot 2.5(\text{H}_2\text{O})$	vesuvianite	9	$\text{Ca}_{10}\text{Mg}_2\text{Al}_4(\text{Si}_2\text{O}_7)_2(\text{SiO}_3)_2(\text{OH})_4$
chrysocolla	9	$\text{Cu}_{1.75}\text{Al}_{0.25}\text{H}_{1.75}(\text{Si}_2\text{O}_7)(\text{OH})_4 \cdot 0.25(\text{H}_2\text{O})$			

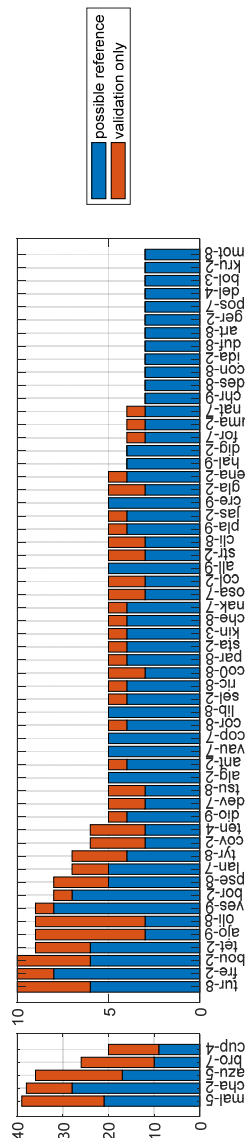


Figure S1. The number of measurement points (MPs) taken per mineral together with their partitions into possible reference and validation only ones.

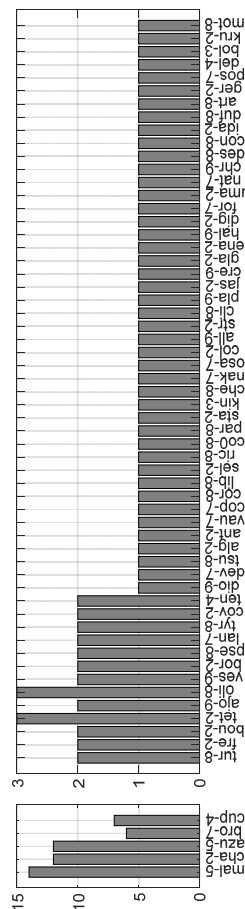


Figure S2. The number of different (exclusive) rock samples on which MPs were taken – listed per mineral.

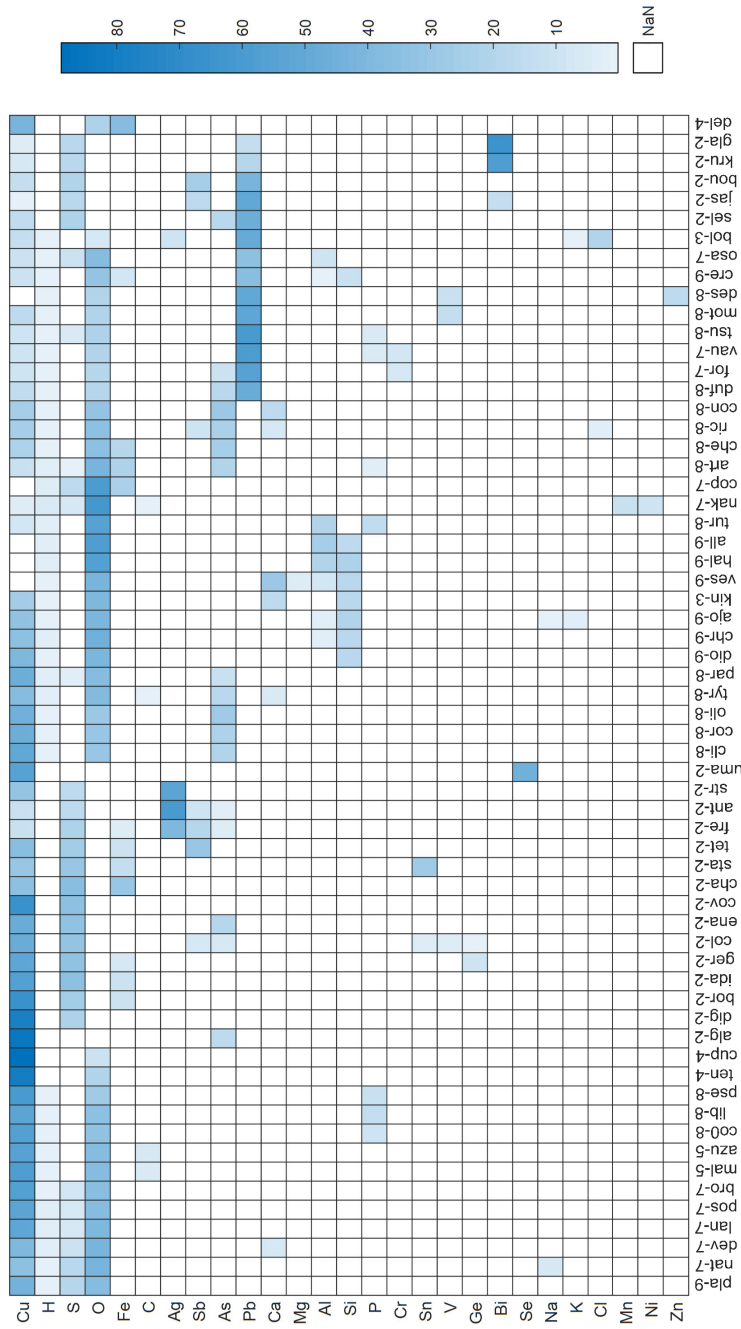


Figure S3. Heatmap representing elements composition among all 62 minerals in accordance to mineral's empirical formula [24]. The sorting from left to right was done based on the shortest N-dimensional distance between each pair of minerals where the first **pla-9** is the closest to the origin of PCA space.

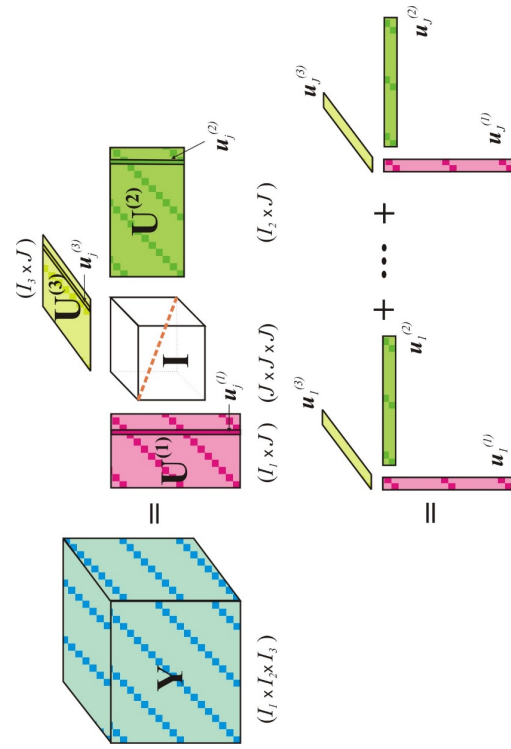


Figure S4. NTF model: a nonnegative 3-way tensor Y is represented by a tensor-matrix product of the latent component matrices $\{U^{(1)}, U^{(2)}, U^{(3)}\}$ with an identity tensor I or equivalently by a sum of J rank-1 tensors obtained by an outer product of the latent components.

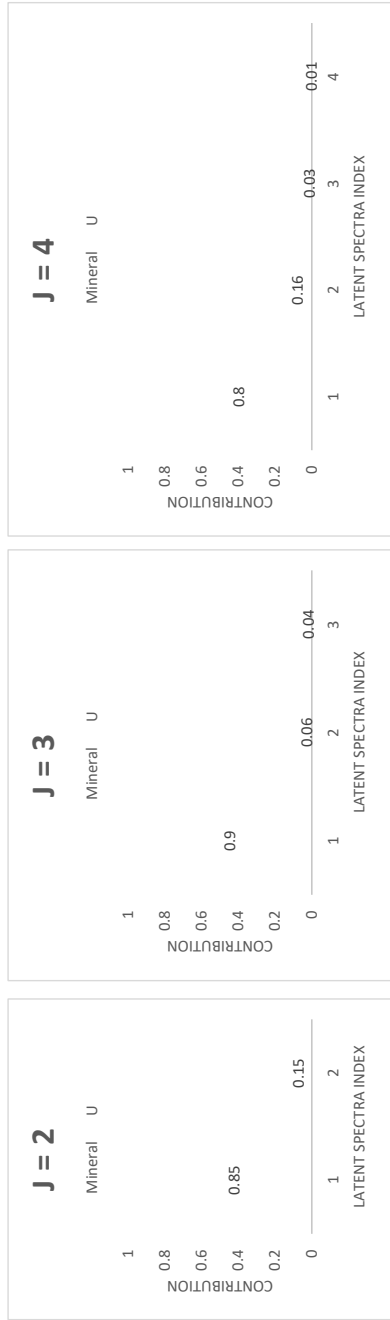


Figure S5. Example of J_m number of latent spectra decision for one m -th object. The L is set to 0.95 and R to 2.0 and maximum latent division $J_m^{(max)}$ to 10:

- First iteration - J_m equals 2. Assuming one spectra always needs to be left as U then the sum of coefficients $\{\alpha_j^{(m)}\}$ of the first K latent spectra (K equals only 1 in that case) covers only 0.85 and the L -condition is not met - J_m is incremented.
- Second iteration - J_m equals 3. Assuming one spectrum always needs to be left as U then the minimum sum of coefficients $\{\alpha_j^{(m)}\}$ of the first K latent spectra that meets L is 0.96 for $K=2$, so the L -condition is met. Spectra 1 and 2 are assigned as mineral while spectrum 3 is assigned as U . Then the second R parameter is checked. The ratio between last mineral (blue) and first U spectrum (orange) is 1.5 which does not meet the R value - hence J_m is incremented. It is worth noting here that the difference between contributions of the coefficients of second spectrum (6 percent) and third spectrum (4 percent) compared to the first spectrum (90 percent) makes it unclear if maybe both second and third spectra are rather U . But in that case, the L -condition is not met, so a given pair of L/R will not work for $J_m = 3$.
- Third iteration - J_m equals to 4. Assuming again that one spectrum always needs to be left as U , then the minimum sum of coefficients $\{\alpha_j^{(m)}\}$ of the first K latent spectra that meets the L minimum contribution is 0.96 for $K=2$, so the L -condition is met. Spectra 1 and 2 are assigned as mineral while spectra 3 and 4 are assigned as U . Then the second R -condition is checked. The ratio between the last mineral (blue) and the first U spectrum (orange) is 5.3 which meets the minimum R -condition = 2.0. The loop is broken and the solution for the m -th object with $J_m = 4$ is given. In such a case, we have two spectra that are assigned as mineral (and later in regression will be labeled as m -th object) and two spectra with minor contribution labeled as U .

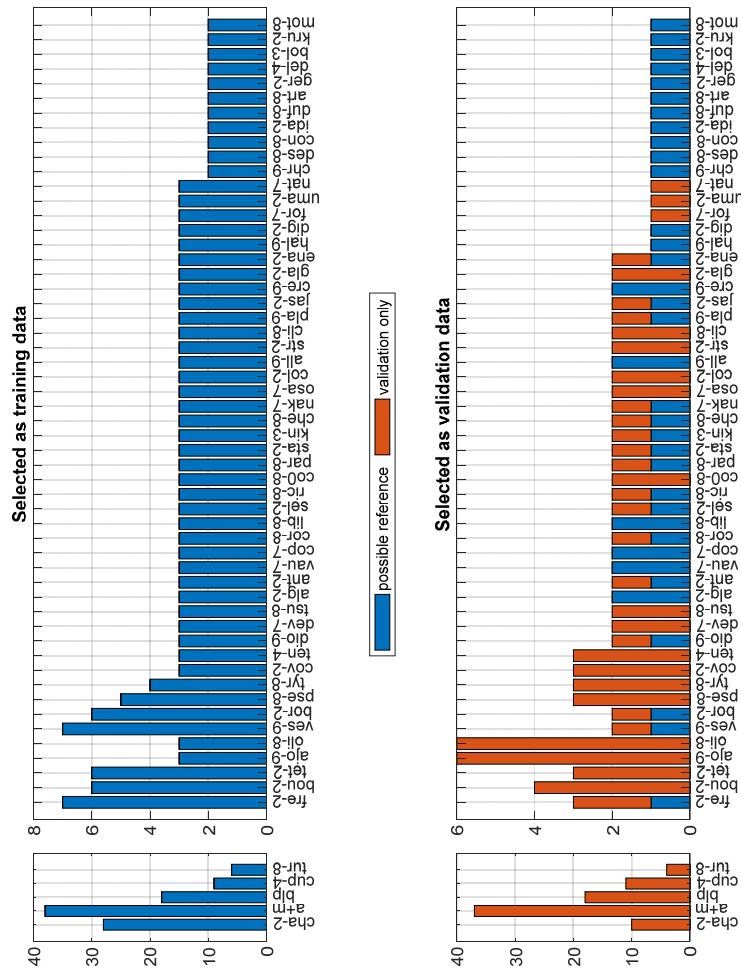


Figure S6. Dataset partitioning of the training and validation data using some of the possible reference MPs as validation data to evenly match the dataset among mineral classes (Section 5.3).

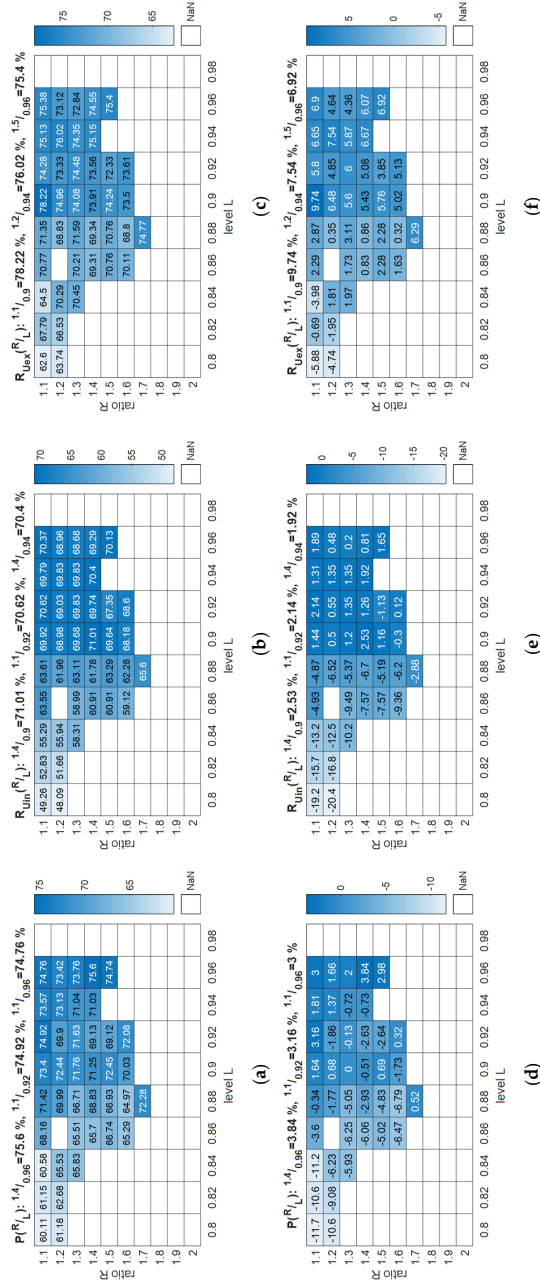


Figure S7. Classification precision and recall output for the parametrisation of the NTF method with R and L variables (validation dataset): **(a)** P – precision scores; **(b)** R_{lim} – recall scores with the U class included as an error; **(c)** R_{lax} – recall scores with the U class excluded from the measure; **(d)** P – precision scores in accordance with the best SVM result from Table 3; **(e)** R_{lim} – recall scores with the U class included as an error in accordance with the best SVM result from Table 3; **(f)** R_{lax} – recall scores with the U class excluded from the measure in accordance with the best SVM result from Table 3.

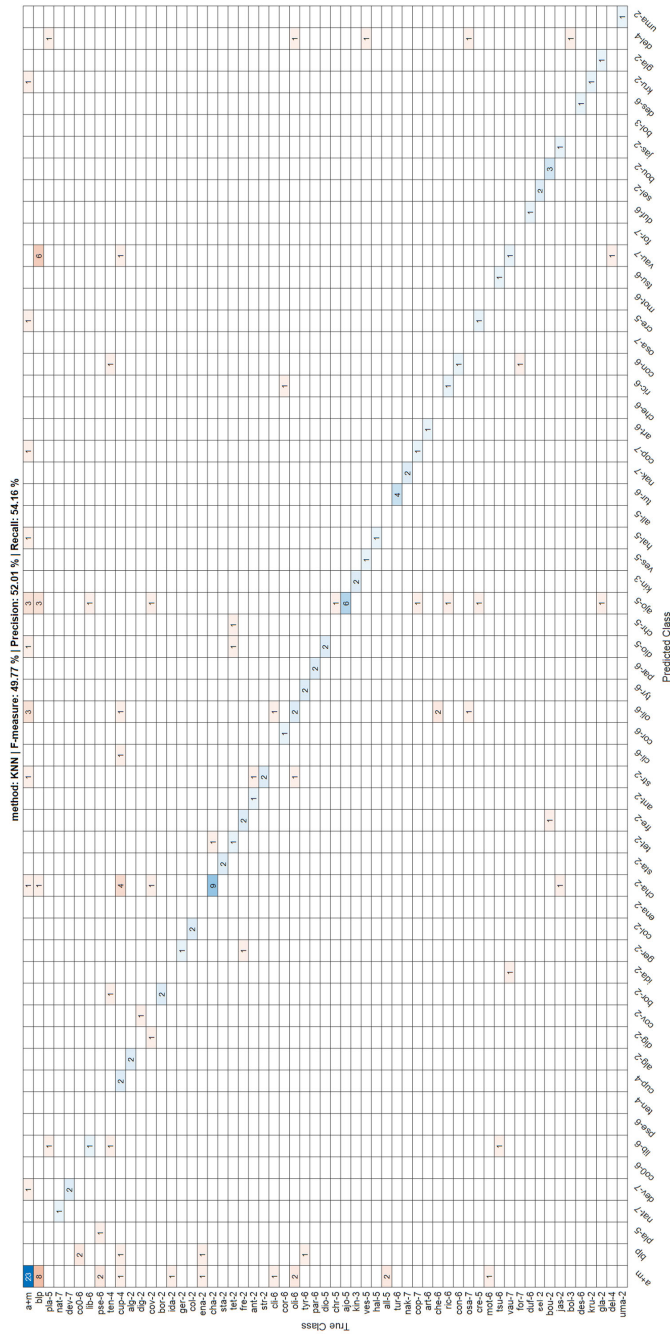


Figure S9. Confusion matrix of the KNN classifier for validation dataset (Section 5.3).

Table S12. CPU time and disk space usage of the compared methods. The desktop hardware was Intel i7-4790k (4.00GHz) with 32 GB RAM.

Method	CPU learning time (s)	CPU validation time for 188 MFs (s)	Disk space usage (MB)
SVM	1274.50	1134.26	1378.59
LDA	199.52	27.23	1440.37
KNN	4.26	462.27	964.46
NTF	1389.22	574.65	25.19

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Mein Dank gilt Prof. Dr. Hans-Gerd Löhmannsröben für die Möglichkeit dieses facettenreiche Thema in seiner Arbeitsgruppe erforschen und diese Dissertation schreiben zu können. Danke für die stete und gute Betreuung, die vielfältigen Anregungen und konstruktiven Ratschläge. Danke auch für die Akquise und Administration des Projektes LIBSqORE, in dessen Rahmen diese Arbeit entstanden ist.

Bedanken möchte ich mich auch bei Prof. Dr. Uwe Altenberger für die Unterstützung bei den geowissenschaftlichen Fragestellungen dieser Arbeit und der Bereitstellung von Proben sowie deren Analytik und für das stets offene Ohr.

Für die Übernahme der Gutachten danke ich Prof. Dr. Hans-Gerd Löhmannsröben, Prof. Dr. Uwe Altenberger und Dr. Jeannette Meima.

Danken möchte auch den IMSlern und LIBSlingen, allen voran Dr. Toralf Beitz für die beständige Unterstützung und die fachlichen Ratschläge. Danke an Alex, Chris, Daniel und Martin für die tolle Arbeitsatmosphäre, die tatkräftige Unterstützungen, die intensiven Gespräche und lustigen Gruppenausflüge sowie Weihnachtsfeiern. Danke auch an die gesamte UPPC für die unterhaltsamen Zusammenkünfte bei Feierlichkeiten, Spieleabenden oder anderen Gelegenheiten im Kommunikationsraum.

Für die lehrreichen Intermezzos in Sachen Petrologie und Geologie möchte ich mich herzlich bei Nicole Köllner bedanken. Auch an Nina Rethfeldt ein großes Dankeschön für die tolle Zusammenarbeit im Labor.

Abschließend möchte ich mich bei meiner Familie und meinen Freunden für die wunderbare Abwechslung und Motivation bedanken.

Bei Krüschi und Emilia möchte ich mich für eure immerwährende Ermutigung und Kraft, ohne die ich hier nicht angekommen wäre, bedanken.

Danke!

Eidesstattliche Erklärung

Hiermit versichere ich, Pia Brinkmann, geboren am 29.11.1989 in Potsdam, die vorliegende Dissertation selbstständig unter Zuhilfenahme der ausgewiesenen Hilfsmittel und Quellen verfasst und an keiner anderen Hochschule eingereicht zu haben.

Kleinmachnow, den 10.05.2022

Pia Brinkmann