MASARYK UNIVERSITY

FACULTY OF SCIENCE

Applications of Advanced Data Analysis and Chemometrics in Archaeometry, Geochemistry, and Mass Spectrometry

Habilitation thesis

LUBOMÍR PROKEŠ

Brno 2026

Bibliographic record

Author: Lubomír Prokeš
Department of Chemistry
Faculty of Science
Masaryk University

Název práce: *Applications of Advanced Data Analysis and Chemometrics in Archaeometry, Geochemistry, and Mass Spectrometry*

Rok: 2026

Počet stran: 117

Klíčová slova: analytical chemistry, archaeometry, chemometrics, elemental imaging, machine learning, mass spectrometry, methods comparison, multivariate statistics, spatial data

Abstract

The habilitation thesis presents the author's research in the Department of Chemistry at Masaryk University, which was focused on the field of chemometrics. The thesis is conceived as a commentary on annotated scientific papers. The thesis focusses on specific applications of chemometrics in archaeometry, geochemistry, and environmental chemistry. The text is divided into six sections, which are listed below with brief descriptions: first, problems related to analysis of archaeological, geochemical, and environmental data; secondly, multivariate data classification in archaeometry; third, analysis of spatial geo-archeological data; fourth, elemental imaging; fifthly, multivariate analysis of mass spectra; and sixth, application of method comparison techniques. The commentary to the selected scientific papers presents the author's contribution to the research via applications of multivariate data analysis and machine learning.

Acknowledgements

I want to express my gratitude to my parents, relatives and friends for their support.

My special thanks go to my supervisors and consultants for their invaluable help, advice and support:

Prof. RNDr. Viktor Kanický, DrSc.

Prof. RNDr. Vítězslav Otruba, CSc.

Prof. RNDr. Josef Havel, DrSc.

Prof. Dr. Eladia Maria Peña Méndez, Ph.D.

I would like to thank my colleagues with whom I have had the opportunity to work for a long time, especially to

Assoc. Prof. Mgr. Jan Čech, Ph.D.

Mgr. Jitka Hegrová, Ph.D.

Mgr. Aleš Hrdlička, Ph.D.

Assoc. Prof. Mgr. Karel Novotný, Ph.D.

Mgr. Lukáš Pečinka, Ph.D.

Mgr. Jan Petřík, Ph.D.

Mgr. Markéta Procházková, Ph.D. (†)

Prof. RNDr. Jiří Příhoda, CSc. (†)

Assoc. Prof. Mgr. Pavel Sťahel, Ph.D.

Assoc. Prof. Mgr. Michaela Vašinová Galiová, Ph.D.

Table of Contents

1 Introduction

Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house.

Henri Poincaré

Modern analytical chemistry produces a large amount of numerical data, including concentrations, signal intensities, mass spectra, and chromatographic profiles. Without sophisticated statistical analysis, these data are simply numbers. Advances in analytical instrumentation can lead to the 'Data Rich, Information Poor' (DRIP) syndrome – accumulation of vast amounts of data without derivation of meaningful insights from it (Timmerman et al., 2010; Ward et al., 1986). Data analysis thus serves as the conduit between measurement and scientific knowledge. The evolving capabilities of modern computing technology have enabled the emergence of sophisticated, computer-intensive methods of data processing and visualisation (e.g. machine learning, ensemble methods, Bayesian methods, etc.), which places ever greater demands on the choice of the right data analysis procedures. A solid grasp of data analysis methodologies is paramount to enhance the reliability and validity of results, ensure their precise interpretation, and draw pertinent conclusions (Szymańska, 2018). Consequently, data analysis has evolved into a crucial skill in analytical chemistry in the 21st century (Adams & Adriaens, 2020).

The research presented in this thesis aims to demonstrate the indispensable role of chemometrics and advanced data analysis in modern analytical chemistry and also to point out some methodological problems associated with the analysis of results in the fields of archaeometry, geochemistry, and environmental chemistry.

1.1 Structure of the thesis

The text of this thesis summarises the author's experience with selected methods of data analysis using specific examples from the fields of archaeometry, geochemistry, and environmental chemistry.

Currently, the author has published 68 papers in impacted journals, 21 papers in non-impacted journals, one scholarly book publication (as a coauthor), 6 chapters in scholarly book publications, and 30 papers in conference proceedings. These publications relate to analytical chemistry (ICP-OES, GC-MS, LDI-TOF MS, etc.), archaeology, archaeometry, applications of data analysis and chemometrics, plasma deposition, and plasma treatment of air and water. For his professional work, the author was awarded the Dean's Award of the Faculty of Education, Masaryk University, for long-term and outstanding achievements in science, research, and education in the academic year 2024/2025.

Chemometrics and analytical data processing are one of several areas in which the author has specialised in his scientific work and on which he intends to focus during his tenure at the Faculty of Education of Masaryk University. This involves the use of modern statistical

methods and machine learning in the fields of archaeometry, geochemistry, and environmental sciences. The author has created a set of scripts in the R and Python programming languages for chemometrics and mass spectra processing.

For this thesis, 18 research articles (see numbered citations in Appendix B) related to the the applications of analytical-chemical data analysis methods in the fields of archaeometry, geochemistry, and environmental sciences (11 articles), data analysis in mass spectrometry (5 articles), and quantitative methods comparison applications (2 articles) were selected.

The entire thesis is focused on five themes:

1. Selected problems of the analysis of archaeological, geochemical, and environmental data.
2. Classification and provenance determination in geochemistry and archaeometry
3. Statistical methods in elemental bioimaging
4. Spatial data analysis in archaeology
5. Treatment of LDI-TOF MS mass spectra
6. Comparison of analytical methods using statistical procedures

Chemometrics is a common language that connects archaeometry, geochemistry, environmental chemistry, and mass spectrometry. This is because all these fields work with complex, multidimensional data whose interpretation requires advanced statistical and multivariate methods. These methods enable the conversion of extensive analytical outputs into comprehensible relationships and decision-making models that provide a better understanding of the origin, development, and interactions of materials in natural and anthropogenic systems.

The data analysis in this thesis and the selected articles is performed, with few exceptions, by freely available statistical software (*https://www.r-project.org/*). In article **[15],** the commercial statistical package *Statistica software* (TIBCO Software) was used on request of the coauthors. For ANN calculations, the *Trajan* neural network simulator (Trajan software Ltd.) was applied in the articles **[14-15]**, or *Statistica* (TIBCO Software) in the article **[6]**. The spatial interpolation in the articles **[9]** and **[10]** was calculated by the coauthors with commercial *ArcGIS* (ESRI), commonly used by archaeologists. The mass spectra of the articles **[12-15]** were treated with *Shimadzu Biotech Launchpad* 2.9 software (Kratos Analytical Ltd.).

1.2 List of Annotated Papers

For the habilitation thesis, research articles dedicated to the research of chemometrics and mass spectrometry were selected. The list of publications is ordered according to their appearance in the thesis. The author's contribution to annotated papers is estimated in terms of content and quality in the respective tables and the list below.

[1] Petřík, J.; **Prokeš, L.**; Všianský, D.; Salaš, M.; Nikolajev, P. Organization of ceramic production at a fortified Early Bronze Age settlement in Moravia (Czech Republic) inferred from minimally destructive archaeometry. *Archaeological and Anthropological Sciences*, 2018, 10 (3), 697-709. DOI:10.1007/s12520-016-0370-8

Document Type: Article, IF = 2.2; JCR Category + Category Quartile: ANTHROPOLOGY Q1 + GEOSCIENCES, MULTIDISCIPLINARY Q3; AIS = 0.740

Author's contribution to the publication: The author participated in the processing and interpretation of the statistical data and in the preparation of the manuscript. The proposal included the utilisation of multidimensional statistical methodologies, including correspondence analysis (whose results were excluded from the final version), using the R software. He performed the necessary calculations, actively participated in the evaluation of the data obtained, and collaborated on the creation of the manuscript intended for publication.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	30%	30%

[2] Prokeš, L.; Vašinová Galiová, M.; Hušková, S.; Vaculovič, T.; Hrdlička, A.; Mason, A.Z.; Neff, H.; Přichystal, A.; Kanický, V. Laser microsampling and multivariate methods in provenance studies of obsidian artefacts. *Chemical Papers*, 2015, 69 (6), 761-778. DOI:10.1515/chempap-2015-0019

Document Type: Article, IF = 1.3; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.235

Author's contribution to the publication: The author participated in the analysis using LA-ICP-MS and designed and computationally implemented multidimensional statistical methods. Some of these methods have not yet been applied in archaeometry (for example, ICA, correspondence analysis, fuzzy clustering). In addition to these responsibilities, he undertook the interpretation of the results, the implementation of the relevant procedures using the R software and made a substantial contribution to the preparation of the manuscript.

Experimental work	Supervision	Manuscript	Research direction
50%	10%	80%	70%

[3] Hrdlička, A.; **Prokeš, L.**; Vasinová Galiová, M.; Novotný, K.; Vitešníková, A.; Helešicová, T.; Kanický, V. Provenance study of volcanic glass using 266-1064 nm orthogonal double pulse laser induced breakdown spectroscopy. *Chemical Papers*, 2013, 67 (5), 546-555. DOI:10.2478/s11696-013-0332-x

Document Type: Article, IF = 1.2; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.194

Author's contribution to the publication: The author participated in the performance of LIBS analyses and proposed the use of appropriate statistical data processing methods and visualisation techniques that are optimised for small data sets. Examples of such techniques include radar charts and Chernoff faces. In addition to these responsibilities, he was tasked with interpreting the results obtained and subsequent collaboration on the preparation of the manuscript, including formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
50%	10%	40%	60%

[4] Burgert, P.; Přichystal, A.; **Prokeš, L.**; Petřík, J.; Hušková, S. The origin and distribution of obsidian in prehistoric Bohemia. *Bulgarian e-Journal of Archaeology | Българско е-Списание за Археология*, 2017, 7 (1), 1-15.

Document Type: Article, IF = N/A; JCR Category + Category Quartile: N/A; AIS = N/A

Author's contribution to the publication: The author participated in the measurement of measurements using pXRF and in the subsequent statistical processing of the data obtained. He made a significant contribution to the preparation of a reference database of Carpathian obsidian and to the implementation of the necessary calculations in the R environment. Furthermore, he actively participated in the interpretation of the results and collaborated on the preparation of the manuscript, including the formulation of conclusions and the overall structure of the text intended for publication.

Experimental work	Supervision	Manuscript	Research direction
40%	10%	50%	60%

This article is an extended version of the article not included in the thesis.

Burgert, P.; Přichystal, A.; Prokeš, L.; Petřík, J.; Hušková, S. The origin of obsidian in prehistoric Bohemia. *Archeologické rozhledy*, 2016, 68 (2), 224-234.

[5] Vítková, G.; **Prokeš, L.**; Novotný, K.; Pořízka, P.; Novotný, J.; Všianský, D.; Čelko, L.; Kaiser, J. Comparative study on fast classification of brick samples by combination of principal component analysis and linear discriminant analysis using stand-off and table-top laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B - Atomic Spectroscopy*, 2014, 101, 191-199. DOI:10.1016/j.sab.2014.08.036

Document Type: Article, IF = 3.3; JCR Category + Category Quartile: SPECTROSCOPY Q1; AIS = 0.542

Author's contribution to the publication: The author participated in the preparation of samples and the selection of appropriate methods for data processing, which he subsequently

applied in the R environment. The scope of his work encompassed the execution of analyses, of which a proportion was subsequently excluded from the final version of the article. Furthermore, he actively participated in the interpretation of the obtained data and collaborated on the preparation of the manuscript for publication, including the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
40%	10%	30%	40%

[6] Vítková, G.; Novotný, K.; **Prokeš, L.**; Hrdlička, A.; Kaiser, J.; Novotný, J., Malina, R.; Prochazka, D. Fast identification of biominerals by means of stand-off laser-induced breakdown spectroscopy using linear discriminant analysis and artificial neural networks. *Spectrochimica Acta Part B - Atomic Spectroscopy*, 2012, 73, 1-6. DOI:10.1016/j.sab.2012.05.010

Document Type: Article, IF = 3.1; JCR Category + Category Quartile: SPECTROSCOPY Q1; AIS = 0.748

Author's contribution to the publication: The author selected and proposed appropriate data analysis methods that had not yet been used in connection with LIBS and ensured their implementation in the R environment and in the Statistica software (for artificial neural networks). The subject also participated in the statistical processing of the data, actively participated in the interpretation of the results, and made a significant contribution to the preparation of the manuscript for publication.

Experimental work	Supervision	Manuscript	Research direction
30%	10%	30%	40%

[7] Vyskočilová, G.; Ebersbach, M.; Kopecká, R.; **Prokeš, L.**; Příhoda, J. Model study of the leather degradation by oxidation and hydrolysis. *Heritage Science*, 2019, 7, 26. DOI:10.1186/s40494-019-0269-7

Document Type: Article, IF = 1.9; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q3 + SPECTROSCOPY Q2 + MATERIALS SCIENCE, MULTIDISCIPLINARY Q3; AIS = N/A

Author's contribution to the publication: The author participated in the selection of suitable data analysis methods (CART and Random Forest) and their implementation in the R environment. In addition to his contributions to the data processing itself, he participated actively in the interpretation of the results and made significant contributions to the preparation of the manuscript, particularly in the formulation of conclusions and the overall structure of the text intended for publication.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	20%	20%

[8] Loun, J.; Novák, M.; Cempírek, J.; Škoda, R.; Vašinová Galiová, M.; **Prokeš, L.**; Dosbaba, M.; Čopjaková, R. Geochemistry and secondary alterations of microlite from eluvial deposits in the Numbi mining area, South Kivu, Democratic Republic of the Congo. *Canadian Mineralogist*, 2018, 56 (2); 203-220. DOI:10.3749/canmin.1700091

Document Type: Article, IF = 1.4; JCR Category + Category Quartile: MINERALOGY Q3; AIS = 0.44

Author's contribution to the publication: The author participated in the data processing and selection of appropriate statistical and visualisation methods (CART, Random Forest, etc.) to determine the provenance of microliths. These were subsequently implemented using the R software. The results obtained were only partially used for the purposes of the published article, with a significant part of the analyses concerning the provenance of microliths not yet being published. The author also participated in the interpretation of the results and contributed to the preparation of the manuscript, in particular to the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	20%	30%

[9] Petřík, J., Adameková, K., Goláňová, P., Tencer, T., **Prokeš, L.** The potential of low-destructive characterization of archaeological sites with stony and eroded soils through geostatistics at the Celtic oppidum of Bibracte (France). *Journal of Archaeological Science: Reports*, 2024, 55, 104509. DOI:10.1016/j.jasrep.2024.104509

Document Type: Article, IF = 1.6; JCR Category + Category Quartile: ARCHAEOLOGY Q1; AIS = 0.547

Author's contribution to the publication: The author participated in the data processing and proposed appropriate methods for their analysis and visualisation, including local indicators of spatial autocorrelation (LISA). In addition, he participated in the implementation of these models within the R environment. It is evident that a proportion of the results obtained from the data analysis were not included in the final version of the article. The author also participated in the interpretation of the results and contributed to the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	20%	30%

The article is an extensively expanded version of a book chapter not included in the thesis.

Petřík, J.; Prokeš, L. Geostatistical evaluation of the chemical dataset from Terrasse. In: *Oppidum as an urban landscape. A multidisciplinary approach to the study of space organisation at Bibracte* (Ed. Goláňová, P.; Milo, P.; Hajnalová, M.), Collection Bibracte – 33, Bibracte – Centre archéologique européen, Glux-en-Glenne, 2023, 303-308. ISBN: 978-2-490601-14-1

10] Prišťáková, M.; Adameková, K.; Petřík, J.; Dresler, P.; **Prokeš, L.** Tracing the spatial organization and activity zones of an Early Mediaeval homestead at the Pohansko stronghold (Czechia) by combining geophysics and geochemical mapping. *Archaeological Prospection*, 2023, 30 (4), 449-464. DOI:10.1002/arp.1907

Document Type: Article, IF = 1.8; JCR Category + Category Quartile: GEOSCIENCES, MULTIDISCIPLINARY Q3; AIS = 0.492

Author's contribution to the publication: The author participated in data processing, proposed appropriate methods for data analysis and visualisation, and collaborated on their implementation using R software. The author also participated in the interpretation of the results and contributed to the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
10%	10%	20%	25%

[11] Vašinová Galiová, M.; Száková, J.; **Prokeš, L.**; Čadková, Z.; Coufalík, P.; Kanický, V.; Otruba, V.; Tlustoš, P. Variability of trace element distribution in Noccaea spp., Arabidopsis spp., and Thlaspi arvense leaves: The role of plant species and element accumulation ability. *Environmental Monitoring and Assessment*, 2019, 191 (3), 181. DOI:10.1007/s10661-019-7331-5

Document Type: Article, IF = 3; JCR Category + Category Quartile: ENVIRONMENTAL SCIENCES Q3; AIS = 0.445

Author's contribution to the publication: The author designed and applied procedures for statistical data processing, including the use of spatial autocorrelation methods (LISA), which had not previously been applied to elements' composition maps (some of the results were not included in the manuscript). Concurrently, he participated in the implementation of these methods in the R environment, the interpretation of the obtained data, and the preparation of the manuscript for publication, particularly in the formulation of conclusions and the structuring of the text.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	25%	25%

[12] Prokeš, L.; Pečinka, L. Leveraging R for advanced interpretation of LDI-TOF mass spectra: A computational approach *Analytical Methods*, 2025, 17 (41), 8502-8506. DOI: 10.1039/d5ay01246e

Document Type: Article, IF = 2.6; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q3 + FOOD SCIENCE & TECHNOLOGY Q3 + SPECTROSCOPY Q2; AIS = 0.405

Author's contribution to the publication: The author participated in the measurement and processing of mass spectra, and the results of this work were published in articles that are not part of the Annotated Papers collection. In addition, he participated in the selection of suitable methods for processing spectra, created the majority of the scripts used in the R software, and contributed to their practical implementation. Furthermore, he actively participated in the interpretation of the results obtained and made a significant contribution to the preparation of the manuscript for publication.

Experimental work	Supervision	Manuscript	Research direction
50%	60%	50%	60%

[13] Kolářová, L.; Prokeš, L.; Kučera, L.; Hampl, A.; Peña-Méndez, E.; Vaňhara, P.; Havel, J. Clusters of monoisotopic elements for calibration in (TOF) mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2017, 28 (3), 419-427. DOI:10.1007/s13361-016-1567-x

Document Type: Article, IF = 2.9; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q2 + BIOCHEMISTRY & MOLECULAR BIOLOGY Q3; AIS = 0.791

Author's contribution to the publication: The author participated in the measurement and processing of the mass spectra. He developed his own scripts in R software for fitting peaks using the least-squares method for various models (Gauss, Cauchy, pseudo-Voigt, and Pearson VII) and applied these procedures to the measured data, although some of the results were not included in the manuscript. Furthermore, he actively participated in the interpretation of the results obtained and collaborated on the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
40%	10%	40%	25%

[14] Valletta, E.; Kučera, L.; **Prokeš, L.**; Amato, F.; Pivetta, T.; Hampl, A.; Havel, J.; Vaňhara, P. Multivariate calibration approach for quantitative determination of cell-line cross contamination by intact cell mass spectrometry and artificial neural networks. *PLOS One*, 2016, 11 (1), e0147414. DOI:10.1371/journal.pone.0147414

Document Type: Article, IF = 2.8; JCR Category + Category Quartile: MULTIDISCIPLINARY SCIENCES Q1; AIS = 1.053

Author's contribution to the publication: The author participated in data processing, selection of appropriate statistical methods and their application in R (for PLS regression models) and Trajan (for artificial neural networks). Furthermore, he actively participated in the interpretation of the results and contributed significantly to the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
30%	10%	30%	25%

[15] Vaňhara, P.; Kučera, L.; **Prokeš, L.**; Jurečková, L.; Peña-Méndez, E.M.; Havel, J.; Hampl, A. Intact cell mass spectrometry as a quality control tool for revealing minute phenotypic changes of cultured human embryonic stem cells. *Stem Cells Translational Medicine*, 2018, 7 (1), 109-114. DOI:10.1002/sctm.17-0107

Document Type: Article, IF = 6; JCR Category + Category Quartile: CELL & TISSUE ENGINEERING Q1; AIS = 1.472

Author's contribution to the publication: The author participated in the design of suitable methods for data analysis (PCA, ANNs) and in the statistical processing of mass spectra in

Statistica and Trajan software. Furthermore, he actively participated in the interpretation of the results obtained and contributed significantly to the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
20%	10%	30%	20%

[16] Pečinka, L.; Pantůčková, J; Vlachová, M.; Moráň, L.; Růžičková, T; Weselá, P.; Prokeš, L.; Havel, J.; Pour, L.; Ševčíková, S.; Vanhara, P. End-to-end workflows for liquid biopsy biotyping analysis using combined MALDI MS and machine learning approach. *Analytical Methods*, 2025, 17(48), 9909-9914. DOI: 10.1039/d5ay01299f

Document Type: Article, IF = 2.6; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q3 + FOOD SCIENCE & TECHNOLOGY Q3 + SPECTROSCOPY Q2; AIS = 0.405

Author's contribution to the publication: The author participated in the creation of the workflow, the selection of appropriate methods for data analysis and visualisation, and their application in statistical data processing in the R environment. It is evident that he participated in the interpretation of the results obtained and made significant contributions to the preparation of the article manuscript. This contribution included the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
10%	5%	15%	15%

[17] Janotková, I.; **Prokeš, L.**; Vaculovič, T.; Holá, M.; Pinkas, J.; Steffan, I.; Kubáň, V.; Kanický, V. Comparison of inductively coupled plasma optical emission spectrometry, energy dispersive X-ray fluorescence spectrometry and laser ablation inductively coupled plasma mass spectrometry in the elemental analysis of agricultural soils. *Journal of Analytical Atomic Spectrometry*, 2013, 28 (12), 1940-1948. DOI:10.1039/c3ja50169h

Document Type: Article, IF = 3.4; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q1 + SPECTROSCOPY Q1; AIS = 0.782

Author's contribution to the publication: The author participated in the selection of suitable samples and subsequently prepared scripts to calculate various regression methods, including York regression, BLS regression, and Ripley–Thompson regression. Furthermore, he developed methodologies for the calculation of confidence intervals and confidence ellipses through the use of the bootstrap method and facilitated the generation of a Bland–Altman plot. He thoroughly processed the data obtained in the R environment, actively participated in the interpretation of the results, and contributed significantly to the preparation of the manuscript for publication, particularly in the formulation of conclusions and the overall structure of the text.

Experimental work	Supervision	Manuscript	Research direction
30%	10%	50%	50%

[18] Prokeš, L.; Hegrová, J.; Kanický, V. Analysis of means (ANOM) as a tool for comparison of sample treatment methods: testing various mineralization procedures for selenium determination in biological materials. *Journal of AOAC International*, 2017, 100 (1), 236-240. DOI:10.5740/jaoacint.16-0258

Document Type: Article, IF = 1.1; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q4 + FOOD SCIENCE & TECHNOLOGY Q3; AIS = 0.234

Author's contribution to the publication: The author participated in the following: sample preparation, statistical data evaluation, interpretation of results, and manuscript preparation. The ANOM method was proposed for the first time in this study, with the author creating his own scripts in R software to calculate the mean, standard deviation, and number of measurements. These functions were not available in existing libraries (e.g., ANOM), and so it was necessary to implement them separately. The author also processed the data obtained and participated in the creation of the final version of the manuscript intended for publication.

Experimental work	Supervision	Manuscript	Research direction
30%	10%	50%	40%

2 Selected problems of the analysis of archaeological, geochemical and environmental data

Archaeometrical, environmental, clinical, chemical, and geochemical data often exhibit certain characteristics that severely limit the ability to apply classical methods of statistical data analysis. These include, in particular:

- Presence of missing or unmeasured data (NA values).
- Presence of values below the detection/quantification limit (the so-called 'left-censored data')
- Normalised data (as ratios or percentages).
- Data involving repeated measurements (in many cases in the form of $\bar{x} \pm sd$)
- Dependent (autocorrelated) data (temporally and/or spatially), often heteroscedastic (variance of the residuals is unequal over a range of measured values).
- Biased datasets, especially due to selection bias. Violation of the assumption of random selection significantly limits the use of statistical induction methods.
- Heteroscedasticity in the data and in measurement errors.
- Nonhomogeneous data sets (mixtures), showing the presence of outliers, multimodality, and skewed or platykurtic mixture distributions.
- Data sets with a non-Gaussian distribution of the data.
- Small data file size or measurements with low number of repetitions per sample

These factors impose certain limitations on the usage of traditional statistical methods and also on the application of special methods, which are often not included sufficiently in common textbooks or are not a standard part of common statistical software.

Selection bias in sampling

Selection bias occurs when the method or criteria used to select samples cause the sample (data set) to be unrepresentative of the entire population or study area it is intended to examine. As a result, conclusions drawn from the data are incorrect or biased (Hegedus & Moody, 2010; Pannucci & Wilkins, 2010). For example, in archaeometry it occurs when samples selected for analysis lack representativeness within the broader archaeological context, due to cost, accessibility, or perceived importance. This may lead to issues such as false clustering, misinterpretation of chemical signatures, and distorted interpretations of provenance, chronology, or technology. In the field of archaeometry, these variants of selection bias are of significant importance:

Sampling bias: It occurs when only certain types of artefacts are selected for analysis. It leads to over-representation of specific artefacts, materials, or styles (Banning, 2021; Purtill, 2023).

- *Spatial bias*: Inconsistent research efforts in different geographic areas can result in incomplete data. In environmental analysis, samples may only be taken from easily accessible locations (e.g., near roads, at shallow depths, or only from locations with visible contamination).
- *Excavation bias*: The choice of inappropriate sampling methods (convenience sampling instead of random sampling) has the potential to introduce bias (methodological bias). It is influenced by excavation methods and recording practices (small or fragmentary items may be missed or discarded).
- *Analytical bias*: It occurs when only certain samples are chosen for instrumental analysis (e.g. LA-ICP-MS, XRF). Furthermore, samples with concentrations below the limit may be ignored or incorrectly processed statistically.

Survivorship bias is focused only on available objects, published studies, or datasets. For example, the preservation of skeletal remains is significantly influenced by soil properties, bone type, age, and sex of individuals (Gordon & Buikstra, 1981; Paine & Harpending, 1998; Walker et al., 1988), therefore only preserved skeletal remains are available for analysis. On the other hand, bioarchaeological growth studies are conducted principally on the remains of nonsurvivors (Saunders & Hoppa, 1993).

The *Berkson paradox* is a statistical phenomenon in which two independent variables appear to be correlated due to a selection bias in the data (Snoep et al., 2014; Westreich, 2012), and vice versa. For example, two variables that are positively correlated in the general population appear to show a poor correlation when they are analysed only in a subpopulation (Barchard & Russell, 2023) (Fig. 1).

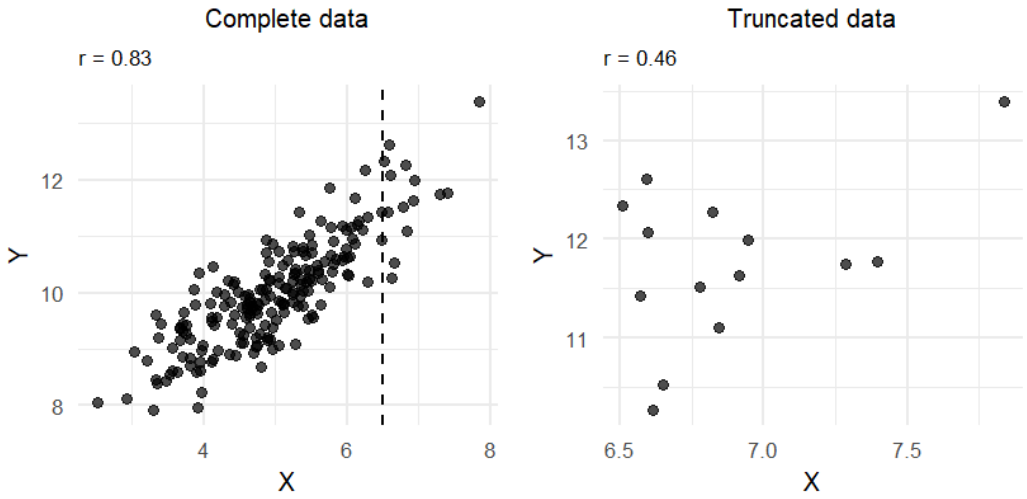

Fig. 1. An example of the Berkson paradox. Complete data (left) and truncated subset of the data (right).

Inference from absence (the absence of evidence as evidence of absence) highlights the distinction between not finding evidence for something and proving that it does not exist.

To constitute strong evidence of absence, a key condition must be met: evidence must be highly expected if the phenomenon exists (Brewer, 2025; Thompson & Scurich, 2018; Wallach, 2019). In the case of insufficient sample size, statistical tests often do not show a significant difference between the compared sets for the selected level of significance (type II error), which incorrectly suggests evidence of a non-existent difference, while in fact it has only been proven that there is no evidence of a difference (Altman & Bland, 1995).

Cognitive biases and errors in judgment in data analysis

Cognitive biases are mental shortcuts that reflect adaptive mechanisms that prioritise efficiency over accuracy (Kahneman, 2011). These biases can distort how researchers perceive, analyse, and report data, even when using rigorous methods (Bertrand et al., 2022; Kahneman, 2011; Kliegr et al., 2021). In archeometry, especially, cognitive biases are critical to correct interpretation, particularly in cases where data are partial, preservation is selective, and sampling is uneven. They are related to small, unrepresentative data sets, which may occur particularly in connection with archaeometric data sets. Preventing cognitive biases is crucial for maintaining objectivity, accuracy, and credibility of results.

Availability bias judgment based on information that is most easily recalled, e.g., decisions are influenced by using more available data, rather than by comprehensive or representative data set (Kahneman, 2011; Tversky & Kahneman, 1973). It is usually more difficult to notice something is missing than to notice something is present. The absence of certain attributes is less noticeable than the presence of other attributes (feature-positive effect) (Allison & Messick, 1988). Examples of availability bias are e.g. selection of the most striking features as descriptors, or selection of elements and methods of data analysis based on published articles.

The *representativeness bias* is a cognitive bias that leads us to make judgments about how likely an object is to belong to a particular category based on how closely it resembles our mental prototype or stereotype of that category (Kahneman, 2011; Kahneman & Frederick, 2002; Kahneman & Tversky, 1972). In other words, people tend to put things into categories based on how they look or act and on what they already know about them (AlKhars et al., 2019). The presence of selected characters or descriptors may be attributed to availability bias. For instance, some obsidian sources (e.g. Tikal in Maya area) have so much variation that pieces from the same source might appear macroscopically distinct while obsidian from different sources may look identical (Moholy-Nagy, 2003; Moholy-Nagy & Nelson, 1990), instrumental methods provide much better results (Moholy-Nagy et al., 2013). Generally, in the case of obsidian, the success rate of visually determining provenance can be more than 90% (Braswell et al., 2000; P. Moore, 2025; Pierce, 2015; Stroth et al., 2019), while for other raw materials, such as chert (Parish & Durham, 2015), the reliability is lower.

The existence of inter-individual and intra-individual differences in the expert opinions can be attributed to the following sources (C. Beck & Jones, 1989; Fish, 1978; Lyman & VanPool, 2009): (1) the explicitness of class definitions, (2) differences in perception among analysts, and (3) changes in a single analyst's perception over time. To test interrater

reliability Cohen kappa is a method for quantifying the degree of agreement between two raters (or observers) when they classify a set of items into a fixed number of predefined nominal categories (McHugh, 2012; Zec et al., 2017).

The effect of representativeness is reflected in the fact that different researchers or methods can often come to significantly different conclusions (Daumas et al., 2023; Gnaden & Holdaway, 2000; Lyman & VanPool, 2009). Problems with using a forensic database based on recent osteological material for ethnicity determination of ancient skeletal remains (F. L. Williams et al., 2005) are an example of combined availability and representativeness biases.

The *law of small numbers* means that people mistakenly assume that small samples accurately represent the larger population (Kahneman, 2011; Tversky & Kahneman, 1971). When working with small sample sets, it is important to be cautious, as conclusions drawn from limited data may be overly confident. Randomly generated small data sets may show a higher correlation than random sets with a larger number of individuals (Fig. 2).

The *clustering illusion* is a cognitive bias by which people perceive patterns or clusters in random data (Kahneman, 2011). This phenomenon stems from our tendency to underestimate randomness and misinterpret apparent groupings as significant (Fig. 2).

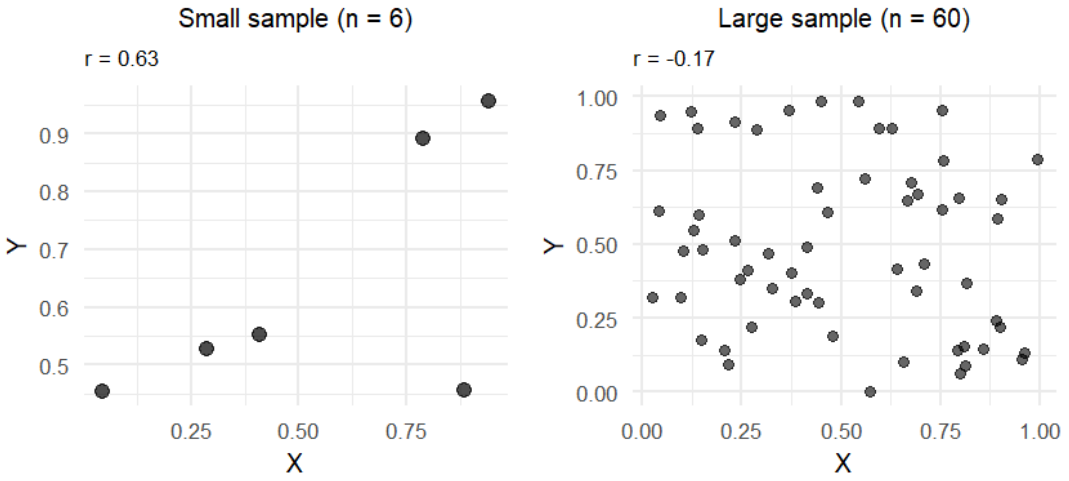

Fig. 2. An example of the law of small numbers and the clustering illusion. Small sample (left) and great sample (right). Values were generated randomly from uniform distribution.

Illusory correlation is a psychological phenomenon in which individuals perceive a relationship between two variables that is not actually present (Chapman, 1967). This cognitive bias frequently emerges due to the fact that uncommon or distinctive occurrences capture our attention, resulting in the formation of false associations.

Confirmation bias (myside bias, disconfirming evidence) is the tendency to formulate a hypothesis and then look for ways to confirm it (Born, 2024; R. S. Nickerson, 1998). It is important to be aware of the tendency to disregard facts that contradict our beliefs (opinions) and to interpret ambiguous information in a way that is consistent with them. Confirmation

bias can have particularly serious consequences in forensic chemistry, e.g., in the Patricia Stallings case (Shoemaker et al., 1992; Zurer, 1991).

The *anchor effect* (Furnham & Boo, 2011; Kahneman, 2011) is a phenomenon in which individuals focus intently on one piece of information or feature, even if it is irrelevant to the decision at hand. This tendency can lead to a narrow focus on the anchor, potentially missing other crucial factors that could influence the decision more significantly.

To limit subjectivity and the influence of cognitive biases, it is necessary to reflect on the results of the data analysis and understand them correctly. For example, you can ask yourself the following questions:

- What do the data actually tell us?
- What do the results mean?
- What are we missing?
- What do not we not know?
- Are we thinking about them correctly?
- How can we look at them or understand them differently?

Formal fallacies in data analysis

The *ecological fallacy* occurs when conclusions about the characteristics of one person are formed based on information about the group that that person belongs to (Fig. 3). This is a case of erroneous judgment, where conclusions about individuals are derived from aggregated data (Roumeliotis et al., 2021; Winzar, 2015).

The *Berkson error* occurs when individual values are replaced with aggregated averages or predictions (Haber et al., 2021). Consequently, the model maintains the correct average but may fail to capture actual individual relationships. Berkson's error may prevail over classical error in cases where exposure data are highly aggregated.

Simpson's paradox is related to the ecological fallacy. It occurs when a trend observed in several groups disappears or reverses when the data are aggregated and vice versa (Goltz & Smith, 2010; Kievit et al., 2013) (Fig. 4). The paradox can be resolved when confounding variables and causal relations are adequately addressed in the relevant statistical modelling.

Lord's paradox (Holland & Rubin, 1982; Kane & Mroch, 2020; Pearl, 2016) occurs when two common statistical methods applied to the same data yield opposite conclusions, typically when comparing change (gain score) with analysis of covariance (ANCOVA) (control of initial value). Although both analyses are statistically sound, they address different aspects of the question. In the absence of an explicit causal model, it is not possible to determine which answer is the "correct" one. In addition, the analysis of change from baseline is burdened by *mathematical coupling* (see below), while ANCOVA does not contain it (Y. Tu et al., 2005).

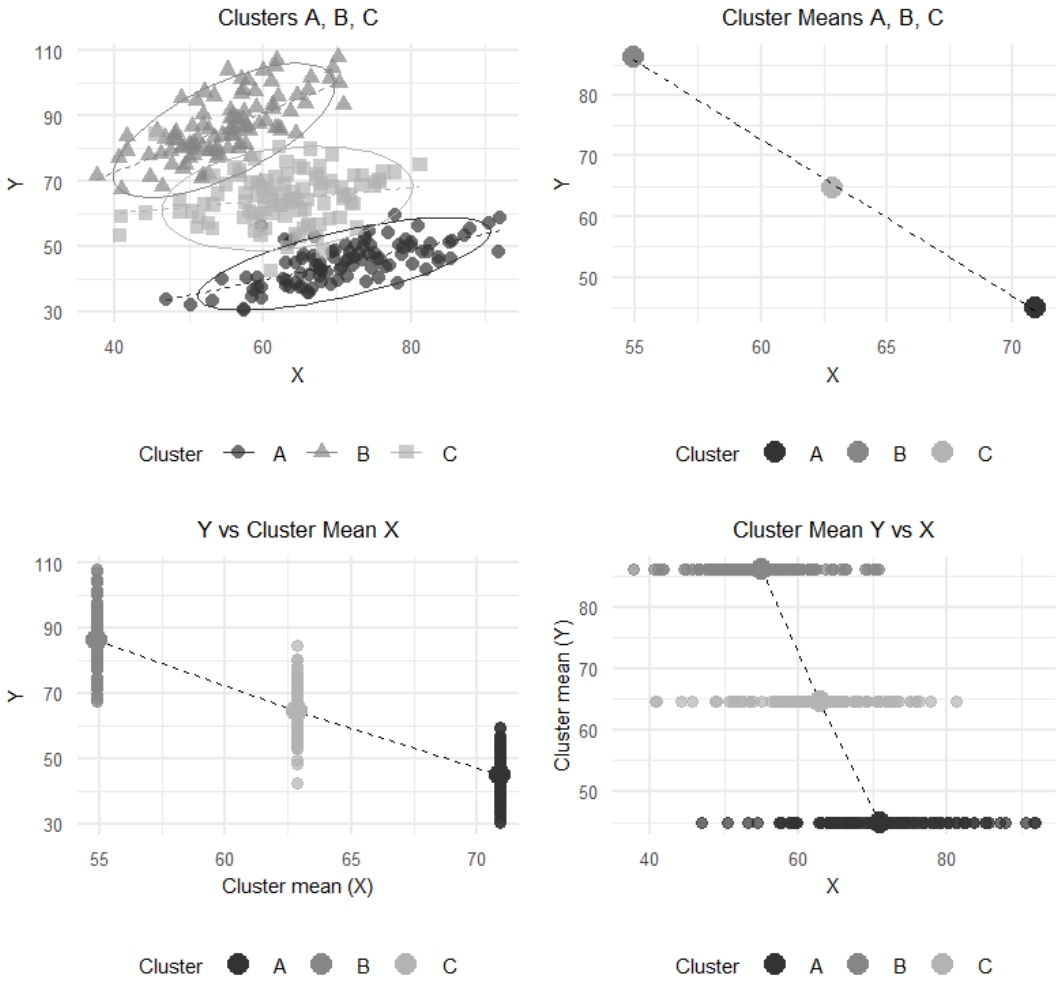

Fig. 3. An illustration of the ecological fallacy and the Berkson error. Original data (topleft), mean values of the clusters (topright), mean values on the x-axis (bottoleft), and mean values on the y-axis (bottom right).

Lord's paradox, Simpson's paradox and the ecological fallacy have one thing in common: they all demonstrate how misinterpreting statistical relationships can lead to misleading or even opposite conclusions (Y.-K. Tu et al., 2008). The necessity to differentiate between descriptive statistics and causal inference is the key factor that binds them.

Spurious correlation is defined as a statistical relationship between two variables that appears to be meaningful, but is actually coincidental or influenced by a hidden third factor (Ghouse et al., 2024; Haig, 2003). These correlations frequently mislead people into assuming causation when no direct connection exists.

Regression toward the mean (regression to the mean, regression towards mediocrity) (Barnett, 2004; Kahneman, 2011; Streiner, 2001) is a statistical phenomenon whereby extreme values tend to return closer to the mean when measured repeatedly. It is purely a consequence of random variability, not an actual change in performance or characteristics.

Missing data

Missing data is a pervasive issue in real-world datasets, and how it is handled can dramatically affect the validity of statistical analyses, machine learning models, and scientific conclusions.

There are three types of missingness (Alwateer et al., 2024; Austin et al., 2021; Heymans & Twisk, 2022; Walczak & Massart, 2001b):

MCAR (missing completely at random): No systematic pattern; least problematic.

MAR (missing at random): Missingness depends on observed data; manageable with proper modelling.

MNAR (missing not at random): Missingness depends on unobserved data; most challenging.

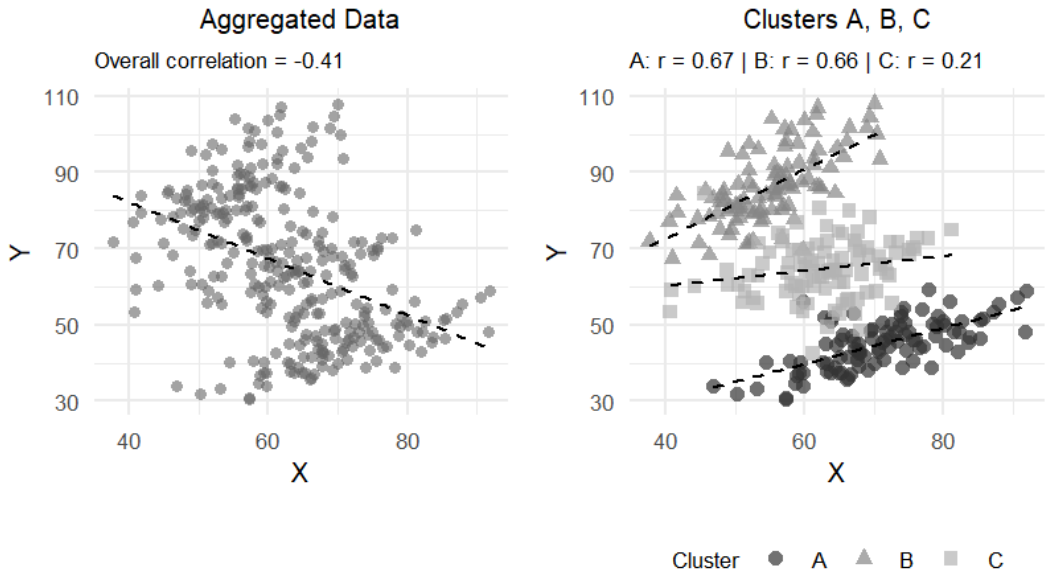

Fig. 4. An illustration of the Simpson paradox. Aggregated data (left) and clusters (right).

Diagnostics of missing values can be performed in several ways:

Visual diagnostics (Templ et al., 2012; Tierney & Cook, 2023; Z. Zhang, 2015) visualise missingness patterns.

Little's test (Little, 1988) is a statistical procedure used to assess whether missing data in a data set are MCAR (that is, the probability of missingness is not related to observed and unobserved data).

Logistic regression of missingness (Austin et al., 2021) models missingness (TRUE/FALSE) as a function of observed variables. Significant predictors suggest that MAR is plausible.

The fraction of missing information (FMI) reflects the proportion of total variability in a parameter estimate that arises from missing data (Wagner, 2010). It is a more informative metric than the raw proportion of missing data, especially when auxiliary variables are available (Madley-Dowd et al., 2019).

The most used methods of missing data treatment are:

Complete case analysis uses only rows with no missing values. It is simple and widely used, but it reduces sample size, and it is biased if not MCAR.

Simple imputation methods (Alwateer et al., 2024; Gómez-Carracedo et al., 2014; Heymans & Twisk, 2022; Walczak & Massart, 2001a; Wissler et al., 2022; Zhang, 2016) are not recommended because most of these methods lead to an artificial decrease standard deviation in the variables to be analysed and, therefore, result in too small standard errors (Heymans & Twisk, 2022).

Multiple imputation methods have several key advantages over single imputation (SI), particularly in terms of statistical validity, preservation of variability, and robustness to uncertainty (Alwateer et al., 2024; Austin et al., 2021; Chen & Chang, 2024; Walczak & Massart, 2001b; Wissler et al., 2022).

Sensitivity analysis (Cro et al., 2020; Thabane et al., 2013) is recommended to assess the robustness of the findings to plausible alternative assumptions about missing data.

In archaeology, missing data treatment can be applied for the imputation of unavailable data (Crema, 2025). Recently, they were used to treat incomplete measured data from damaged human skeletons (Pang & Liu, 2023; Wissler et al., 2022). The author and his colleagues have used multiple data imputation for non-available soil samples from Bibracte in the article **[9]**.

Censored data

Censored data have unknown values outside the measurement range (at one or both ends). Examples of the lower censoring limit (left censoring) are values below a certain threshold, for example the detection limit (such a concentration is usually reported as "<LOD" or "<DL"), while values above the upper censoring limit (right censoring) may be values lying above a given concentration X, for example corresponding to detector saturation (such a concentration is reported as ">X").

If the value is left censored, its value is unknown and can be anywhere between zero and the detection limit value (Helsel, 2009). This, of course, causes difficulties in exploratory and descriptive and inductive statistical data analysis (e.g., describing data using location and scatter estimates and testing them, diagnosing the shape of distributions using graphs and statistical tests, regression and correlation, etc.), requiring a complete ordering and alignment of the data that cannot be done in the presence of <LOD values.

Treatment methods of <LOD values are the most commonly used:

Deleting of <LOD values is not acceptable, as it shifts all statistical estimates toward higher values (Baccarelli et al., 2005; Hornung & Reed, 1990).

Replacement of <LOD values with constant, especially with LOD / 2 (eg, US EPA) or LOD / 2 (Baccarelli et al., 2005; Hites, 2019; Hornung & Reed, 1990) for a small number of nondetects (Zhang et al., 2004).

Kaplan-Meier estimation (KM), *regression on order statistics* (ROS), and *maximum likelihood estimation* (MLE), or MLE with bootstrap, are commonly used to calculate descriptive statistics (mean, median, percentiles, etc.) for left-censored data (Haslauer et al., 2017; Helsel, 2012; A. Singh & Nocerino, 2002; Zhao & Frey, 2006). Estimation of <LOD values by modelling the underlying distribution (e.g., Kaplan-Maier or Maximum Likelihood estimation, etc.) based on existing data (Helsel, 2012; Huybrechts et al., 2002; Shoari & Dubé, 2016, 2017) gives the most statistically satisfactory results (Tab. I).

Modern methods of censored data treatment are based on regression methods (Ahmadi et al., 2021; Boss et al., 2019; Lubin et al., 2004; Ortega-Villa et al., 2021), or on Bayesian estimations (Feroze & Aslam, 2022; Montes, 2024; Suzuki et al., 2020, 2025).

Tab. I Effect of <LOD values treatment on parameters of the univariate dataset (according to Baccarelli et al., 2005, modified).

Treatment	**Effect**	**Validity**	
		Arithmetic mean	**Standard deviation**
Discarding of values	No values below LOD	overvaluation	undervaluation
Substitution LOD/2	The values under LOD are equal to LOD/2	Significant bias does not occur for a small number of values below LOD for skewed distributions (skewness of 3 or more)	
Substitution LOD/√2	The values under LOD are equal to LOD/√2	No significant bias occurs for small number of values below LOD with less skewed distributions	
Imputation based on distribution	Values under LOD are calculated from a given distribution	No bias occurs for the number of values below the LOD below 60-70% + a slight to moderate deviation of the data from the chosen distribution	

Box plots EDA plots, or bar charts may be used as visualisation tools for univariate censored data (Barrett et al., 2014; Helsel, 2012; Shoari & Dubé, 2017).

In case of multidimensional data analysis, *substitution with LOD/2* is commonly used (Aruga, 1997, 2004; Farnham et al., 2002). *Random substitution* use values are randomly chosen from the interval between zero and the detection limit. It is recommended for unsupervised methods, such as cluster analysis (Aruga, 1997, 2004). None of the methods is applicable if the proportion of values <LOD exceeds approximately 25% (Farnham et al., 2002). For PCA data sets with <LOD below 10%, they are applicable; in the case of classification methods, the proportion of these values may be higher (Aruga, 2004).

Modern approaches in this area include covariance matrix estimation with maximum likelihood (Chung, 1993; Pesonen et al., 2015), or mixture model-based clustering (Liu & Brown, 2014; Wang et al., 2019). Methods for treatment datasets with both non-detects and missing data are also available (Faucheux et al., 2021; Stanimirova, 2013).

To visualise nondetects in the dataset (Grünfeld, 2005), similar graphical methods can also be used as for imputing missing values. The influence of <LOD values is more pronounced if they are distributed unevenly (inhomogeneously) in the matrix (Aruga, 2004).

Sensitivity analysis (Ahmadi et al., 2021; Haslauer et al., 2017; Thabane et al., 2013) should be performed to assess the robustness of the findings to plausible alternative assumptions about missing data.

In archaeometry, nondetects treatment occurs only rarely (Aruga, 2004), unlike the analysis of environmental and geochemical data (Helsel, 2005, 2009, 2012; Shoari & Dubé, 2017). Left-censored data treatment is used in several articles, not included in the thesis (Prokeš et al., 2004; Prokeš, 2012), where, however, only a small number of non-detects appeared in the datasets.

Relative data and compositional data analysis

Calculated variables are typically used in geochemistry (Nicholls & Russell, 2016). For example, in the case of columbites/tantalites values are commonly used, e.g. sums, differences, products and ratios of the variables, e.g. Mn/(Mn+Fe) or Ta/(Ta+Nb) (see e.g. article **[8]**), or more complex expressions (Thorpe et al., 1984). For these values, *mathematical coupling* (Archie, 1981; Curran-Everett, 2013; Kenney, 1982; Lenahan et al., 2011; M. R. Williams et al., 2022), should always be taken into account. Mathematical coupling (Fig. 5) usually causes spurious correlations (Brett, 2004; Härtel et al., 2022; Harter, 1963; Kenney, 1982; Lenahan et al., 2011; Pollman & Axelrad, 2014).

The dilution effect occurs often in geochemical, archaeometrical and environmental data. It is the result of an addition of inert diluent, which leads to unaltered ratios between the constituents of the mixture, but absolute quantities are changed (Aruga et al., 1993; Feenstra, 2006) (Fig. 6):

1. heigts or integrals of peaks in gas chromatography changes due to different volumes of injected sample.
2. the concentrations of water pollution concentrations at different distances from the source are changed due to the increasing dilution.
3. different absolute concentrations in pottery due to the amount of added ‘temper’ (extraneous substance, generally of nonplastic nature, quartz, calcite) to the original clay to improve the properties of the manufactured product.
4. differences in absolute concentrations of elements in soils due to different quartz content (Bern, 2009).

The dilution effect induces spurious correlation between variables and also one or more elongated clusters in related scatter plots. Thus, they appear in mixtures where the individual components differ from each other by values of constant ratios.

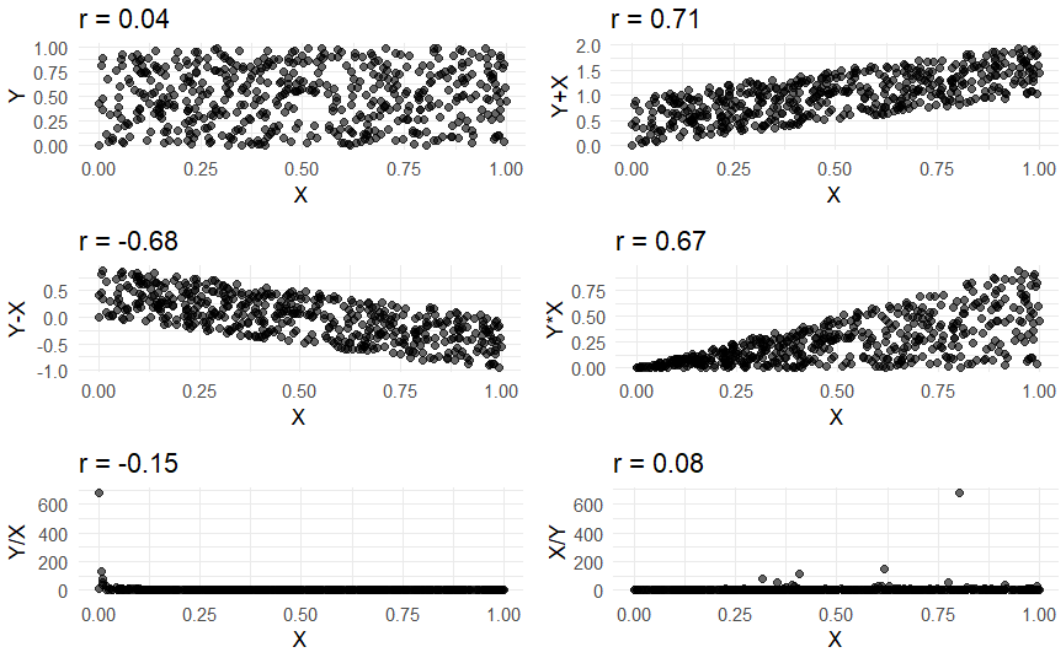

Fig. 5. Mathematical coupling and spurious correlations of two uniformly distributed variables (X,Y) and calculated variables. Values were generated randomly from uniform distribution.

This is very evident, for example, in scatter plots of elemental composition of Olmec pottery, where pottery samples from different sites are correlated with each other (Hancock et al., 2008). Similarly in scatter plots, the and in PCA scores plots V-shape of the data indicates more or less constant ratios, or correlation between individuals belonging to the same group, and the V-shape is thus at best indicative of constant ratio mixtures (Massart et al., 2001).

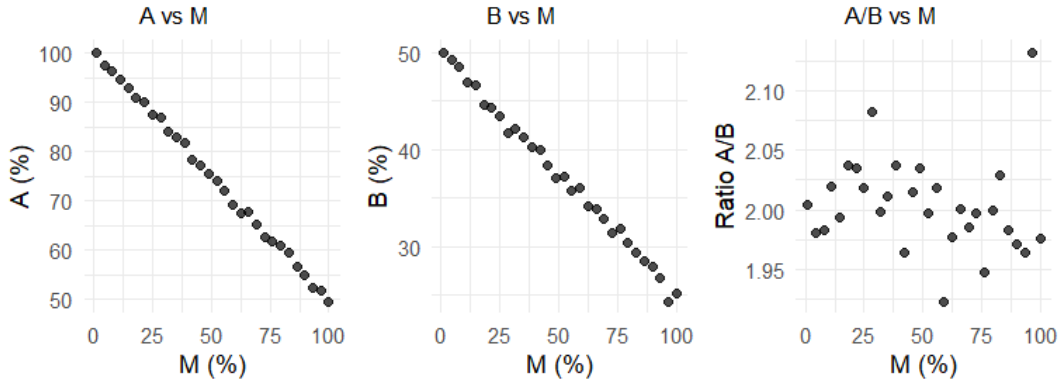

Fig. 6. An illustration of the dilution effect. Dependence of the concentrations of component A (left) and B (middle) and their ratio (right) on the inert component M content.

Ratios

For elimination of dilution effect elemental or isotopic ratios or various indices are commonly used in geochemistry, environmental chemistry and archaeometry (Anders, 1972; Barnes, 1990; Burton et al., 1999; Feenstra, 2006; Hancock et al., 2019; Michelaki & Hancock, 2011; Nicholls, 1988; Stanley, 2020; Terrill et al., 2022).

As in the simplest case with two normal distributions, estimations of mean and variance of the ratio are biased (Van Kempen & Van Vliet, 2000), various methods of mean (Rao & Beegle, 1967; Marsaglia, 2006; Srinivas et al., 2013), variance (Schuessler, 1974; Pham-Gia et al., 2006; Holmes & Buhr, 2007), and confidence intervals (Beyene & Moineddin, 2005) calculation methods were developed. Another method for estimating ratios is total least squares (TLS) linear regression (Curran-Everett, 2013; Moreno, 1996).

The use of ratio variables in correlation and regression analysis can give rise to spurious results due to inappropriate model specification and mathematical coupling, leading to serious misinterpretation of data and consequently to incorrect study conclusions (Y.-K. Tu et al., 2004). The use of ratios induces spurious correlation: if the correlations among X, Y, and Z are zero (i.e. $r_{XZ} = r_{YZ} = r_{XY} = 0$), the correlation among the ratios X/Z and Y / Z is not zero (Fig. 7). In practice, although the correlations of such ratios give apparently reasonable and interpretable results (Burton et al., 1999), the correlation coefficient values are spuriously biased. The Pearson formula and its modifications are used for the calculation (Atchley et al., 1976; Dunlap et al., 1997; D. A. Jackson & Somers, 1991; Kim, 1999; Van Der Weijden, 2002). The use of logarithmically transformed ratio vs. ratio scatter plots, also known as log-log plots, is advised in order to suppress this effect (Feenstra, 2006; Isles, 2020), however, this method is employed only on sporadic occasions (Burton et al., 1999; Feenstra, 2006).

The author of the thesis has used ratio variables in the results of the skeletal remains analysis results in articles not included in the thesis (Prokeš et al., 2004; Prokeš, 2012; Prokeš & Hegrová, 2006).

Percentage (closed) data

Closed data (or percentages) represent the relative proportions of the different parts that make up the whole. The parts themselves are not independent, and therefore their relative proportion is constrained to be a constant in the sum, usually 1 or 100%. In addition, compositional data includes all vectors representing parts of the whole that carry only relative information and thus includes not only parts per unit or percent, but also molar composition. Percentage transformation (closure) is manifested in changing units, e.g. from molar concentration to weight percents or ppm (Butler, 1979), but also for mg/L (Praus, 2019) or molarities (Otero et al., 2005). More often, however, compositional data appear less prominently: different components may be given in different physical units, different cases may give different sums, and almost never are all relevant components reported.

Closure (constant sum) also induce spurious negative correlations between major variables (with high %) and positive correlations between minor variables. If variables have different means and/or variances, the spurious correlations might be strong even if the number of variables is large. Spurious correlations also increase as the number of variables decreases (Chayes, 1960, 1962; Skala, 1979). In addition to the closure, the correlation also involves error propagation (Rietjens, 1995).

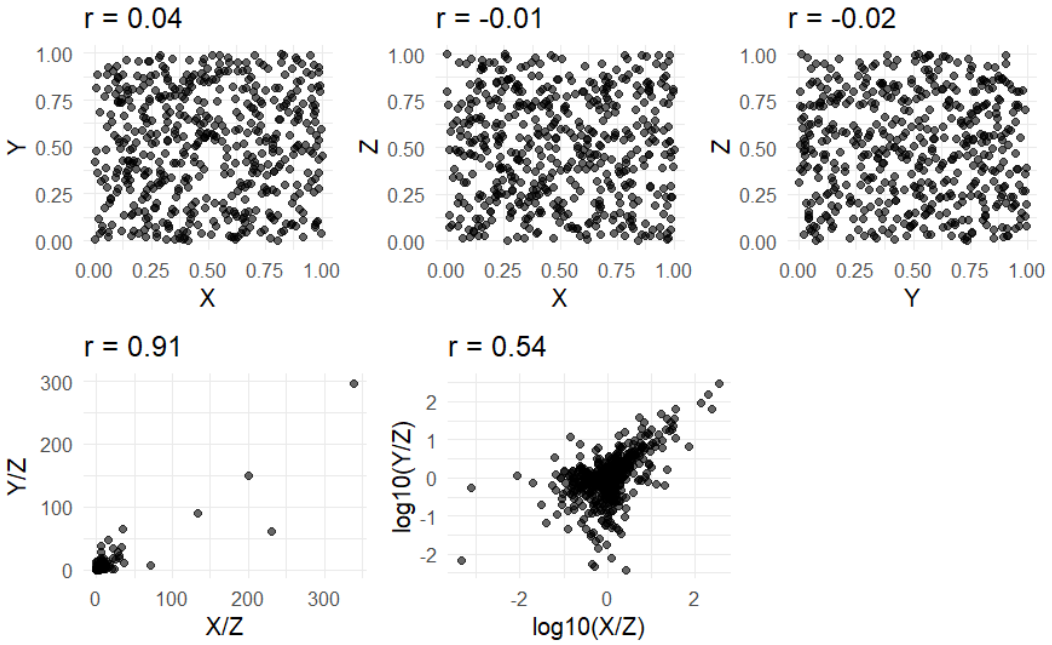

Fig. 7. Mathematical coupling influence on correlation of X, Y, Z and ratios X/Z and Y/Z. before and after logarithmic transformation. Values were generated randomly from uniform distribution.

Percentages can be calculated also from count data (Gloor et al., 2016, 2017; Lovell et al., 2020), typical for archaeology (see Shennan, 1988, Chapter 3; Solheim, 1960).

Compositional data concept and logratio transformation

Compositional data analysis (CoDA) is a set of methods for analysing data made of parts of a whole, like percentages, proportions, or concentrations, that always sum to a constant (e.g., 100% or 1) (Alenazi, 2023; S. N. Deming et al., 1993; Egozcue et al., 2024; Greenacre, 2021; Greenacre & Wood, 2024; Grunsky & Caritat, 2020). Compositional data do not belong to the classical Euclidean space, but they have their own Euclidean geometry on the simplex (Aitchison, 1986), even for univariate data analysis. Therefore, all classical methods based on Euclidean distance yield faulty results, if applied to compositional data.

Additive logratio (alr) transformation (Aitchison, 1986) – as logratio of all other components to one selected component (e.g. SiO_2 or Al_2O_3 in geochemistry), the number of columns in the data matrix is thus reduced by one.

Centered logratio (clr) transformation (Aitchison, 1986) – each variable divided by the geometric mean of all components before log-transformation. However, clr-transformed variables have some information overlap due to geometric mean usage.

Isometric logratio (ilr) transformation (Egozcue, 2003) - created orthonormal coordinates, preserving distances and geometry in the dataset, the number of columns in the data matrix is thus reduced by one, and direct relation to components is lost.

The pairwise log-ratio transformation (pwlr) (Greenacre et al., 2021) produces logarithms of the two-component ratios of a composition. They preserve relative information, but do not form an orthonormal basis. To link each pwlr to an orthogonal coordinate system, *backwards pivot coordinates* were developed (Hron et al., 2021), and the OPLAS algorithm can select orthonormal pwlr using Latin squares (Jašková et al., 2025).

Due to the logarithmic transformation in logratio computation, the concentration values should be positive. Zero or censored (<LOD) values should be imputed before the logratio transformation (Lubbe et al., 2021; Palarea-Albaladejo & Martín-Fernández, 2013, 2015). In addition, missing data can also be imputed (Hron et al., 2010).

Correspondence analysis of mean-normalised columns of the data matrix (Underhill & Peisach, 1985), which can handle zeros in the data without difficulty and give analogical results like PCA of logratio transformed data (Baxter, 1991; Baxter et al., 1990). If a power transformation is introduced into the correspondence analysis algorithm, then the limit of the power-transformed correspondence analysis is exactly the log-ratio analysis (Greenacre, 2009, 2010).

Some authors warn against mindless application of compositional data concept (logratio analysis):

- Trace element concentrations generally do not suffer from closure, but major oxide concentrations commonly do (Stanley, 2019).
- Only open systems in which an addition of a component physically crowds out others are subject to compositional closure (Pingitore & Engle, 2022).

Logratio analysis can be dominated by variables having low absolute presence and high relative variation that do not contribute to structure in the data but can obscure it. Traditional methods can detect certain kinds of structure in the data that correspond to structure on a ratio scale more directly than log-ratio analysis (Baxter, Beardah, et al., 2005; Baxter & Freestone, 2006).

Log-ratio analysis uses relative rather than absolute values of component concentrations. However, absolute differences may reflect the application of different raw materials or formulations. If these differences occur between variables with higher absolute values, they may not show high relative variability after the log-ratio transformation. Therefore, traditional analysis based on standardised concentrations can identify these differences quite directly (Baxter & Freestone, 2006). For example, temper content can be a feature that

distinguishes different types of ceramics, which can then be distinguished only using classical methods.

The concept of compositional data is now widely used in archaeometry (Baxter, 2008; Greenacre & Wood, 2024), and in geochemistry and environmental science (Grunsky & Caritat, 2020; Reimann et al., 2012). This concept was applied in articles **[9]** and **[10]**. Correspondence analysis was applied to obsidian data in the article **[2]**. This technique is very effective for visualisation of small datasets with $n_{col} > n_{row}$, where classical multivariate methods cannot usually be used.

Data with measurement uncertainty

'Uncertainty of measurement is a parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measure' (Ellison & Williams, 2013). Analytical data often contain the arithmetic mean and analytical uncertainty as standard deviation in the form mean ± sd, rarely as raw repeatedly measured values. In some cases, the uncertainty may be dependent on the relevant concentration (heteroscedastic uncertainty, small concentrations are more accurate than large ones), and incorrectly processed measurement samples may manifest as data heterogeneity (Stanley, 2006).

Measurement uncertainties can greatly attenuate correlations between variables, reduce statistical power of ANOVA, distort regression coefficients, underrate the explanation contributions of the most important factors in factor analysis (Duewer et al., 1976), and depreciate the significance of discriminant function and discrimination abilities of individual variables in discrimination analysis (Y. Liu & Salvendy, 2009). In PCA, the error-induced increase in variability is small for the components corresponding to the largest eigenvalues, but the measurement error may contribute to a large variability in loadings relative to the loading values (interpretation based on the loadings can be difficult) (Hellton & Thoresen, 2014). The results of the cluster analysis may also be affected by measurement uncertainties (Pankowska et al., 2025).

For the calculation of mean, variance, and confidence intervals for a data set, the grand (or pooled) mean and pooled variance are usually applied. For more complex calculations, uncertainty propagation methods (Chew & Walczyk, 2012; Farrance & Frenkel, 2012; Kragten, 1994; Stanley et al., 2010; Stanley & Lawie, 2007) should be used (Theodorsson, 2017; Zauner et al., 2025). Uncertainty can also be calculated for ratios (Hibbert, 2003; Holmes & Buhr, 2007).

Measurement uncertainty is usually neglected in multivariate data analysis, because it is assumed to be negligible compared to other sources of variability (Duewer et al., 1976; Zauner et al., 2025). However, its omitting can leads to overstated precision, misleading trends, unreliable classifications and clusterings, or invalid comparisons (Wentzell, 2013; Wentzell et al., 2018).

Weighting procedures can be used to include the effect of measurement error, for example, for Pearson's correlation coefficient (Saccenti et al., 2020), t-test (Durvasula et al., 2012), regression, and PCA (Andrews et al., 1996; Park & Klabjan, 2018). Measurement uncertainty may also be included in the calculation of the Mahalanobis distance for sample similarity estimation and data filtering (Beier & Mommsen, 1994) or factor analysis (Faber et al., 1993) by adding the uncertainty to the main diagonal of the covariance matrix.

Modern methods of uncertain data treatment are:

- *Simulation methods* (Cook & Stefanski, 1994; DeBruine & Barr, 2021; Duewer et al., 1976).
- *Maximum likelihood methods* (Tipping & Bishop, 1999; Wentzell & Lohnes, 1999).
- *Bayesian methods* (Gille et al., 2024; Nounou et al., 2002; Zuanetti et al., 2019).
- *Uncertainty-aware methods* (Chen et al., 2015; Gortler et al., 2020; Hagele et al., 2022).

Unfortunately, most of these methods are still not available in software packages.

Comparison of results for with and without measurement uncertainty included in the calculations for various data analysis methods (Duewer et al., 1976), and application of sensitivity analysis (Innes et al., 2021) are suitable for determining the influence of analytical uncertainty on the results of applied data analysis methods.

If raw repeatedly measured values are available instead of mean and standard deviation, calculations can be made with mixed linear models (Yu et al., 2022), for example, for Pearson correlation coefficient (Bakdash & Marusich, 2017; Hamlett et al., 2003; Shan et al., 2020), t-test or ANOVA (Muhammad, 2023; Yu et al., 2022).

Duplicate measurements are often used for the estimation of analytical precision (Hyslop & White, 2009; Stanley & Lawie, 2007; Thompson & Howarth, 1976) in geochemistry and medicine, especially.

In archaeology and archaeometry (Beier & Mommsen, 1994; Crema, 2025; Scott et al., 2007; Szpak et al., 2017), geochemistry (Stanley, 2006; Stanley et al., 2010; Stanley & Lawie, 2007), and environmental science, data with measurement uncertainties are typical for analytical results. The analysis of data with measurement error was treated by the author of the thesis in the article (Prokeš et al., 2025), not included in the annotated papers. For multivariate data in the annotated papers selected for the thesis measurements, uncertainties were omitted due to low values and a lack of software available for uncertainty treatment at the time of the article publication.

Methods for treating non-homogeneous data sets (e.g. outliers and multimodality), non-normally distributed data, and small data sets are commonly used and will not be discussed here.

3 Multivariate data analysis in archaeometry

Multivariate data analysis methods are essential in archaeometry because archaeological materials – ceramics, obsidian, metals, glass, pigments – are typically described by chemical or mineralogical compositions, which are inherently multivariate. These methods facilitate the identification of provenance, production groups, technological choices, and exchange networks.

Contemporary trends in multivariate data analysis emphasise the use of robust resampling techniques, such as bootstrap and cross-validation, to ensure reliable validation. Ensemble methods are used to enhance accuracy and robustness, while the integration of classical statistical methods with modern machine learning techniques facilitates interpretability and predictive power.

Data normalisation is one of the preprocessing approaches where the data is either scaled or transformed to make an equal contribution of each feature (Ahsan et al., 2021; De Amorim et al., 2023; V. N. G. Raju et al., 2020; D. Singh & Singh, 2020).

Dimensions reduction methods

Dimension reduction methods help reveal structure, provenance groups, and technological patterns.

Principal component analysis (PCA) (Aruga et al., 1993; Baxter, 2006, p. 20; P. A. López-García & Argote, 2023; Miriello & De Luca, 2025; Papageorgiou, 2020; Rácz et al., 2013; Remolá et al., 1993; Remolà et al., 1996), correspondence analysis (CA) (Peisach et al., 1982; Underhill & Peisach, 1985) and multidimensional scaling (MDS) (Peisach et al., 1982), or Sammon mapping (B. R. Kowalski et al., 1972) are among the most widely used. In contrast, factor analysis (FA) is not commonly used in archaeometry, unlike its use in the visualisation of archaeological data (Johnson, 1977; Lischka, 1975; Vierra & Carlson, 1981). In terms of modern methods, independent component analysis (IDA) was applied in archaeometry by Baxter (Baxter, 2006). Recently, this technique used for analyses of paintings (Cerrillo-Cuenca et al., 2021, 2024; Y. Liu et al., 2018; Schmitt et al., 2023), together with non-negative (or positive) matrix factorisation (NMF) (Y. Liu et al., 2018). The method of nonlinear principal component analysis based on auto-associative artificial neural networks (Kramer, 1991) has also been applied in archaeometry (Bitetto et al., 2016). Some of these procedures were also applied in the article **[2]**.

Other modern methods, like nonlinear kernel principal component analysis (Schölkopf et al., 1998; W. Wu et al., 1997), t-distributed stochastic neighbour embedding (t-SNE), uniform manifold approximation and projection (UMAP) (Healy & McInnes, 2024; Milošević et al., 2022) and Autoencoder deep learning (Mienye & Swart, 2025) have not yet been applied in the archaeometry area.

The selection of the optimal number of dimensions varies between methods. PCA relies on explained variance, correspondence analysis uses *inertia*, and MDS evaluates solution quality through *stress*. The optimal dimensionality is selected where additional dimensions bring only minimal improvement in these metrics. The results of the Procrustes analysis can also be used to select the optimal number of dimensions. The optimal number of dimensions

in ICA and NMF is determined on the interpretability of latent factors (the initial estimate may be based on PCA), while 2D / 3D visualisation is used as standard for the t-SNE and UMAP methods. Bootstrap confidence ellipses in CA (Ringrose, 2012) and PCA (Babamoradi et al., 2013) help decide whether dimensions are stable enough to interpret.

For quantitative comparison of object (or score) plots obtained with different visualisation methods (scatter plots, PCA- PLS- or CA- score plots, MDS and Sammon mapping plots) some similarity coefficients of topological preservation like RV-coefficient (Robert & Escoufier, 1976), or ITP-index (Bezdek & R. Pal, 1995) and Procrustes analysis results (Carlosena et al., 1995; Gonçalves et al., 2023; Vermeesch et al., 2023) are useful, but generally not used very often.

Clustering methods

Cluster analysis methods can be used to divide heterogeneous sets into a suitable number of groups.

Hierarchical clustering methods (Aruga et al., 1993; Baxter, 2006; Müller et al., 2018; Papageorgiou, 2020; Remolá et al., 1993) that reduce the dimension of the distance matrix using a dendrogram are very popular in archaeometry. To select a suitable clustering method, it is advisable to use the cophenetic correlation coefficient (Sokal & Rohlf, 1962). Other methods used in archaeometry include partitioning clustering methods, like k-means, k-medoids, and k-medians (Baxter, 2006), ISODATA (B. R. Kowalski et al., 1972), k-nearest neighbour (Peisach et al., 1982), fuzzy c-means clustering (Baxter, 2006), model-based clustering (or Gaussian mixture models) (Baxter, 2006; Papageorgiou et al., 2001; Papageorgiou & Liritzis, 2007), Kohonen self-organising maps (SOM) (Baxter, 2006; Lletí et al., 2003; Remolà et al., 1996) and most recently also databionic swarm clustering (P. A. López-García & Argote, 2023).

Modern clustering methods not implemented in archaeometry until now are, for example, density-based spatial clustering of applications with noise (DBSCAN) (Schubert et al., 2017), ordering points to identify the clustering structure (OPTICS) (Ankerst et al., 1999), genetic algorithm (GA) clustering (Robles-Berumen et al., 2024; Sheikh et al., 2008), or possibilistic c-means (PCM) clustering (Krishnapuram & Keller, 1993, 1996).

Cluster membership is either explicit (natural groups) or implicit (if membership is identified based on clustering results). For the fuzzy c-means, model-based, and possibilistic c-means clustering methods, calculated membership functions are available for each object.

Classification methods

Classification methods are used for the evaluation of either natural or isolated groups, or to predict membership of unknown samples within specified groups.

One-class classification methods (Brereton, 2011; Khan & Madden, 2014; Perera et al., 2021; Seliya et al., 2021; Strani et al., 2025) are often used to identify anomalies and outliers.

Two-class and multiclass classification differ in the number of categories into which the studied objects are classified.

The ensemble methods in classification methods combine several models to make better predictions than any single model alone.

- *Bagging:* trains models on different bootstrap samples and averages their results (e.g., random forest).
- *Boosting*: builds models sequentially, each focussing on errors of the previous one (e.g., XGBoost, AdaBoost).
- *Stacking*: combines diverse models and uses another model (meta-learner) to blend their outputs.

To prevent overfitting in classification methods (Defernez & Kemsley, 1997), cross-validation is a commonly used technique applied. For datasets where the number of descriptors >> number of cases, e.g. whole spectra, these methods are used:

Regularised methods are employed to control the complexity of the model. In order to prevent overfitting and improve generalisation, regularised classifiers add a penalty term to the loss function.

Sparse methods aim to produce models that use only a small subset of predictors. These measures improve interpretability and reduce noise. Sparse methods go further by forcing many coefficients to zero, creating simpler and more interpretable classifiers.

Data augmentation is a technique used to artificially increase the diversity of unbalanced datasets by creating modified copies of existing data, which helps to improve the performance of the machine learning model. By applying various transformations to the original data, data augmentation helps models generalise better and reduces the risk of overfitting.

Feature selection chooses the most informative variables for building a classifier. The goals are to improve accuracy, reduce overfitting, speed up computation, and increase interpretability (Bain et al., 2025; Pudjihartono et al., 2022; Sanchez-Pinto et al., 2018). Feature selection methods also prevent subjective variable selection due to representativeness bias.

- *Filter methods* select variables before training the classifier, based on statistical criteria (correlation / mutual information, variance thresholding, Relief / ReliefF, etc.).
- *Wrapper methods* evaluate subsets of variables using the classifier itself (forward selection, backward elimination, recursive feature elimination).
- *Embedded methods* perform feature selection inside the classifier during training, e.g. Lasso (L1) logistic regression (sparse model), *elastic net*, *random forest,* and *gradient boosting* (provide feature importance values), or regularised SVM.

A confusion matrix is used to evaluate the performance of a classification model for binary and multiclass classification problems. In addition to the confusion matrix, other techniques can also be applied to evaluate the classification. For two-category classification, metrics derived from the 2x2 confusion matrix (accuracy, precision, recall, F1 a F2 score), ROC curve, AUC-ROC score, Cohen kappa, etc. (Walsh et al., 2024). For multicategorial classifications, related modifications are available (Amirajlo et al., 2025; Ben Dor et al., 2023; Grandini et al., 2020; Hand & Till, 2001).

In archaeometry, the following techniques are mainly used for two-class and multiclass classification: logistic regression (LR) (Baxter, 2006; Ruschioni et al., 2023), linear discriminant analysis (LDA) (Anglisano et al., 2019; Baxter, 2006; Gajić-Kvaščev et al.,

2012; Miriello & De Luca, 2025; Rácz et al., 2013; Ruschioni et al., 2023) or quadratic discriminant analysis (QDA) (Baxter, 2006), Gaussian mixture discriminant analysis (Baxter, 2006), or flexible discriminant analysis (Baxter, 2006), soft independent modelling of class analogy (SIMCA) (Remolà et al., 1996; Remolá et al., 1993), partial least squares - discriminant analysis (PLS-DA) (P. López-García et al., 2018; Rácz et al., 2013), learning vector quantisation (LVQ) (Baxter, 2006), k-nearest neighbours (k-NN) classification (Anglisano et al., 2019; Baxter, 2006; Ruschioni et al., 2023; H. Sun et al., 2020), Naive Bayes (NB) classification (Eberl et al., 2023; Yrarrazaval et al., 2024), classification and regression trees (CART) (Baxter, 2006; Morisaki et al., 2024; Rácz et al., 2013; Ruschioni et al., 2023), support vector machine (SVM) (Baxter, 2006; Ruschioni et al., 2023; H. Sun et al., 2020) and artificial neural networks (ANN) using backpropagation, feedforward, radial basis or counter-propagation algorithms (Anglisano et al., 2019; Baxter, 2006; Kvascev et al., 2012; Ma et al., 2000; Ramil et al., 2008; Remolà et al., 1996; Ruschioni et al., 2023), convolutional neural networks (CNN) (Pawlowicz & Downum, 2021) and glmnet classification using generalised linear models with Elastic net regularisation (Eberl et al., 2023). The UNEQ method (Derde & Massart, 1986; Oliveri, 2017), similar to the SIMCA method, was applied for the identification of binders in historical paintings (Lletí et al., 2003).

From ensemble methods random forest (RF) (Anglisano et al., 2019; Ruschioni et al., 2023; H. Sun et al., 2020), adaptive boosting (AdaBoost) (H. Sun et al., 2020), and extreme gradient boosting (XGBoost) (Eberl et al., 2023) were used in archaeometry.

3.1 Classification of archaeological pottery

Petrography together with physical, chemical, and isotopic analyses carried out to solve important issues on provenance, production technology, dating, functionality, and conservation state of ancient ceramic materials. As for the provenance issue, constraining relationships between far away communities/societies can be evaluated. This complex methodological approach gives rise to the new concept of archaeometric analysis on ancient ceramic materials defined as “Archaeo-ceramic 2.0” (Maritan, 2019). Statistical methods plays a key role in ceramics archaeometry (Angourakis et al., 2018; Papageorgiou, 2020; Wilson, 1978).

Compositional variability of pottery from a particular production site is affected by (Hein & Kilikoglou, 2017):

- the natural variability of the raw materials used (Gualtieri, 2020; Hein & Kilikoglou, 2020)
- variation in the ceramic production process, such as clay processing and mixing of raw materials (Arnold et al., 1991);
- potential post-depositional alterations (see Maritan, 2020)
- limited number of samples, possible selection bias (Wilson, 1978)

Therefore, the main sources of compositional variability S_T^2 (overall variance of one group of samples originated from a given source) (Bieber et al., 1976; Bishop et al., 1990; Hazenfratz Marks, 2016; Hein & Kilikoglou, 2017; Pashkova et al., 2023) can be expressed as:

$$S_T^2 = S_N^2 + S_M^2 + S_S^2 + S_A^2$$

S_N^2 - "natural" variance due to raw material compositions
S_M^2 - manufacturing process
S_P^2 - potential post-depositional alteration,
S_S^2 - sampling variance due to the sample selection (Stanley, 2003)
S_A^2 - variance introduced in the analytical process (Stanley et al., 2010)

It is generally expected that $S_A^2 + S_S^2$ has no large effect on S_T^2 (Bieber et al., 1976; Pashkova et al., 2023), however, for limited number of samples, the sampling strategy also has a significant impact on the results (Hein & Kilikoglou, 2017).

Additionally, post-depositional processes cause contamination of both archaeological material (Buxeda I Garrigós, 1999).

The dilution effect in archaeological ceramics may occur due to temper addition (Aruga, 1998, 2003) or postdepositional contamination (Buxeda I Garrigós, 1999). Therefore, if the dilution effect is not negligible, the application of the compositional data approach is appropriate.

It is usually not possible to link archaeological ceramics directly to the clay sources used to make them. This is in part because clay is processed and mixed with other raw materials. The way the clay is mixed affects how easy it is to tell the difference between ceramics made in different places (Müller et al., 2018). In addition to the fact that there are many different types of clay deposits, the way they are analysed is also important to see how useful the results are. It is important to note that the ability to analyse distinguishing elements is crucial, particularly the precision with which these elements are determined. This is essential to avoid obscuring their natural variability (Bishop et al., 1982, 1990; Müller et al., 2018).

In the article **[1],** the organisation of Early Bronze Age pottery production in South Moravia (Czech Republic) is explored in terms of ethnography-based models, inferred jointly by pXRF multi-elemental analysis, micropetrography and powder X-ray diffraction. Petrographic groups were established by the Ward method of hierarchical cluster analysis with Euclidean distance (Fig. 2 of the article) of semiquantitative micropetrographical data (three groups: locally produced, from around the Brno batholith, and from the Svratka alluvium and the Carpathian Foredeep, see Fig. 1 of the article).

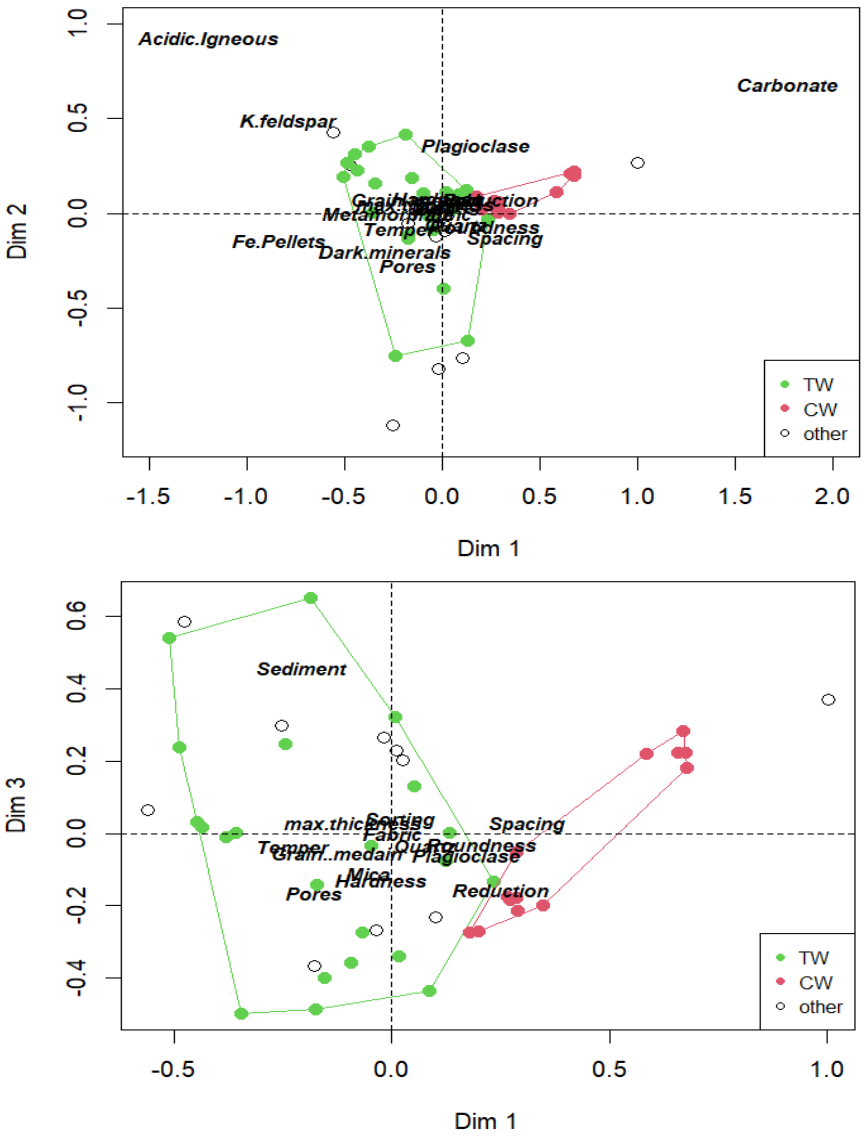

Fig. 8. Correspondence analysis of micropetrographical results: TW, tableware, CW, coarseware, other, unclassified samples.

Multielemental pXRF data was $\log_{10}$ transformed prior the analysis with linear discriminant analysis (LDA). LDA was applied first to the results obtained from samples of known micropetrography/provenance, and then the obtained LDA model was applied to samples of unknown provenance (Fig. 5 of the article). The concentration values of Ca, Ti, Mn, Fe, Ni, Cu, Zn, As, Rb, Sr, Y, Zr and Pb were chosen for the subsequent principal component analysis (PCA) of the $\log_{10}$ transformed and normalised data (Fig. 6 in the article). Archaeometry-based evidence indicates that tableware (TW) of a specific shape was produced using a specific technology at the site or in its close vicinity and that coarse ware (CW) was brought to the site from elsewhere. Correspondence analysis was used to visualise

the relationship between the pottery groups and micropetrographical parameters (Fig. 8), not included in the article.

Subgroups in multivariate data sets are usually established via hierarchical clustering (Badawy et al., 2022; Barone et al., 2005), model-based clustering (Jasiewicz et al., 2021; P. A. López-García & Argote, 2023), or SOM (Troiano et al., 2024). Other methods could also be used for this purpose, such as fuzzy clustering, DBSCAN, and others. For one-dimensional data, discretization can be used alongside cluster analysis to categorise continuous data (Garcia et al., 2013).

For estimation of stability of clusters various indexes and parameters, e.g. silhouette score, gap statistic, Calinski-Harabasz index, Davies-Bouldin index, etc., are used (Ashari et al., 2023; Ikotun et al., 2025). These values can be used for estimation of optimal number of clusters (Chicco et al., 2025), together with methods differ according to clustering methods: silhouette plot (k-means clustering), Akaike (AIC) and Bayes (BIC) information criteria (for model-based clustering), parameters *eps*, *minPts,* and k-distance plot (DBSCAN/OPTICS), and dendrogram inconsistency (hierarchical clustering). The Rand index, the adjusted Rand index and mutual information value can be used to compare the true assignments with the clustering assignments.

Bootstrap resampling can provide a more stable solution for cluster analysis (Levine & Domany, 2001). Ensemble methods can also be used to obtain more stable and accurate clustering solutions.

- *Bagging* (Bootstrap Aggregating) (T. Liu et al., 2022; Lubbe, 2024) trains multiple models on bootstrap samples and averages their output.
- *Consensus clustering* (Bodinier et al., 2025; Monti et al., 2003) is an ensemble method used to aggregate results from multiple clustering algorithms to produce a single, more robust clustering solution.

As an alternative to the above-mentioned methodology, methods using mixed mode data (continuous + categorical/binary) may be applied for joint analysis of chemical and petrographic data - clustering with Gower similarity coefficient or latent class models analysis (Baxter et al., 2008; Moustaki & Papageorgiou, 2005; Ownby et al., 2014; Rice & Saffer, 1982).

The results are interpretable without using compositional data methodology, as the dilution effect is related to the ceramic technology, and may be one of the factors distinguishing the two ceramic groups (see Baxter & Freestone, 2006). Compositional data methodology (for pottery analysis, see Barone et al., 2005; Buxeda I Garrigós, 1999; P. López-García et al., 2018; Wood & Greenacre, 2021) reflects rather the provenance in terms of the composition of the clay raw material. PCA and LDA (e.g., with ilr-transformation) may bring additional information to the results interpretation (Baxter, 1991; Baxter & Freestone, 2006; Hazenfratz et al., 2024).

Instead of the contents or concentrations of elements, selected spectral line intensities were used for provenance estimation of Roman pottery with ANN (Ramil et al., 2008). Unfortunately, direct usage of spectral line intensities excludes application of the compositional data approach. Whole spectra (XRF) were also used as input to radial basis-ANN (Kvascev et al., 2012) or random forest (Qi et al., 2018) for the classification of archaeological pottery samples.

3.2 Classification of obsidian

Obsidian is a naturally occurring volcanic glass formed when acid felsic lava extruded from a volcano cools rapidly with minimal crystal growth. It is hard and brittle (Ericson et al., 1975), therefore, it was used to manufacture chipped stone industry, especially blades. The provenance of obsidian has been extensively studied for more than 50 years (Frahm, 2025; Freund, 2013; Glascock, 2020; Kuzmin et al., 2020) in various parts of the world.

In the article **[2]**, the applicability of various multivariate visualisation and unsupervised classification techniques was tested on a set of obsidian samples from various regions. Principal component analysis and correspondence analysis combined with scatter plots and cluster analysis have been proven as efficient tools for the interpretation of multivariate geochemical data on obsidians. Independent component analysis, multidimensional scaling, and Sammon mapping have been shown to be only supplementary techniques, appropriate for visualisation and classification in this work. Hierarchical and fuzzy cluster analysis can be used to evaluate clusters in the data. The methods used were found to be to a large extent complementary.

Principal component analysis (Agha-Aligol et al., 2015; Bonizzoni et al., 2023; Harmon et al., 2023; Seccaroni et al., 2008; Syvilay et al., 2019; Thorpe et al., 1984), Sammon non-linear mapping (B. R. Kowalski et al., 1972) and hierarchical cluster analysis (Bonizzoni et al., 2023) involved in the article **[2]**, together with the supervised k-NN method (Mameli et al., 2023) or LDA (Seccaroni et al., 2008) were also used for obsidian provenance studies. Analysis of whole spectra (LIBS and XRF) combined with PCA and PLS-DA (Harmon et al., 2023), QDA (Syvilay et al., 2019) and the GA variables selection method (Mameli et al., 2023) was also applied for the estimation of obsidian provenance.

Until now, the concept of compositional data (CoDa) was not applied for obsidian provenance study, although it has been involved in archaeometric studies of flint or radiolarite chipped stone industry (Ben Dor et al., 2023; Brandl et al., 2014) or artificial archaeological artificial glass (Baxter, Cool et al., 2005). The suitability of this concept for obsidian study (see Pingitore & Engle, 2022; Stanley, 2019) should be consulted with experts on volcanic rocks in the future.

In article **[3]** LIBS spectrometry was applied on various samples of natural glassy material, tachylite (Czech Republic), pitchstone (Saxony, FRG) and obsidian (Czech Republic, Slovakia, Hungarian, Greek, Turkey and Ukraine) to obtain simple and fast method for their identification. Both spectral lines intensities (or their ratios) and element concentration (calculated from calibration lines) were used for this purpose. Simple graphical visualisation tools (radar graph, and Chernoff faces) and scatter plots (with error bars based on quantiles) indicate a significant compositional difference between tachylite and other glassy materials. Scatter plots also indicate the possibility of recognition of obsidians of various origin. Due to the limited number of samples, multivariate techniques were not applied here. This research should be verified on a larger sample set.

Carpathian obsidian represents the only significant source of naturally occurring volcanic glass in continental Europe, distributed across three distinct geological sources located in the eastern Carpathian Mountains (Hungary, Slovakia, and Ukraine). Exploited from at least 28,000 years BP through the Bronze Age, this material was traded across distances exceeding 400 kilometres (Bonsall & Kohút, 2025; Kohút et al., 2021). In the Czech Republic, obsidian artefacts appear in the Late Paleolithic period and in the Moravian painted pottery culture in the Early Neolithic period (Přichystal, 2018).

In the article **[4]** obsidian samples from Carpathian sources (see Fig. 1 in the article) and archaeological samples from Bohemia (upper Paleolithic / Mesolithic and Neolithic) were analysed with pXRF method. The importance of variables (elements) was determined with the random forests (RF) method (Fig. 9).

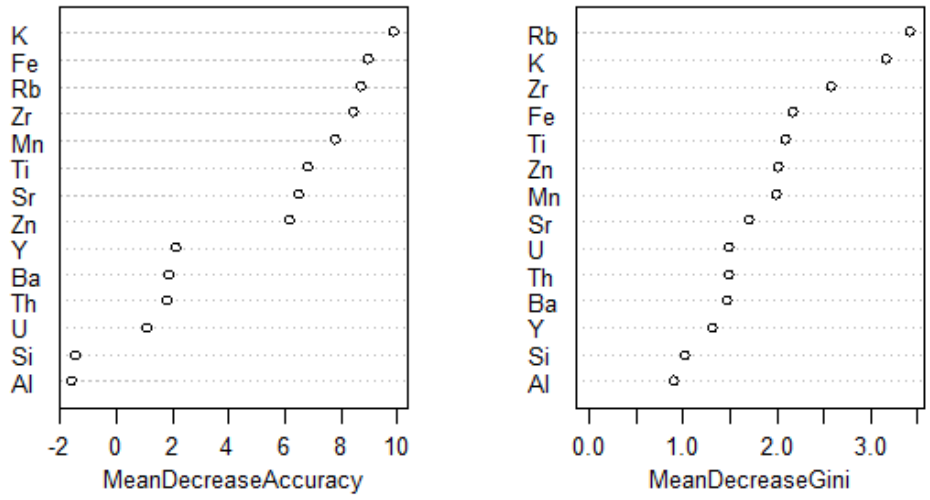

Fig. 9. Importance of variables using the random forest method

Scatterplots in the article indicate Slovakian origin of Czech obsidian artefacts, except for two samples from Kolín site. For them, Hungarian origin was proposed (Fig. 5 and 6 in the article). For differentiation between the Slovakia obsidian sources, nonlinear multivariate methods (e.g. ANN) should be applied on larger sample set.

Carpathian obsidian sources played a key role in the obsidian procurement in Paleolithic and Neolithic. Information concerning obsidian provenance is necessary for application of advanced mathematical models (e.g. Barge & Chataigner, 2003; Golitko et al., 2012; Ibáñez et al., 2016; Ortega et al., 2014), even reinforced with AI (Kerig et al., 2023), for characterisation of obsidian exchange and long-distance trade. Aggregation and meta-analysis of existing data from publications, reports, and databases (Eder et al., 2013; Frahm, 2014; Hancock & Carter, 2010) is also desirable.

3.3 Classification and identification of cultural heritage materials

Chemometrics also plays a role in the preservation of cultural heritage (Coccato & Caggiani, 2024; Lletí et al., 2003; Musumarra & Fichera, 1998; Riu & Giussani, 2024).

In the article **[5]** linear discriminant analysis techniques (LDA, maximum likelihood method, LOO cross-validation) and artificial neural networks (ANN, multilayer perceptron method) are compared for fast and simple identification of archaeological materials merely on the base of their spectra obtained by stand-off LIBS (this method has a big potential for archeological in-field measurements). Both data processing methods were applied on spectral lines intensities of LIBS spectra of 18 different samples (both archaeological and recent), containing seven types of material (i.e., shells, mortar, bricks, soil pellets, ceramic, teeth, and bones). As input data PCA scores (projection pursuit algorithm) were used. The use of PCA reduced the dimensionality of the data while preserving as much information as possible about the ensemble, ensuring independence of the variables and avoiding overfitting of the supervised classification methods used. The original set of 93 spectra was split into a training set (75 spectra) and a test set (8 spectra of independent samples). The remaining 10 spectra were used as a validation set for the ANN to optimise them. Both, LDA and ANN, methods are suitable for classification of archaeological materials based on stand-off LIBS spectrometry. Two other machine learning methods, support vector machines (SVM) and classification and regression trees (CART) methods, were tested, but their prediction success was poor (Tab. II).

Table II. Success rate of classification of various archaeological materials.

Test spectrum	Material	CART	SVM	LDA	ANN
11	soil	soil	soil	soil	**pottery**
12	brick	brick	brick	brick	brick
36	tooth	**bone**	**bone**	tooth	tooth
42	mortar	**bone**	**soil**	**soil**	mortar
43	bone	bone	**tooth**	bone	bone
50	shell	**bone**	shell	shell	shell
51	pottery	**brick**	pottery	pottery	pottery
93	tooth	tooth	tooth	**bone**	tooth
Prediction success	-	50.0 %	62.5 %	75.0 %	87.5 %

Therefore, these results were not included in the article.

Except for CART, most of the misclassifications were due to confusion between biominerals (bone, tooth) and geomaterials (soil pellet, brick, pottery) due to their analogous composition.

LIBS coupled with multivariate statistics has recently been applied as a tool in archaeology and cultural heritage science (Detalle & Bai, 2022; Donais et al., 2023; Pagnin et al., 2020). Recently, LIBS coupled with machine learning (Hao et al., 2024) or deep learning (C. Zhang et al., 2023) techniques is gaining importance for classification of materials.

In the article **[6]** 29 brick samples (with known firing temperature), made from soils from seven different localities, were analysed using stand-off LIBS and table-top LIBS setups. Principal component analysis (projection pursuit algorithm) was applied to selected spectral lines intensities in the LIBS spectra of brick samples. Considerably elongated clusters in PCA score plots (see Figs. 1-3, 5 in the article text) are consequence of different firing temperatures (see Figs. 4 and 6 in the article text). The firing-temperature dependence thus can also be used to estimate the firing temperature using LIBS for unknown samples in the future. Linear discriminant analysis was used as a visualisation tool for elimination of the influence of firing temperature on clusters shape in PCA. The cause of the elongated clusters may be an artefact of the analytical method (LIBS) due to temperature-driven changes in the physical properties of the materials (see Figs. 7-9 in the article text), because chemical composition does not change during pottery firing (Cogswell et al., 1996).

Selected spectral line intensities can be used if no suitable calibration CRMs are available. However, direct usage of spectral line intensities excludes application of the compositional data approach. Spectral line intensities were used, for example, for provenance estimation of Roman pottery with ANN (Ramil et al., 2008). Whole spectra (LIBS and XRF) were also used as input to radial basis-ANN (Kvascev et al., 2012) or random forests (Qi et al., 2018) for classification of archaeological pottery samples.

In the article **[7]** the manifestation of various pathways (photooxidation, chemical oxidation, alkaline and acid hydrolysis) of leather degradation was observed in ATR-FTIR spectra. Nine leather samples were treated with four procedures mentioned above and infrared spectra were taken. Because the whole spectra of a limited number of samples were used for data analysis, sparse method of principal component analysis (SPCA) was employed. Only photooxidated samples are distinguishable according to positive PC1 values (Figs. 7a and b in the article). Acid and alkaline hydrolysed samples were distinguished by PC2 values: positive values for acid hydrolysis and negaitive values for alkaline hydrolysis were found (Fig. 7a in the article). On the basis of PCA loading plots, it was possible to recognise specific regions of spectra related to the above-mentioned treatments. The amide bands (A_I and A_{II}) intensities (see Fig. 4 in the article), their distances $\Delta\nu(A_I–A_{II})$ and their A_I / A_{II} ratios were also employed for classification as boxplots (Fig. 5 in the article), scatterplots (Fig. 8 in the article) and the classification (CART) tree (Fig. 9 in the article). Due to a limited number of samples, these results should be verified on a larger sample set (especially for CART).

Dickinson and High (Dickinson & High, 2022) have obtained interpretable results concerning deterioration of vegetable tanned leather using non-sparse (classical) PCA, PCA-DA, and PLS-DA methods.

3.4 Provenance of conflict minerals

'Conflict minerals' refer to raw materials or minerals that come from a particular part of the world where a conflict occurs and affects the mining and trade of those materials. They currently include minerals like cassiterite, columbite-tantalite (coltan), microlite, and wolframite, sources of metals like tantalum, tin, or tungsten. Armed groups participate in the trade in these raw materials to fund a civil war in the eastern part of the Democratic Republic of Congo (DRC). Conflict minerals mined in the DRC may pass through numerous locations in neighbouring countries as they are shipped to processing facilities. They provide a major source of funding for warlords in the DRC region. *Regulation (EU) 2017/821 of the European Parliament and of the Council of 17 May 2017 establishing supply chain due diligence obligations for EU importers of tin, tantalum and tungsten, their ores, and gold originating in conflict-affected and high-risk areas* (known as Conflict Minerals Regulation) obliges EU importers to identify risks and mitigating responses in their supply chains [http://data.europa.eu/eli/reg/2017/821/oj].

In the article **[8]** classification and regression trees (CART) and random forests (RF) were used for selection of variables (elements in microlite samples) capable of distinguishing between raw material from various mines in the DRC area (see Fig. 8c- e in the article, Fig. 10 and Fig 11 here). The reference microlite grains were obtained from mines in the DRC (Numbi I, Numbi II, Rubaya), Rwanda (Kanyoni) and Mozambique (Morrua).

The results can also be applied for recognition of mixtures of smuggled microlite from DRC and microlite mined in neighbouring countries (especially Rwanda) in commercial samples, where cheaper smuggled raw materials from DRC may be mixed with legal material from other countries. Each commercial sample consists of a collection of columbite-tantalite, microlite, and cassitterite grains. The selected grains were mineralogically identified and then analysed with electron microprobe. The names of mines are used here (in contrast to the geographical names of the localities in the article **[8]**).

Kanyoni (Muhanga, Rwanda)
Numbi I (Mungwe, S. Kivu, DRC)
Numbi II (Misumari II, S. Kivu, DRC)
Morrua (Ile, Mozambique)
Rubaya (N. Kivu, DRC)

Visualization using CART identified key elements for classifying microlites from the monitored mines (Fig. 10).

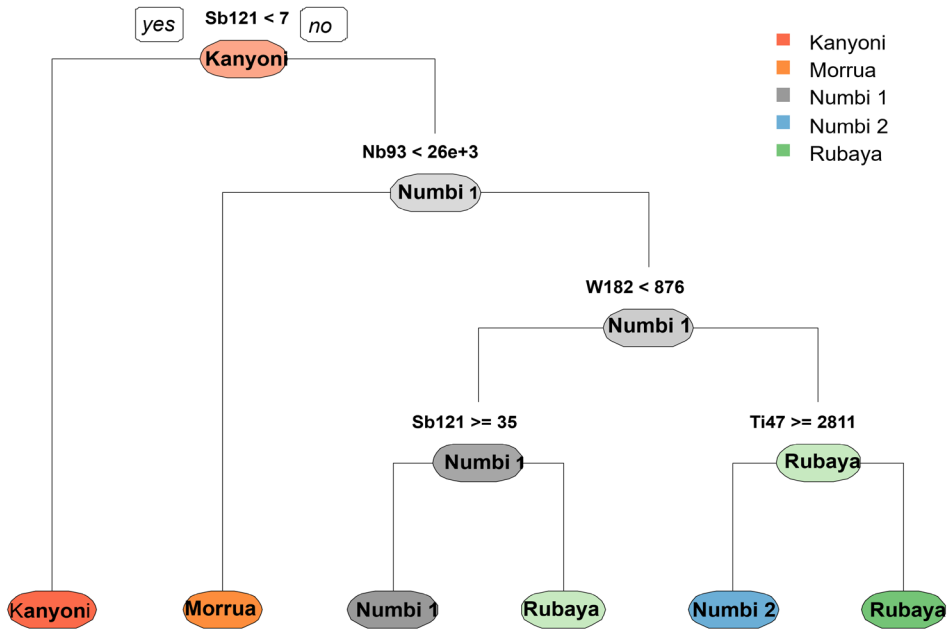

Fig. 10. Classification tree (CART) of microlite grains samples.

The prediction of commercial sample (ref. material) using the CART model indicates a high content of grains from DRC mines (Tab. III).

Tab III. Confusion table of classification of microlite grains samples based on CART

		original					
		Kanyoni	**Morrua**	**Numbi I**	**Numbi II**	**ref. sample**	**Rubaya**
predicted	**Kanyoni**	361	0	0	1	13	4
	Morrua	2	47	0	0	3	0
	Numbi I	0	0	77	0	18	0
	Numbi II	0	0	0	30	0	0
	Rubaya	4	2	2	0	511	99

Classification with random forest (RF) as an ensemble method yields more stable results in variable selection due to bagging (Fig. 11). However, unlike CART, it does not allow tree-like visualisation.

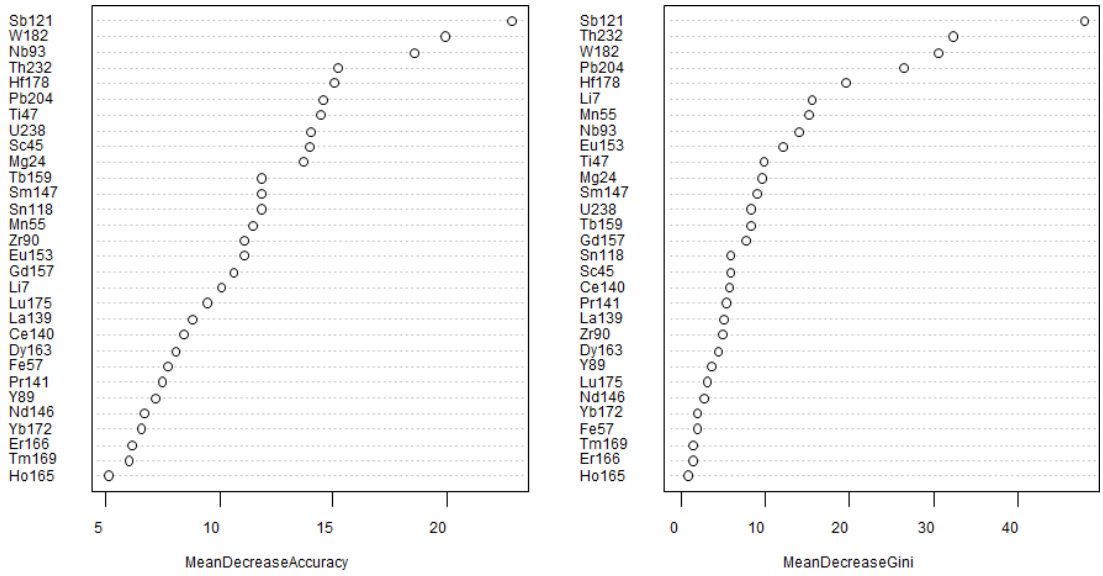

Fig. 11. Importance of variables using random forest method

Comparison of the confusion matrice for CART (Tab. III.) and RF (Tab. IV.) indicate better performance of RF, using all monitored elements.

Tab IV. Confusion table of classification of microlite samples based on the random forest method

		original					
		Kanyoni	**Morrua**	**Numbi I**	**Numbi II**	**ref. sample**	**Rubaya**
predicted	**Kanyoni**	367	0	0	0	9	0
	Morrua	0	49	0	0	2	0
	Numbi I	0	0	79	0	8	0
	Numbi II	0	0	0	31	0	0
	Rubaya	0	0	0	0	526	103

Suggested classification is consistent with data visualisation using scatter plot (Fig. 12) and boxplots (Fig. 13).

Elemental analysis combined with multivariate methods like PCA, PLS-DA, LDA or SVM (Gäbler et al., 2013, 2020; Harmon et al., 2011; Martyna et al., 2018; Savu-Krohn et al., 2011; C. Wang et al., 2022) got involved in attempts for source determination and samples similarities of conflict minerals.

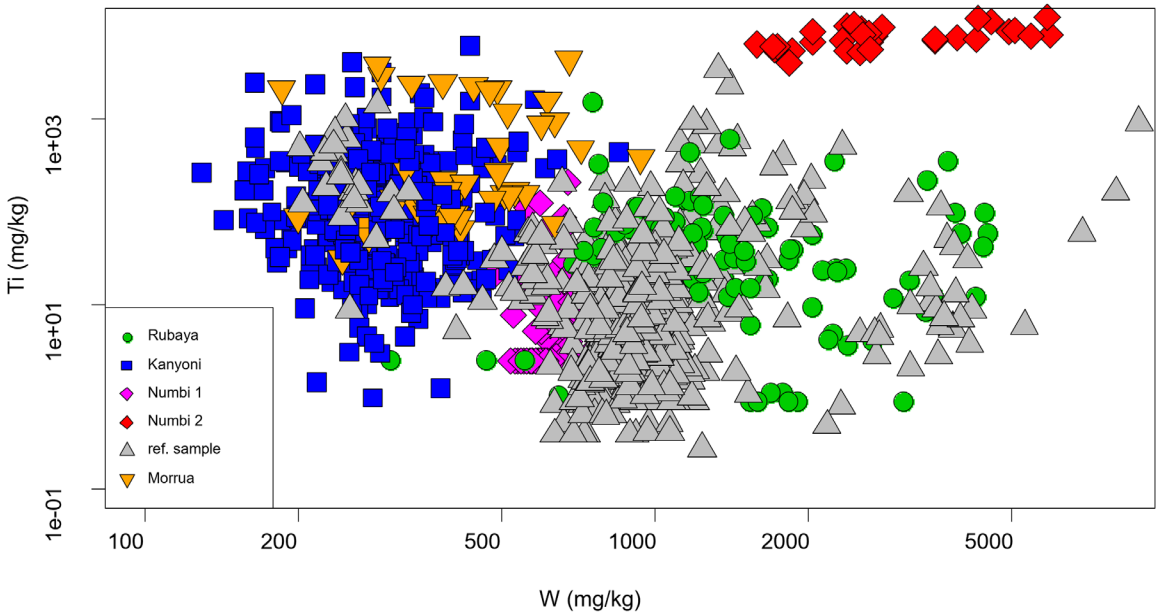

Fig. 12. Scatter plot of Ti andW concentrations (with commercial reference solvent)

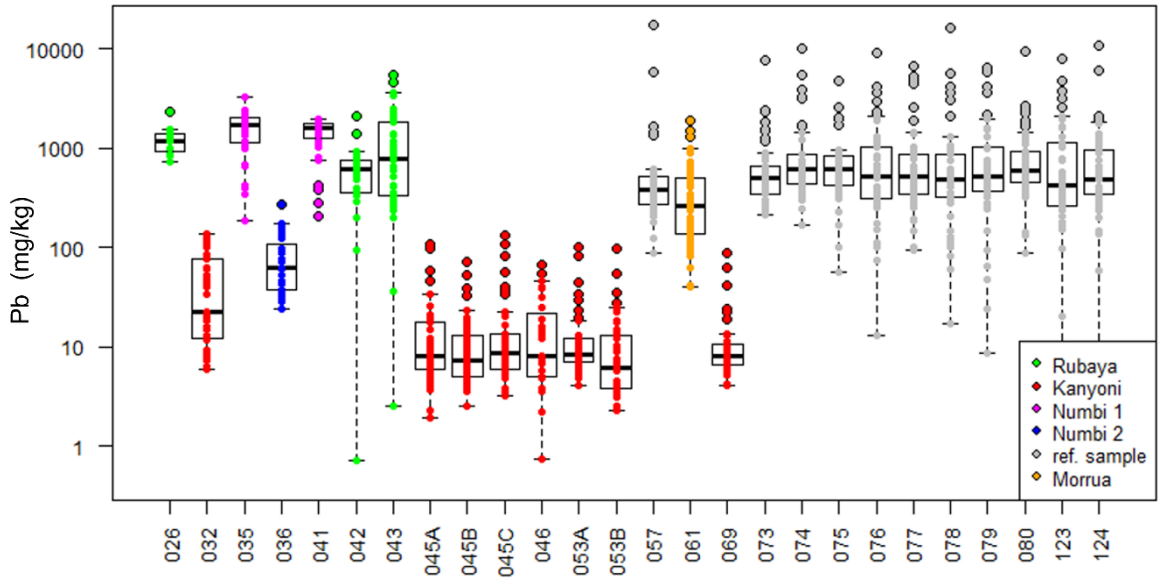

Fig. 13. Box plot of Pb concentrations (comparison with commercial reference standards)

Application of classification methods based on distances (cluster analysis, multidimensional scaling, Sammon mapping) is possible with similarity measures between samples (sets of grains) distributions using Kolmogov-Smirnov distance (Ben Dor et al., 2023; Gäbler et al., 2020), distance between kernel density estimators KDE (Sircombe & Hazelton, 2004), Bray-Curtis and Aitchison central logratio distance (Vermeesch et al., 2016), and many others (see Cha, 2007).

Until now, the compositional data (CoDa) concept was not used for the study of conflict mineral provenance.

Recently, Bayesian methodology (through the calculation of likelihood ratios) has also been applied in this field (Gäbler et al., 2020; Martyna et al., 2018).

When implemented rigorously, the combination of elemental analysis and Bayesian statistics has been shown to enhance the objectivity of forensic science, thus reducing the risk of wrongful conviction resulting from the misinterpretation or misunderstanding of trace evidence. The Bayesian likelihood ratio (LR) methodology represents a significant advance in forensic science, as likelihood ratio frameworks provide principled and transparent methods for communicating the probative value of multivariate elemental evidence (Akmeemana et al., 2021, 2022; Van Es et al., 2017). However, without rigorous uncertainty quantification, assumption transparency, and judicial education on Bayesian interpretation, high LR values risk misleading courts. In light of contemporary best practices, expert reports must now include assumptions lattices, uncertainty pyramids, sensitivity analyses, and explicit disclaimers regarding crime-level conclusions derived from source-level analyses (Lund & Iyer, 2017).

4 Elemental soil analysis in archaeology

Elemental analysis of soils and sediments provides information to trace past human activity at and around archaeological sites. Therefore, soil geochemical sampling is becoming a relatively common feature of archaeological excavation and surface surveys (see Bintliff & Degryse, 2022).

In the article **[9]** magnetic susceptibility and chemical element concentrations were used to identify various historical uses of selected areas of Celtic oppidum Bibracte (France) (Fig. 1-3 of the article). The interpolation of spatial data (Fig. 4, 10-11 in the article) and filling in missing values for the current variable were performed by inverse distance weighting (IDW) method. Missing values (in 5, 15, 25 and 35 cm) were obtained from these interpolated values. Robust compositional principal component analysis was performed after the centred log-ratio (clr) transformation (Fig. 6-9 in the article). Correlation statistics were calculated and visualised using nonparametric Spearman correlation coefficient (Fig. 5 in the article). Multivariate spatial analysis of the cultural layer revealed different zones associated with different activities and functions: PC1 is linked to heavy metals, such as Pb, Zn, and Cu, which may in archaeological contexts indicate some kind of metalworking activity and to Ca, S, and P relating mostly to organic matter, body waste, ashes, and bones. PC2 (Rb, K, Ti, Mn and Al) possibly reflects elements of geogenic nature. PC3 indicates a negative correlation between the hydromorphic / pedogenic feature (Mn with positive values) and some elements (S, Ba, Sb and P) associated with human activities concentrated in specific features. The Global Moran I index indicated specific spatial distribution of PC1 elements, such as Zn or P.

In article **[10]** analogical methodology (except for spatial autocorrelation) was also used for the identification of activity areas in the Pohansko Early Medieval stronghold (South Moravia, Czech Republic).

Projection methods are often used as visualisation and preprocessing techniques for clustering and classification methods in geochemistry. Principal component analysis (PCA) is the most widely used projection method for geospatial data analysis (Gajić-Kvaščev et al., 2012; Ma et al., 2000; Michael et al., 2025; Pîrnău et al., 2020; Sadeghi et al., 2024; Srivastava et al., 2024). However, PCA results may be biased due to spatial autocorrelation; therefore, autocorrelated and spatial weighted PCA (Fernández et al., 2018; Harris et al., 2011) should be preferred. Instead of PCA some other projection methods like correspondence analysis (Ji et al., 1995, 2007; Mellinger, 1987; Teil, 1975; Valenchon, 1982), factor analysis (FA) (Davis, 2002; Miesch, 1979; Reimann et al., 2002; Tripathi, 1979; R. Wu et al., 2020), and multidimensional scaling (MDS) (Pan & Harris, 1991; Song et al., 2022; Vermeesch et al., 2023) are used in geochemistry and geochemical mapping, but much less frequently.

Modern projection methods used for geospatial data processing are, especially, ICA (Shahrestani et al., 2024; Yang & Cheng, 2015a, 2015b), NMF (Kuwatani et al., 2025; Saylor et al., 2019; Zekri et al., 2019), kernel PCA (Esmaeili et al., 2020; B. Liu et al., 2016),

t-SNE (Horrocks et al., 2019; H. Liu et al., 2021), UMAP (Mpaka & Von Der Heyden, 2024; Q. Zhang et al., 2024) and Autoencoder (S. Zhang et al., 2019).

Clustering methods are commonly used in the analysis of geochemical data (Templ et al., 2008). In the case of spatially dependent data analysis, cluster analysis can also be used for spatial segmentation (see elemental imaging). Instead of classical methods like herarchical clustering methods (Templ et al., 2008), partitioning clustering methods (k-means, k-medoids and k-medians) (Esmaeiloghli & Tabatabaei, 2020a) and ISODATA (S. Zhang et al., 2021), methods such as Gaussian mixture models (Y. Chen & Wu, 2019; Lucero-Álvarez et al., 2021; F. Zhang et al., 2025), Expectation Maximisation (EM) (Esmaeiloghli & Tabatabaei, 2020a) and fuzzy c-means (Ghezelbash et al., 2025) are employed for geochemical data clustering. Univariate model-based cluster analysis was used in archaeological phosphate prospecting (Prokeš et al., 2013), as a method for estimation of the backgroung concentration value, instead of the classical distribution modelling (Portier, 2001; Rodríguez et al., 2006). Modern clustering methods such as DBSCAN (Esmaeiloghli & Tabatabaei, 2020b; Ghezelbash et al., 2025; S. Zhang et al., 2019), SOM (Esmaeiloghli & Tabatabaei, 2020a), and the databionic swarm (Engle & Chaput, 2023) are increasingly being used to cluster geochemical data.

Supervised and semi-supervised methods are used for anomalies and outliers detection, k-nearest neighbour (Y. Chen et al., 2021; Y. Chen & Lu, 2023; Oonk & Spijker, 2015), SVM (Oonk & Spijker, 2015), RF (S. Zhang et al., 2024), artificial neural networks (Ginau et al., 2020; Oonk & Spijker, 2015) and one-group classification methods, especially random forests (S. Zhang et al., 2024), isolation forest (Amirajlo et al., 2025; Saremi et al., 2025; Shahrestani & Carranza, 2024; Shahrestani & Sanislav, 2025), one-class SVM (Amirajlo et al., 2025; Shahrestani & Carranza, 2024), and local outlier factor (Shahrestani & Carranza, 2024; Shahrestani & Sanislav, 2025).

Spatial data analysis methods use instead of commonly used variogram and semivariogram (Davis, 2002; Reimann et al., 2008) also local indicators of spatial association (LISA) using local Moran I or Geary c autocorrelation (Anselin, 1995; Yuan et al., 2018), local Getis and Ord G based on distance statistics (Getis & Ord, 1992; Ord & Getis, 1995), and local Lee L (Lee, 2001) integrating Pearson's *r* and Moran I.

The logratio transformation, used in the articles **[9]** and **[10]**, is often applied in geochemical analyses (Ginau et al., 2020; McKinley et al., 2016; Pawlowsky-Glahn & Egozcue, 2020; Reimann et al., 2012), instead of classical normalisation, to prevent the dilution effect. Logratio transformation is suitable for application with multivariate and machine learning methods, especially PCA, and cluster analysis (Brückner & Heethoff, 2017; Greenacre & Wood, 2024).

5 Elemental imaging (mapping)

Elemental mapping is the process of creating images that demonstrate the distribution of various elements within a sample. This technique is of particular importance in fields such as medicine, environmental science, materials science, geology, and chemistry, as it assists researchers in comprehending the composition and structure of materials at the microscopic level. Elemental map is related to the so-called width-format of data table. Before data analysis, transformation to long format (xyz) of the table should be made.

Hyperspectral mapping is used to map spectral data (Fonville et al., 2013; Gut et al., 2015; Palmer et al., 2015; Verbeeck et al., 2020), where typically $p >> n$. Here regularised (sparse) methods (Filzmoser et al., 2012) should be used.

In the article **[11]** data analysis was performed using carbon (^{12}C) normalised values. The de-spiking of the line scans was made using a three-point median Tukey filter, and subsequently a 15-point Hampel filter was applied to remove outliers. Visualisation of the filtered data was performed using two-dimensional elemental maps (Fig. 2 and 3 in the article). The distributions of particular elements were visualised using kernel density estimators (Figs. S1 and S2 in the article). The bimodality of the distributions is due to the different composition of sampling points on the leaf and leaf and microscopic glass slide (Fig. 2 and 3 in the article). Differentiation between microscopic glas and the leaf is also possible also via segmentation. For this purpose, k-means and fuzzy c-means clustering (for 2 clusters) of autoscaled data were applied. Both clustering procedures yield analogical clustering result as compared by means of confusion matrix (Fig. 14).

		fuzzy c-means	
		1	2
k-means	1	906	0
	2	1	2893

Fig. 14. Segmentation of the data for differentiation between microscopic glas and the *Noccaea praecox* leaf. A. k-means clustering, B. fuzzy c-means clustering.

The elemental maps of the $^{112}Cd/^{12}C$ ratio in the leaves were explored using some above-mentioned statistical procedures for spatial data (with "queen adjacency"). These measures give an indication of areas of spatial nonhomogeneity through the extent of significant spatial clustering of similar values. The levels displayed reflect the degree of similarity of the ratio value in a given point with the ratio values at the surrounding points.

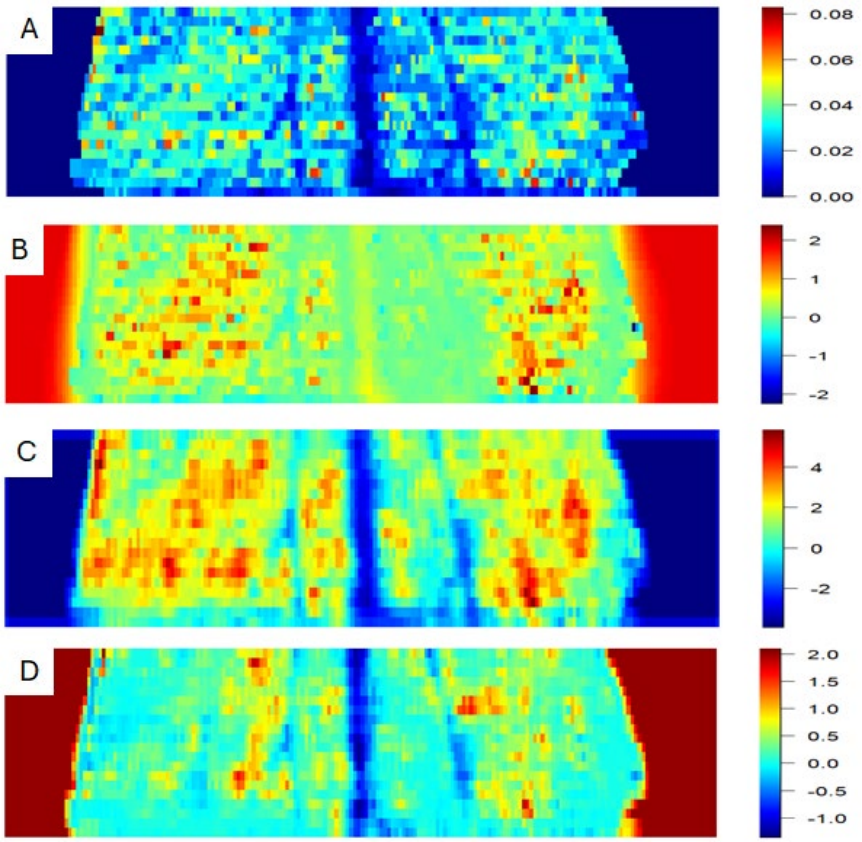

Fig. 15. Spatial autocorrelation of the $^{112}Cd/^{12}C$ ratio in the *Noccaea praecox* leaf. A. $^{112}Cd/^{12}C$ ratio, B. LISA, C, Local G, D. Local Lee L

Local measures indicate a high homogeneity of the $^{112}Cd/^{12}C$ ratio in leaf veining and glass, probably due to the very low level of this ratio. However, no additional information regarding the distribution of heavy elements was provided, and the interpretation of elemental maps was conducted with these indicators.

Principal component analysis was applied to the autoscaled data matrix (Fig. 16). The first principal component is microscopic glass especially, because of the high loading value of Si. The second principal composition represents plant tissue, due to the magnesium content in chlorophyll. The third principal component loading is very low, in comparison with both previous.

Together with logarithmic transformation (Rittner & Müller, 2012), closing data and subsequential logratio transformation can also be used for elemental mapping (Pesenson et al., 2015) if concentration values are used.

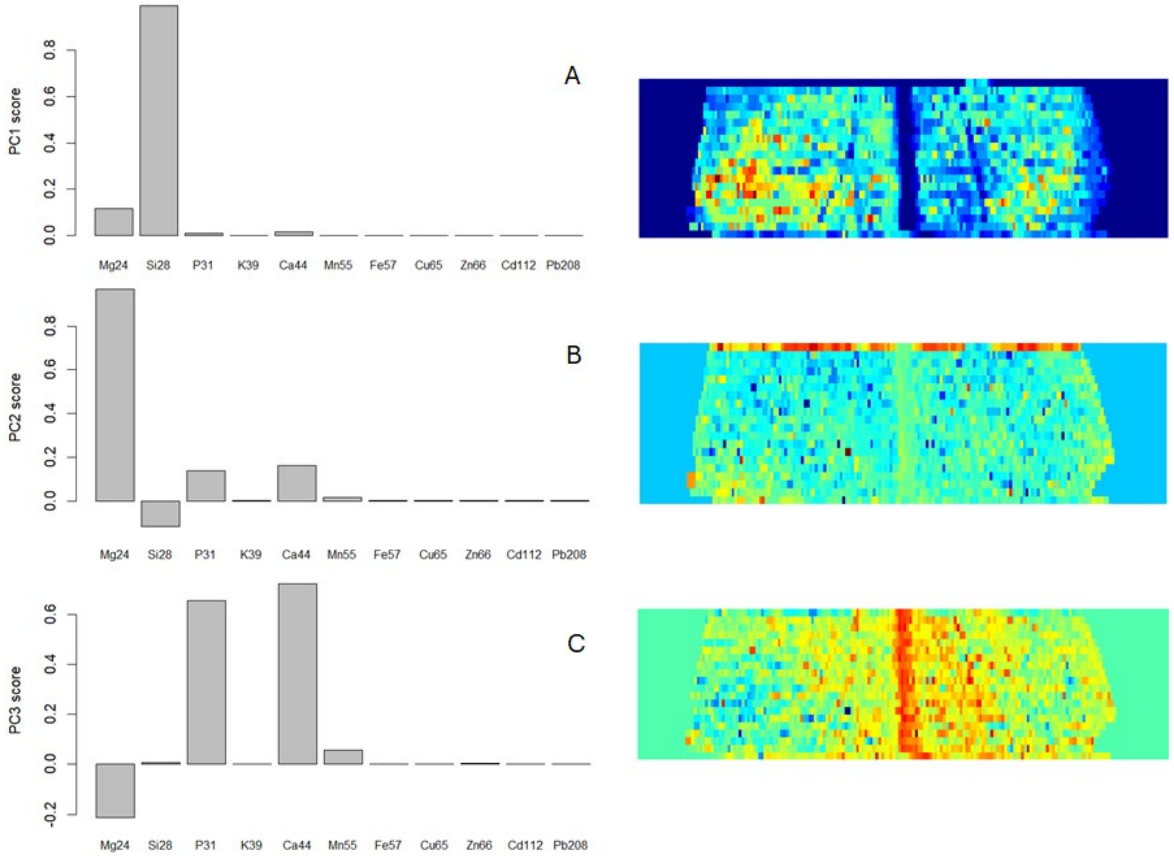

Fig. 16. Results of principal component analysis: loadings plots and score maps.

From the ***projection methods,*** PCA is the most used projection method for chemical imaging (Aoyagi, 2023; Gut et al., 2015; E. A. Jones et al., 2012; Liebich et al., 1992; McCombie et al., 2005; B. Paul et al., 2014; Prasad et al., 2022; Verbeeck et al., 2020), sometimes with Varimax rotation (Verbeeck et al., 2020). Autocorrelated and spatial weighted PCA (Harris et al., 2011), eliminating the influence of spatial autocorrelation on PCs, was not applied for image treatment until now. ICA (Gut et al., 2015; Verbeeck et al., 2020), NMF (Gut et al., 2015; E. A. Jones et al., 2012; Prasad et al., 2022; Verbeeck et al., 2020), maximum autocorrelation factorisation (E. A. Jones et al., 2012; Verbeeck et al., 2020), t-SNE (Balamurali et al., 2019; Prasad et al., 2022), UMAP (Smets et al., 2020; W. Zhang et al., 2021), multivariate curve resolution-alternating least squares (MCR-ALS) (Aoyagi, 2023; Gut et al., 2015; Jaumot & Tauler, 2015; Verbeeck et al., 2020), probabilistic latent semantic analysis (E. A. Jones et al., 2012; Verbeeck et al., 2020), random projection (Palmer et al., 2015; Varmuza et al., 2011) and some other methods (Verbeeck et al., 2020) were used for mapping elemental and hyperspectral data.

For ***segmentation*** of elemental maps or maps of their principal / independent components, various clustering procedures were employed, for example hierarchical clustering (Banas et al., 2013; Castellanos-García et al., 2020; E. A. Jones et al., 2012; Liebich et al., 1989, 1992; McCombie et al., 2005), k-means clustering (Castellanos-García et al., 2020; E. A. Jones et

al., 2012; McCombie et al., 2005; Palmer et al., 2015; Prasad et al., 2022; Verbeeck et al., 2020), ISODATA (McCombie et al., 2005), Sammon nonlinear mapping (Liebich et al., 1992), fuzzy c-means clustering (E. A. Jones et al., 2012; B. Paul et al., 2014), and SOM (Balamurali et al., 2019; Franceschi & Wehrens, 2014; Verbeeck et al., 2020). Other methods, such as DBSCAN (W. Zhang et al., 2021) and Gaussian mixture models (Prasad et al., 2022), could provide interesting results in the field of element mapping. In the case of fuzzy and model-based methods, calculated membership functions can also be retransformed to map (B. Paul et al., 2014). The segmentation results can then be processed using classification methods (Danzer & Singer, 1985; Graham & Castner, 2012; Liebich et al., 1992). Regression methods, like polynomial regression (Liebich et al., 1989; Singer & Danzer, 1984) or random forest (Hanselmann et al., 2009; Shamraeva et al., 2024), can also be applied with intensities or concentrations as dependent variables and x and y coordinates as independent variables. RF regression is analogous to CART (recursive partioning) regression (Vignon, 2015), which can be an interesting alternative to other regression methods (Fig. 17).

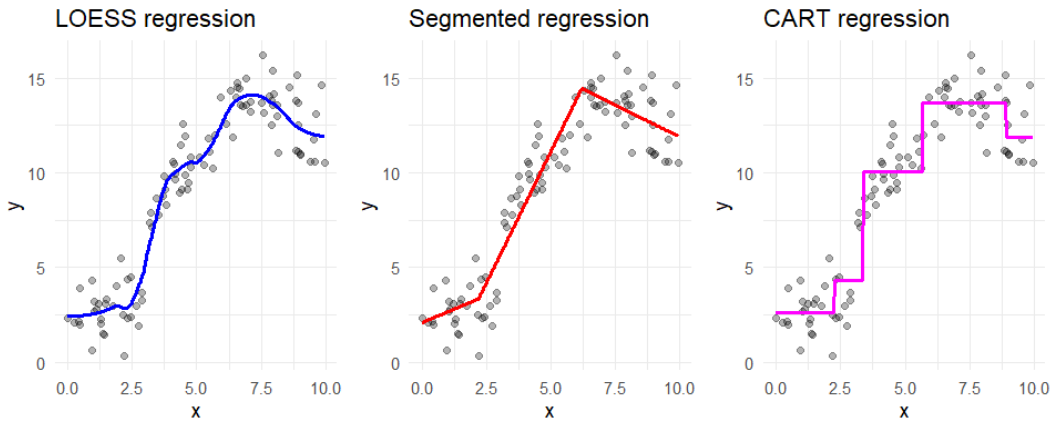

Fig. 17. Comparison of local regression (LOESS), piecewise (segmented) regression, and CART regression.

As for ***spatial data analysis methods***, only LISA was applied in textile industry (Militký & Meloun, 2000). Lag plot and plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). ACF was proposed for homogeneity evaluation with line scans in material analysis (Doerffel et al., 1990; Scholze et al., 1994). The Box test in two variants, Ljung-Box test or Box-Pierce test, serves for evaluation of results obtained from autocorrelation plots. The periodic trend can be detected with a periodogram plot (spectral analysis using FFT) or with a Fisher or Siegel test. The CUSUM indication of tendencies was developed as a statistical tool for material analysis, the Global CUSUM test (Liebich, 1990; Liebich et al., 1989) as an indicator of homogeneity, and the CUSUM shape analysis (Doerffel et al., 1991) is a graphical tool suitable for a high number of measured points ($n >$ 150).

6 Analysis of mass spectra

Laser desorption/ionisation time-of-flight mass spectrometry (LDI-TOF MS) has emerged as a powerful analytical technique for characterising inorganic materials and metal/metalloid clusters.

The article **[12]** presents a new open-source workflow in R designed for the advanced interpretation of LDI TOF mass spectra of inorganic materials. The set of scripts in the R programme includes procedures (e.g., decomposition of mixed spectra into individual components, visualisation methods, and others) that were used in the interpretation of mass spectra and determination of the stoichiometry of inorganic clusters (see the author's publications not included in this thesis).

In the article **[13],** a method for calibrating TOF mass spectrometers using monoisotopic gold and phosphorus is described. This article includes spectral fitting using a Gaussian model (Romanenko et al., 2006) to refine the estimation of *m/z* values corresponding to the maximum spectral line. A simple Gaussian model proved to be sufficient in this case. It was therefore not necessary to use other, computationally more complex models like the Lorenz (or Cauchy) model (Romanenko et al., 2006), pseudo-Voigt model (Meier, 2005), or Pearson VII model (Hall et al., 1977).

Mass spectrometry (MS) has emerged as a powerful analytical platform for high-throughput profiling of biological molecules, generating complex, high-dimensional datasets that require sophisticated statistical and machine learning approaches for meaningful biological interpretation. The integration of multivariate statistics and machine learning into MS workflows has been demonstrated to be an effective solution to critical challenges in data preprocessing (J. Sun & Xia, 2024; Tseng et al., 2025), feature extraction, classification, and biomarker discovery (A. G. Beck et al., 2024; Liebal et al., 2020).

The following methods are the most commonly used for multivariate statistical analysis and machine learning in the context of mass spectrometry results: principal component analysis, logistic regression, linear discriminant analysis, partial least squares discriminant analysis, k-nearest neighbour, support vector machines, classification and regression trees, random forests, gradient boosting, and neural networks (feedforward, convolutional, and recurrent) (A. G. Beck et al., 2024; Godmer et al., 2025; Gromski et al., 2015; Idkowiak et al., 2025; Liebal et al., 2020; Meyer, 2021; Ng et al., 2023; A. Paul & De Boves Harrington, 2021; Rajalahti & Kvalheim, 2011; Rohart et al., 2017; Trygg et al., 2007). For cluster analysis of mass spectra various similarity coefficients (Alfassi, 2005) can be utilised.

The article **[14]** provides a method for evaluating cell culture cross-contamination levels based on mass spectrometric fingerprints of intact mammalian cells. The mass peaks with the highest intrinsic variability in the datasets were selected by:

(1) comparative determination of the standard deviations of individual peaks normalised to the total variance of the dataset.

(2) Lasso regression.

(3) sparse partial least squares regression

PCA was applied to discriminate the pure hESC and MEF cell populations and their 1:1 mixture; plot the PCA scores yield shows 3 clearly separated groups (Fig. 2A and 2B). PLS regression (as a linear technique) and non-linear technique, artificial neural networks (ANN), were used for estimation of the contamination rate of the cell cultures. The prediction precision of PLS was rather low, in comparison to ANN.

In a similar way, in the article **[15]** PCA and ANN of whole cell mass spectra were employed tool for the detection of induced and spontaneous alterations of embryonic stem cell populations in cell cultures.

The article **[16]** presents a complete, open and reproducible workflow in R for liquid biopsy analysis using MALDI mass spectrometry combined with machine learning. The article does not anticipate the latest R library *MsclassifR* (Godmer et al., 2025), which was not available at the time of writing the article **[16]**.

7 Methods comparison techniques

7.1 Comparison of analytical methods

The comparison of two measurement methods is a fundamental task in analytical chemistry, clinical diagnostics, biotechnology, and many other fields. Modern approaches combine graphical, regression, and correlation approaches.

In the article **[17]**, 22 soil samples and 3 CRMs (GBW07405, GBW07406, and GBW07407) were analysed using the LA-ICP-MS method. The results were then compared with those obtained by ICP-OES and ED-XRF analyses. Data visualisation is an effective tool for identifying bias, heteroscedasticity, and outliers. The following visual comparison tools were used: the XY plot, the Bland-Altman plot, and the score plot based on principal component analysis (PCA). A comparison of LA-ICP-MS with both ICP-OES and XRF methods was performed using regression techniques, including ordinary least squares regression (OLS), weighted least squares regression, orthogonal regression, Deming regression (DR), maximum likelihood estimation regression (bivariate least squares regression (BLS), York regression and Ripley-Thompson regression, and Passing–Bablok regression (PBR). With the exception of the Passing–Bablok regression, the confidence intervals for intercepts and slopes were estimated in all other regression methods using the bootstrap approach. To perform a simultaneous test of the slope and intercept of the regression line, confidence ellipses were used.

Traditional approaches, such as the paired t-test, correlation, or ordinary linear regression, lead to conclusions that incorrectly reflect the agreement between methods (Ludbrook, 2010b).

The *paired t-test* is applicable only if $\sigma_X = \sigma_Y$, so when the proportional error is absent (Linnet, 1999; Westgard & Hunt, 1973). To test the conformity of variances for normally distributed paired (dependent) data, the Pitman-Morgan test is often recommended. The main problems with using the t-test are outliers, data heterogeneity, trends, nonnormal distributions of the data, and data recording accuracy (Preece, 1982). Bradley and Blackwood (Bradley & Blackwood, 1989) provides a simultaneous test for both means and variances (C. J. Kowalski et al., 1995; Shoukri, 2024). For data with repeated measurements of each individual, a paired t-test calculation using mixed linear models can be used (Yu et al., 2022). Calculation of the paired t-test for left-censored data requires application of special methods based on maximum likelihood calculations (Helsel, 2012).

The *Pearson correlation coefficient* is not an appropriate method for assessing the agreement between two methods because it depends on both inter-individual variability (i.e., between actual values) and intra-individual variability (measurement error). If the variability between individuals is high compared to the measurement error, the correlation will be high. In contrast, if the variability between individuals is low, the correlation will be low (Altman & Bland, 1983; Ludbrook, 2002). Furthermore, a high correlation between two methods

does not mean that they agree; they could be systematically biased and still be highly correlated. Therefore, linear correlation is indicative only as a measure of the strength of the linear relationship between two variables (Giavarina, 2015; Ludbrook, 1997; Zaki et al., 2012). For repeated measurement correlations (Bland & Altman, 1995c, 1995a, 1996; Shan et al., 2020) special techniques, often based on linear models (Bakdash & Marusich, 2017, 2019; Nguyen & Jiang, 2011; A. Roy, 2006). Heteroscedasticity plays a role in significance testing of the correlation coefficient because it affects the width of the confidence interval. For its estimation, the adjusted percentile bootstrap method is recommended (Wilcox, Muska, 2001). When dealing with left-censored data, it is recommended to calculate the correlation coefficient using either of the following approaches: the first approach involves the imputation of nondetect values based on maximum likelihood (Barchard & Russell, 2023), the second approach involves the application of a special maximum likelihood-based method for correlation coefficient calculation (Helsel, 2012).

The *Lin concordance correlation coefficient* (CCC) combines accuracy and correlation (Lin, 1989). It is a special case of the *intraclass correlation coefficient* (ICC) (C. A. E. Nickerson, 1997). CCC determines how far the observed data deviates from the line of perfect agreement (i.e., a line with zero intercept and unit slope, with an angle of 45 ° in a scatter plot with equal range on both axes). Its value increases depending on the proximity of the reduced main axis of the data to the line of perfect concordance (accuracy) and on the tightness of the data around its reduced main axis (precision). The absolute value of CCC is always lower than the absolute value of Pearson's correlation coefficient. CCC for repeated measurements (Carrasco et al., 2009, 2013; Carrasco & Jover, 2003). For heteroscedastic measurements, a calculation based on linear models was derived (Nawarathna & Choudhary, 2013, 2015). For left-censored data, maximum likelihood imputation of nondetect values is recommended (Domthong et al., 2014).

Regression model I (ordinary least squares, OLS) assumes that the variable x is error-free and that error is present only in the variable y. However, when comparing two measurement methods, both variables are subject to error, so OLS provides biased estimates. *Regression dilution bias* (or *regression attenuation bias*) (Hutcheon et al., 2010; Riggs et al., 1978) is a bias of the linear regression model towards zero (i.e. an underestimation of its absolute value) caused by errors in the independent variable. The OLS regression is not symmetric (Fig. 18): the line of best fit for predicting Y from X is not the same as the line of best fit for predicting X from Y ("regressing y on x" underestimates the true directive and "regressing x on y" overestimates the true directive). Regression dilution bias therefore leads to erroneous regression coefficients due to application of OLS on measured data on both axes, for example, in physical chemistry (Němec, 2015), pharmacology (T. Roy, 1994), or ecotoxicology (Angleton & Bonham, 1995; Halfon, 1985). However, in some special cases (Stöckl et al., 1998), the regression dilution bias is negligible. For linear regression with repeated measurements (Asuero & González, 2007; Sayago et al., 2004; Sayago & Asuero, 2004) mixed linear models can be used. Weighted linear regression should be used for heteroscedasticity treatment (Linnet, 1993). For left-censored data, estimation of regression parameters based on maximum likelihood are suitable (Helsel, 2012).

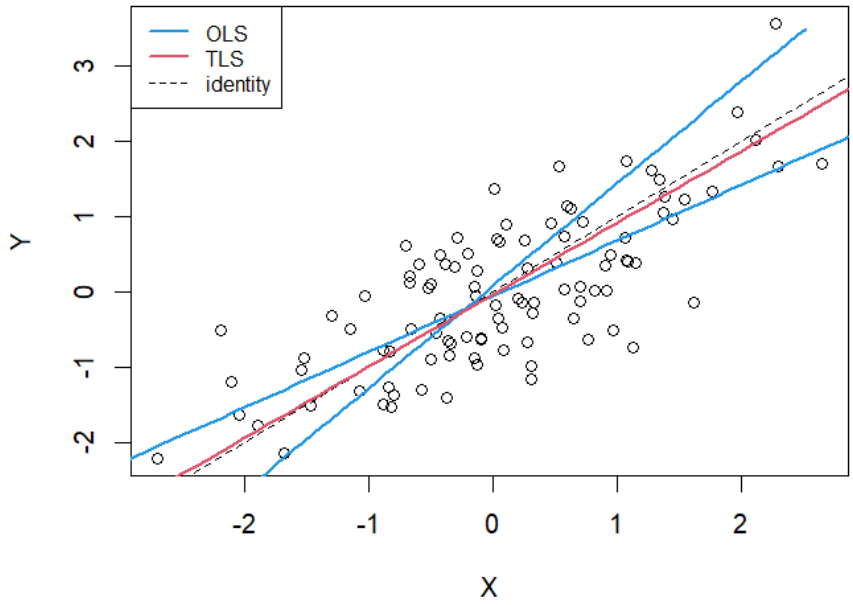

Fig. 18 Regression dilution.*OLS*: ordinary least squares (X vs. Y and Y vs. X), *TLS*: total least squares, *Identity*: line with intercept 0 and slope 1.

Regression model II (errors-in-variables regression) is a regression technique that minimises the perpendicular (orthogonal) distances of points from the regression line instead of just the vertical deviations (Andrews et al., 1996; De Castilho, 2004; Dhanoa et al., 2016; Ludbrook, 2010b; Tellinghuisen, 2020; C. Wu & Yu, 2018). It is especially suitable for situations where both variables are subject to error, not just the dependent variable (Ludbrook, 2010b, Tab. S1 in the Supplementary materials). Errors-in-variables regression methods are commonly used for the comparison of two measurement methods.

1. *Orthogonal regression* (total least squares) (Carroll & Ruppert, 1996; Kane & Mroch, 2020; Pallavi et al., 2022) minimises perpendicular (orthogonal) distances from points to the fitted line, treating both variables symmetrically. Errors in both axes should have approximately equal variances.

Other methods also include measurement errors in the model.

2. *Deming regression* (W. E. Deming, 1964; Linnet, 1993) is the standard method in clinical chemistry and analytical laboratories for method comparison studies (Linnet 1983; Ludbrook 2002). The method requires a known error ratio λ (it can be estimated from data based on repeated measurements, or it can be known from the introduction of the methodology). Incorrect λ leads to substantial bias. If λ is uncertain, sensitivity analysis or York regression is preferable (C. Wu & Yu, 2018). When λ is unknown and difficult to estimate, its iterative estimation is available in some modifications of the Deming regression method, such as *Ripley-Thompson regression* (Ripley and Thompson 1987) or *iteratively reweighted general Deming regression* (Martin 2000). Heteroscedasticity can be eliminated with *weighted Deming regression* (Linnet 1990) or *Bayesian Deming regression* (Pioda 2021). The latest approaches use precision profile models that model heteroscedastic error (Hawkins & Kraker, 2025).

3. The *York-Williamson regression* (Mahon, 1996; Williamson, 1968; York, 1966, 1968) is the most flexible error-in-variable approach. It is suitable for systems where the measurements x and y depend on each other (errors in x and y are correlated).

4. *Bivariate least squares (BLS) regression* (Riu & Rius, 1996, 1997) generalises Deming regression by adding correlation terms. It agrees with the York regression under appropriate variance specifications. BLS is mathematically equivalent to generalised least squares (GLS) and the total variance (TV) method for linear models, and it handles heteroscedastic errors (variance changing with measurement level) and correlated errors between x and y.

In the case of heterosketastic data, Linnet's weighting method (Linnet, 1990) can be used. For censored data, *Akritas-Theil-Sen (ATS) regression* is available (Akritas et al., 1995; Helsel, 2012).

Instead of regression methods, ***principal component analysis*** (PCA) is rarely used for a comparison of two methods (Hartmann and Massart, 1994). In the PC1 vs. PC2 plot, the first principal component represents the variability of the dataset, and the second principal component errors / differences / variability. The value of the second principal component score (PC2) corresponds to the orthogonal distance of the point from the line represented by the first principal component (PC1). PCA is thus analogous to orthogonal (or total least squares) regression (Andrews et al., 1996; J. D. Jackson & Dunlevy, 1988; Leng et al., 2007; Schuermans et al., 2005). The plot of component scores is thus related to the plot of orthogonal residuals (Hartmann & Massart, 1994). Heteroscedasticity also affects the results of principal component analysis (PCA), since PCA assumes homoskedasticity (equal variance) across dimensions. Heteroscedasticity may bias the PCA results towards components with higher variance, potentially overlooking important structures in the data. Weighted (Andrews et al., 1996) or iteratively reweighted (Park & Klabjan, 2018) PCA should be applied in such cases.

Passing-Bablok regression (Bilic-Zulle, 2011; Passing & Bablok, 1983), modification of nonparametric Theil-Sen regression, is a nonparametric, robust method that does not make any assumptions about the distribution of data or errors. It requires a larger sample size than Deming regression (Passing-Bablok has larger confidence intervals than Deming). There is no "weighted" Passing-Bablok (there is no minimisation of residuals that could be weighted).

The ***Bland-Atman (BA) plot*** (Bland & Altman, 1986, 1999; Giavarina, 2015; Ludbrook, 2010a; Mansournia et al., 2021) is a simple, intuitive graphical approach typically used to compare two methods. For $n = 3$ methods compared, the comparison of all methods with each other (Essack et al., 2009; Watkins et al., 2021) or each method with the reference method („gold standard“) (Proschan & Leifer, 2006) is generally used.

Mathematical coupling may occur in BA plots (Fig. 19), which can cause spurious correlation and thus false positive proportional systematic error (Mansournia et al., 2021). Therefore, displaying the difference between the results of both methods (on the y-axis) compared to the results of the reference method is only applicable if the results on the x-axis are not subject to any random error (e.g., when the actual values are known, or it is a so-called "gold standard"), or if the results of the tested method are subject to a much greater random error than the reference method (Bland & Altman, 1995b; Hyltoft Petersen et al., 1997; Krouwer, 2008). In such cases, the influence of spurious correlations is negligible (Hyltoft Petersen et al., 1997), and displaying the reference results on the x-axis is even more appropriate than the classic approach using arithmetic means (Krouwer, 2008; Taffé et al., 2022).

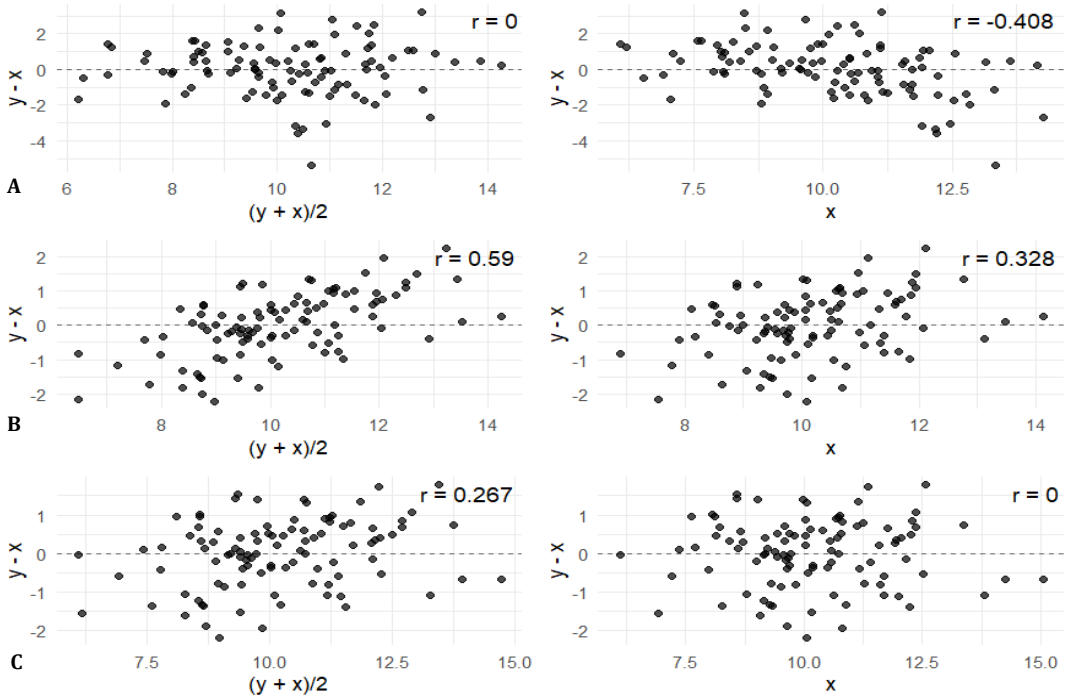

Fig. 19. Spurious correlations due to mathematical coupling in the Bland Altman method for x-values with identical stochastic noise (A), half- (B) and zero (C) in comparison with y-values.

The classical BA plot is only reliable for homoscedastic data, where the mean of two methods is plotted on the x-axis and their difference is plotted on the y-axis. In this case, the bias and 95% LoA can be estimated according to Bland and Altman (Bland & Altman, 1999), or with a number of refinements to consider, including bias-corrected LoA (Vock, 2016), prediction interval-based LoA for small and medium samples (Ludbrook, 2010a), exact LoA according to Francq (Francq et al., 2020), Bayesian estimation (Carkeet & Goh, 2018; Francq et al., 2020; Ludbrook, 2010a; Vock, 2016) and tolerance limits (Carkeet & Goh, 2018; Francq et al., 2020; Ludbrook, 2010a; Vock, 2016). There are several methods

that can be used to calculate the confidence intervals for acceptance limits (LoA). These include the delta method (Bland & Altman, 1999; Carkeet, 2015; Carkeet & Goh, 2018; Donner & Zou, 2012; Shieh, 2018), MOVER (Donner & Zou, 2012), or exact and percentile-based intervals (Shieh, 2018).

For proportional bias (due to different variances of the two methods), it describes detrending via a weighted mean transformation (Proschan & Leifer, 2006) and regression-based LoA using OLS (Ludbrook, 2010a) or correlated errors-in-variable models (Francq & Govaerts, 2016).

For heteroscedasticity treatment, two strategies can be used:

- Transformation of data, for example, logarithmic BA plots (Euser et al., 2008), ratio plots (Eksborg, 1981), mean-normalised plots (Pollock et al., 1992), normalisation using variance function (Sadler, 2019b), variance stabilising transformations like Box–Cox, Yeo–Johnson, or generalised log (Hawkins, 2013), and then backtransforming parallel LoA to obtain V-shaped LoA.
- Direct modelling of non-parallel LoA on the original scale using prediction bands derived from weighted linear regression (Francq & Govaerts, 2014; Ludbrook, 2010a) or Altman's method with linear, quadratic, or fractional polynomial functions (Gerke & Möller, 2023; Ludbrook, 2010a; Sevrukov et al., 2005).

The BA technique can also be applied to data with repeated measurements or data with measurement error (Bland & Altman, 1999, 2007; Taffé, 2020; Taffé et al., 2020).

For left-censored data, is the BA method unsuitable (Sadler, 2019a).

A nonparametric BA approach is proposed for nonnormal differences, using quantile-based LoA (e.g., 0.025 and 0.975 quantiles) (Frey et al., 2020; Twomey, 2006).

For multiple methods, the comparison of all methods with each other or the comparison of each method with the reference method („gold standard ") is usually used.

Principal component analysis can also be used for multiple methods comparison (Carey et al., 1975; Martínez et al., 2001).

Modifications to the Bland-Altman plot were developed for the comparison of $n \geq 2$ methods (M. Jones et al., 2011; Koziol et al., 2008; Möller et al., 2021). Christensen et al. (Christensen et al., 2020) derived LoA confidence intervals for Jones et al. (2011) method.

7.2 Analysis of means

Analysis of means (ANOM) is a simple graphical summary of results for comparative experiments. Its graphical output is more illustrative than the output from the commonly used one-way ANOVA coupled with multiple comparison testing. This methodology was applied not only for comparison of several means (including heteroscedastic and rank tansformed data), but also for variances, proportions, and correlation coefficients (Jayalath & Turner, 2021; Mohammed & Holder, 2012; Nelson et al., 2005; Pallmann & Hothorn, 2016; C. V. Rao, 2005). Similarly to ANOVA (Larson, 1992), the ANOM method allows one to work with summary results (mean, SD, and number of cases).

The application of ANOM for analytical data treatment (Demir et al., 2022; Keskin et al., 2021; Yigit & Mendes, 2017) is still rather sporadic compared to ANOVA.

The article **[18]** compares several mineralisation methods for the determination of selenium. Selenium in various biological samples was repeatedly determined, and the results were analysed with one-way analysis of variance (ANOVA) and analysis of means (ANOM). Both methods gave analogical results; however, the graphical output of ANOM is more illustrative compared to ANOVA.

Bootstrap or permutation randomisation are sporadically applied, e.g., for variances comparison ANOM (Bernard & Wludyka, 2001; Krishna Reddy et al., 2011; Wludyka & Nelson, 1997). The AMOM method can also be generalised for random- and mixed-effects models (Jayalath & Ng, 2020; Jayalath & Turner, 2021; Mahmood et al., 2023; Pallmann & Hothorn, 2016), or for comparison of averages with a reference value (Chakraborti, 1991). The ANOM plot transformed to polar coordinates, so called “compass plot”, was used in toxicology (Budsaba et al., 2000).

8 Conclusions and outlook

The small wisdom is like water in a glass: clear, transparent, pure.
The great wisdom is like the water in the sea: dark, mysterious, impenetrable.

Rabindranath Tagore

The research presented in this thesis demonstrates the indispensable role of chemometrics and advanced data analysis in modern analytical chemistry and addresses a critical gap in modern analytical chemistry: the need for sophisticated data-analysis approaches that account for the inherent complexities of archaeometric and geochemical datasets. Unlike traditional statistical methods that assume ideal data conditions, this work acknowledges and provides practical solutions for the challenges encountered in real-world applications, including missing values, censored data, compositional constraints, and measurement uncertainties.

The key conclusions drawn from the submitted work are as follows:

Robustness in provenance studies: The application of multivariate statistical methods has proven superior to univariate approaches in provenance determination. As demonstrated in studies on obsidian and conflict minerals (coltan), machine learning algorithms can successfully handle complex geochemical fingerprints and mitigate the "dilution effect" often encountered in archaeological ceramics.

Importance of data pre-processing: The work highlights that the quality of statistical output is strictly dependent on data pretreatment. The rigorous application of compositional data analysis (CoDA) using log-ratio transformations is essential for geochemical data to avoid spurious correlations. Furthermore, the correct handling of censored data (values <LOD) and missing values through multiple imputation techniques significantly improves the reliability of environmental and archaeological models.

Spatially dependent data analysis: In the field of soil chemical data analysis, geostatistical methods have enabled the reconstruction of past human activities. Spatial data methodology can be also used in conjunction with elemental imaging, as well as with multivariate and machine learning methods.

Pattern recognition of mass spectra: In the field of mass spectrometry, the use of computational methods has led to significant advancements. These methods have enabled the elucidation of the composition of complex inorganic clusters and the classification of biomedical samples. This is a task that would not be possible through manual interpretation of mass spectra.

Critical evaluation of methods: The comparative studies included in this thesis emphasise that statistical validation (e.g., Deming regression, Bland-Altman plots) is required when cross-validating analytical techniques. Simple correlation coefficients are often insufficient and misleading when comparing method performance.

The work also confirms that in the era of high-throughput instrumentation, the bottleneck of scientific discovery has shifted from data acquisition to data interpretation:

The *primacy of data quality and the "GIGO" principle:* Data quality plays a key role in the validity of any chemometric model. The thesis underscores the concept of GIGO (Garbage In, Garbage Out); just as in computer science, incorrect or meaningless input data produce worthless output. This principle applies generally to all analysis, as arguments are unreliable if their premises are incorrect (Kilkenny & Robinson, 2018). Sampling bias is particularly prevalent in archaeometry and to a lesser extent in geochemical and environmental research. This largely precludes the use of certain statistical methods that assume random selection. Consequently, sampling bias must be explicitly accounted for when interpreting results to avoid deriving definitive truths from non-representative samples.

Avoidance of "Cargo cult" data analysis: With the democratisation of statistical software, there is an increasing risk of "black box" analysis, where complex algorithms are applied without verifying assumptions. This thesis challenges the uncritical application of statistics, often referred to as "Cargo cult" data analysis (Gelman & Higgs, 2025; Stark & Saltelli, 2018). This term describes the superficial imitation of analytical procedures - running regressions and citing p-values - without grasping the underlying principles or assumptions. This "form over function" approach results in practices that appear rigorous but lack epistemic substance. The thesis advocates against blind replication of workflows and overconfidence in model output, urging researchers to acknowledge uncertainty and bias.

To reduce the risk of choosing an inappropriate method or misinterpreting results, it is essential to perform exploratory data analysis (EDA) and interactive, user-driven data exploration. These practices help to reveal structure, outliers and assumptions early in the workflow and guide analysts towards methods that truly fit the data (Bondu et al., 2020; Hroch, 2025; Idkowiak et al., 2025; Meloun & Militký, 2012).

Cognitive biases can subtly influence data interpretation and model creation in analytical chemistry and chemometrics. Availability bias leads analysts to overestimate easily retrievable patterns or recent experiences, while representativeness bias encourages hasty assignment of samples to known classes simply because they "look similar." Together with other biases, such as confirmation bias or anchoring, these tendencies can cause incorrect evaluation of spectra, incorrect choice of preprocessing, or overestimation of model quality. Awareness of these mechanisms and working with objective validation, blind testing, and transparent documentation is key to reliable chemometric analysis.

For the end user, who is often not an expert, it is crucial to present analytical results in a simple and understandable way. This means using clear graphs, highlighting the main results and conclusions (Gemignani, 2014). Visualisations should be clean, intuitive, and accompanied by a short story explaining what was measured, why it is important, and what the user should do next.

Although significant progress has been made, the field of chemometrics and data analysis is undergoing rapid transformation. Future research will need to address several emerging trends and challenges:

Integration of Large Language Models (LLMs): Recent developments have seen the integration of LLMs into data analytics, transforming workflows by augmenting traditional methods with natural language capabilities (Miriello & De Luca, 2025; J. H. Moore & Tatonetti, 2025). LLMs act as assistants that can lower the barrier for nontechnical users, explain statistical concepts, and automate code generation in R or Python. This leads to the democratisation of data evaluation, making powerful tools accessible to users with limited programming expertise. However, this comes with significant risks. Researchers must be wary of hallucinations (factually incorrect generation of statistics), the black box problem (lack of transparency in reasoning) and deskilling (reduction in critical thinking due to overreliance). Future work must focus on validation frameworks to ensure that LLMs are used to assist, not replace, human oversight.

Advanced methodological implementation: Beyond machine learning and LLMs, future development will include the implementation of Bayesian methods (Armstrong & Hibbert, 2009, 2009; Brereton, 2022), which offer a probabilistic approach to uncertainty, often better suited for complex environmental data than frequentist statistics. Also, modelling dependencies and multidimensional distributions using copulas (Aminullah et al., 2023; Genest & Favre, 2007; Üçer, 2011) represents a promising technique for handling the non-Gaussian, correlated datasets typical in geochemistry. Meta-analysis serves as a tool that can consolidate often fragmented data from many studies and reveal patterns that would remain hidden in individual works (Field & Gillett, 2010; Hansen et al., 2022; Mikolajewicz & Komarova, 2019).

In summary, this work establishes that advanced data analysis is not merely a postprocessing step, but a fundamental component of the analytical methodology. Whether through established chemometric techniques or emerging AI-augmented workflows, rigorous data treatment is essential for extracting legally and scientifically defensible conclusions from complex data.

9 Bibliography

Adams, F., & Adriaens, M. (2020). The metamorphosis of analytical chemistry. *Analytical and Bioanalytical Chemistry*, *412*(15), 3525–3537. https://doi.org/10.1007/s00216-019-02313-z

Agha-Aligol, D., Lamehi-Rachti, M., Oliaiy, P., Shokouhi, F., Farahani, M. F., Moradi, M., & Jalali, F. F. (2015). Characterization of Iranian obsidian artifacts by PIXE and multivariate statistical analysis. *Geoarchaeology*, *30*(3), 261–270. https://doi.org/10.1002/gea.21509

Ahmadi, H., Granger, D. A., Hamilton, K. R., Blair, C., & Riis, J. L. (2021). Censored data considerations and analytical approaches for salivary bioscience data. *Psychoneuroendocrinology*, *129*, 105274. https://doi.org/10.1016/j.psyneuen.2021.105274

Ahsan, M., Mahmud, M., Saha, P., Gupta, K., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, *9*(3), 52. https://doi.org/10.3390/technologies9030052

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Springer Netherlands.

Akmeemana, A., Corzo, R., & Almirall, J. (2022). Calculation of likelihood ratios in forensic glass comparisons; introduction to a R code and Shiny app applied to existing background glass elemental databases. *Forensic Chemistry*, *27*, 100390. https://doi.org/10.1016/j.forc.2021.100390

Akmeemana, A., Weis, P., Corzo, R., Ramos, D., Zoon, P., Trejos, T., Ernst, T., Pollock, C., Bakowska, E., Neumann, C., & Almirall, J. (2021). Interpretation of chemical data from glass analysis for forensic purposes. *Journal of Chemometrics*, *35*(1), e3267. https://doi.org/10.1002/cem.3267

Akritas, M. G., Murphy, S. A., & Lavalley, M. P. (1995). The Theil-Sen estimator with doubly censored data and applications to astronomy. *Journal of the American Statistical Association*, *90*(429), 170–177. https://doi.org/10.1080/01621459.1995.10476499

Alenazi, A. (2023). A review of compositional data analysis and recent advances. *Communications in Statistics - Theory and Methods*, *52*(16), 5535–5567. https://doi.org/10.1080/03610926.2021.2014890

Alfassi, Z. B. (2005). Vector analysis of multi-measurements identification. *Journal of Radioanalytical and Nuclear Chemistry*, *266*(2), 245–250. https://doi.org/10.1007/s10967-005-0899-y

AlKhars, M., Evangelopoulos, N., Pavur, R., & Kulkarni, S. (2019). Cognitive biases resulting from the representativeness heuristic in operations management: An experimental investigation. *Psychology Research and Behavior Management*, *Volume 12*, 263–276. https://doi.org/10.2147/PRBM.S193092

Allison, S. T., & Messick, D. M. (1988). The feature-positive efect, attitude strength, and degree of perceived consensus. *Personality and Social Psychology Bulletin*, *14*(2), 231–241. https://doi.org/10.1177/0146167288142002

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Statistician*, *32*(3), 307. https://doi.org/10.2307/2987937

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, *311*(7003), 485–485. https://doi.org/10.1136/bmj.311.7003.485

Alwateer, M., Atlam, E.-S., El-Raouf, M. M. A., Ghoneim, O. A., & Gad, I. (2024). Missing data imputation: A comprehensive review. *Journal of Computer and Communications*, *12*(11), 53–75. https://doi.org/10.4236/jcc.2024.1211004

Aminullah, S. Z., Novita, M., & Fithriani, I. (2023). Vine copula model: Application to chemical elements in water samples. *Proceedings of The International Conference on Data Science and Official Statistics*, *2023*(1), 572–585. https://doi.org/10.34123/icdsos.v2023i1.346

Amirajlo, P., Hassani, H., Beiranvand Pour, A., & Habibkhah, N. (2025). Detection of multivariate geochemical anomalies using machine learning (ML) algorithms in Dehaq Pb-Zn mineralization, Sanandaj-Sirjan zone, Isfahan, Iran. *Earth Science Informatics*, *18*(1), 124. https://doi.org/10.1007/s12145-024-01565-3

Anders, O. U. (1972). Ratio matching. Statistical aid for discovering generic relations among samples. *Analytical Chemistry*, *44*(12), 1930–1933. https://doi.org/10.1021/ac60320a015

Andrews, D. T., Chen, L., Wentzell, P. D., & Hamilton, D. C. (1996). Comments on the relationship between principal components analysis and weighted linear regression for bivariate data sets. *Chemometrics and Intelligent Laboratory Systems*, *34*(2), 231–244. https://doi.org/10.1016/0169-7439(96)00031-7

Angleton, G. M., & Bonham, C. D. (1995). Least squares regression vs. Geometric mean regression for ecotoxicology studies. *Applied Mathematics and Computation*, *72*(1), 21–32. https://doi.org/10.1016/0096-3003(94)00161-V

Anglisano, A., Casas, L., Anglisano, M., & Queralt, I. (2019). Application of supervised machine-learning methods for attesting provenance in Catalan traditional pottery industry. *Minerals*, *10*(1), 8. https://doi.org/10.3390/min10010008

Angourakis, A., Martínez Ferreras, V., Torrano, A., & Gurt Esparraguera, J. M. (2018). Presenting multivariate statistical protocols in R using Roman wine amphorae productions in Catalonia, Spain. *Journal of Archaeological Science*, *93*, 150–165. https://doi.org/10.1016/j.jas.2018.03.007

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Record*, *28*(2), 49–60. https://doi.org/10.1145/304181.304187

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Aoyagi, S. (2023). Application of machine learning to spectrum and image data. *Journal of Vacuum Science & Technology A*, *41*(6), 063202. https://doi.org/10.1116/6.0002858

Archie, J. P. (1981). Mathematic coupling of data: A common source of error. *Annals of Surgery*, *193*(3), 296–303. https://doi.org/10.1097/00000658-198103000-00008

Armstrong, N., & Hibbert, D. B. (2009). An introduction to Bayesian methods for analyzing chemistry data. *Chemometrics and Intelligent Laboratory Systems*, *97*(2), 194–210. https://doi.org/10.1016/j.chemolab.2009.04.001

Arnold, D. E., Neff, H., & Bishop, R. L. (1991). Compositional analysis and "sources" of pottery: An ethnoarcheological approach. *American Anthropologist*, *93*(1), 70–90. https://doi.org/10.1525/aa.1991.93.1.02a00040

Aruga, R. (1997). Treatment of responses below the detection limit: Some current techniques compared by Factor Analysis on environmental data. *Analytica Chimica Acta*, *354*(1–3), 255–262. https://doi.org/10.1016/S0003-2670(97)00463-7

Aruga, R. (1998). Closure of analytical chemical data and multivariate classification. *Talanta*, *47*(4), 1053–1061. https://doi.org/10.1016/S0039-9140(98)00126-X

Aruga, R. (2003). The problem of multivariate classification of samples with radial (or V-shaped) chemical data. *Talanta*, *60*(5), 937–944. https://doi.org/10.1016/S0039-9140(03)00157-7

Aruga, R. (2004). The problem of responses less than the reporting limit in unsupervised pattern recognition. *Talanta*, *62*(5), 871–878. https://doi.org/10.1016/j.talanta.2003.10.036

Aruga, R., Mirti, P., & Casoli, A. (1993). Application of multivariate chemometric techniques to the study of Roman pottery (terra sigillata). *Analytica Chimica Acta*, *276*(1), 197–204. https://doi.org/10.1016/0003-2670(93)85056-P

Ashari, I. F., Dwi Nugroho, E., Baraku, R., Novri Yanda, I., & Liwardana, R. (2023). Analysis of elbow, silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-index evaluation on k-means algorithm for classifying flood-affected areas in Jakarta. *Journal of Applied Informatics and Computing*, *7*(1), 89–97. https://doi.org/10.30871/jaic.v7i1.4947

Asuero, A. G., & González, G. (2007). Fitting straight lines with replicated observations by linear regression. III. Weighting data. *Critical Reviews in Analytical Chemistry*, *37*(3), 143–172. https://doi.org/10.1080/10408340701244615

Atchley, W. R., Gaskins, C. T., & Anderson, D. (1976). Statistical Properties of Ratios. I. Empirical Results. *Systematic Zoology*, *25*(2), 137. https://doi.org/10.2307/2412740

Austin, P. C., White, I. R., Lee, D. S., & Van Buuren, S. (2021). Missing data in clinical research: A tutorial on multiple imputation. *Canadian Journal of Cardiology*, *37*(9), 1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

Babamoradi, H., Van Den Berg, F., & Rinnan, Å. (2013). Bootstrap based confidence limits in principal component analysis—A case study. *Chemometrics and Intelligent Laboratory Systems*, *120*, 97–105. https://doi.org/10.1016/j.chemolab.2012.10.007

Baccarelli, A., Pfeiffer, R., Consonni, D., Pesatori, A. C., Bonzini, M., Patterson, D. G., Bertazzi, P. A., & Landi, M. T. (2005). Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. *Chemosphere*, *60*(7), 898–906. https://doi.org/10.1016/j.chemosphere.2005.01.055

Badawy, W. M., Dmitriev, A. Yu., Koval, V. Yu., Smirnova, V. S., Chepurchenko, O. E., Lobachev, V. V., Belova, M. O., & Galushko, A. M. (2022). Formation of reference groups for archaeological pottery using neutron activation and multivariate statistical analyses. *Archaeometry*, *64*(6), 1377–1393. https://doi.org/10.1111/arcm.12793

Bain, C., Shi, D., Banad, Y., Ethridge, L., Norris, J., & Loeffelman, J. (2025). A tutorial on supervised machine learning variable selection methods in classification for the social and health sciences in R. *Journal of Behavioral Data Science*, *5*(1), 103–147. https://doi.org/10.35566/jbds/bain

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456. https://doi.org/10.3389/fpsyg.2017.00456

Bakdash, J. Z., & Marusich, L. R. (2019). Corrigendum: Repeated measures correlation. *Frontiers in Psychology*, *10*, 1201. https://doi.org/10.3389/fpsyg.2019.01201

Balamurali, M., Silversides, K. L., & Melkumyan, A. (2019). A comparison of t-SNE, SOM and SPADE for identifying material type domains in geological data. *Computers & Geosciences*, *125*, 78–89. https://doi.org/10.1016/j.cageo.2019.01.011

Banas, K., Banas, A. M., Gajda, M., Kwiatek, W. M., Pawlicki, B., & Breese, M. B. H. (2013). Analysis of synchrotron radiation induced X-ray emission spectra with R environment. *Radiation Physics and Chemistry*, *93*, 82–86. https://doi.org/10.1016/j.radphyschem.2013.04.026

Banning, E. B. (2021). Sampled to death? The rise and fall of probability sampling in archaeology. *American Antiquity*, *86*(1), 43–60. https://doi.org/10.1017/aaq.2020.39

Barchard, K. A., & Russell, J. A. (2023). Distorted correlations among censored data: Causes, effects, and correction. *Behavior Research Methods*, *56*(3), 1207–1228. https://doi.org/10.3758/s13428-023-02086-5

Barge, O., & Chataigner, C. (2003). The procurement of obsidian: Factors influencing the choice of deposits. *Journal of Non-Crystalline Solids*, *323*(1–3), 172–179. https://doi.org/10.1016/S0022-3093(03)00301-6

Barnes, S.-J. (1990). The use of metal ratios in prospecting for platinum-group element deposits in mafic and ultramafic intrusions. *Journal of Geochemical Exploration*, *37*(1), 91–99. https://doi.org/10.1016/0375-6742(90)90084-N

Barnett, A. G. (2004). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Barone, G., Lo Giudice, A., Mazzoleni, P., Pezzino, A., Barilaro, D., Crupi, V., & Triscari, M. (2005). Chemical characterization and statistical multivariate analysis of ancient pottery from Messina, Catania, Lentini and Siracusa (Sicily). *Archaeometry*, *47*(4), 745–762. https://doi.org/10.1111/j.1475-4754.2005.00230.x

Barrett, T. J., Sharpe, R. L., & Azim, M. E. (2014). A novel approach for graphing censored environmental data. *Integrated Environmental Assessment and Management*, *10*(4), 595–596. https://doi.org/10.1002/ieam.1563

Baxter, M. J. (1991). Principal component and correspondence analyses of glass compositions: An empirical study. *Archaeometry*, *33*(1), 29–41. https://doi.org/10.1111/j.1475-4754.1991.tb00683.x

Baxter, M. J. (2006). A review of supervised and unsupervised pattern recognition in archaeometry. *Archaeometry*, *48*(4), 671–694. https://doi.org/10.1111/j.1475-4754.2006.00280.x

Baxter, M. J. (2008). Mathematics, statistics and archaeometry: The past 50 years or so. *Archaeometry*, *50*(6), 968–982. https://doi.org/10.1111/j.1475-4754.2008.00427.x

Baxter, M. J., Beardah, C. C., Cool, H. E. M., & Jackson, C. M. (2005). Compositional data analysis of some alkaline glasses. *Mathematical Geology*, *37*(2), 183–196. https://doi.org/10.1007/s11004-005-1308-3

Baxter, M. J., Beardah, C. C., Papageorgiou, I., Cau, M. A., Day, P. M., & Kilikoglou, V. (2008). On statistical approaches to the study of ceramic artefacts using geochemical and petrographic data. *Archaeometry*, *50*(1), 142–157. https://doi.org/10.1111/j.1475-4754.2007.00359.x

Baxter, M. J., Cool, H. E. M., & Heyworth, M. P. (1990). Principal component and correspondence analysis of compositional data: Some similarities. *Journal of Applied Statistics*, *17*(2), 229–235. https://doi.org/10.1080/757582834

Baxter, M. J., Cool, H. E. M., & Jackson, C. M. (2005). Further studies in the compositional variability of colourless Romano-British vessel glass. *Archaeometry*, *47*(1), 47–68. https://doi.org/10.1111/j.1475-4754.2005.00187.x

Baxter, M. J., & Freestone, I. C. (2006). Log-ratio compositional data analysis in archaeometry. *Archaeometry*, *48*(3), 511–531. https://doi.org/10.1111/j.1475-4754.2006.00270.x

Beck, A. G., Muhoberac, M., Randolph, C. E., Beveridge, C. H., Wijewardhane, P. R., Kenttämaa, H. I., & Chopra, G. (2024). Recent developments in machine learning for mass spectrometry. *ACS Measurement Science Au*, *4*(3), 233–246. https://doi.org/10.1021/acsmeasuresciau.3c00060

Beck, C., & Jones, G. T. (1989). Bias and archaeological classification. *American Antiquity*, *54*(2), 244–262. https://doi.org/10.2307/281706

Beier, Th., & Mommsen, H. (1994). A method for classifying multidimensional data with respect to uncertainties of measurement and its application to archaeometry. *Naturwissenschaften*, *81*(12), 546–548. https://doi.org/10.1007/BF01140001

Ben Dor, Y., Finkel, M., & Ben-Yosef, E. (2023). A probabilistic approach to provenance studies using whole object elemental composition: Chert (flint) as a case study. *Journal of Archaeological Science*, *153*, 105767. https://doi.org/10.1016/j.jas.2023.105767

Bern, C. R. (2009). Soil chemistry in lithologically diverse datasets: The quartz dilution effect. *Applied Geochemistry*, *24*(8), 1429–1437. https://doi.org/10.1016/j.apgeochem.2009.04.013

Bernard, A. J., & Wludyka, P. S. (2001). Robust *I* -sample analysis of means type randomization tests for variances. *Journal of Statistical Computation and Simulation*, *69*(1), 57–88. https://doi.org/10.1080/00949650108812082

Bertrand, A., Belloum, R., Eagan, J. R., & Maxwell, W. (2022). How cognitive biases affect XAI-assisted decision-making: A systematic review. *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 78–91. https://doi.org/10.1145/3514094.3534164

Beyene, J., & Moineddin, R. (2005). Methods for confidence interval estimation of a ratio parameter with application to location quotients. *BMC Medical Research Methodology*, *5*(1), 32. https://doi.org/10.1186/1471-2288-5-32

Bezdek, J. C., & R. Pal, N. (1995). An index of topological preservation for feature extraction. *Pattern Recognition*, *28*(3), 381–391. https://doi.org/10.1016/0031-3203(94)00111-X

Bieber, A. M., Brooks, D. W., Harbottle, G., & Sayre, E. V. (1976). Application of multivariate techniques to analytical data on Aegean ceramics. *Archaeometry*, *18*(1), 59–74. https://doi.org/10.1111/j.1475-4754.1976.tb00145.x

Bilic-Zulle, L. (2011). Comparison of methods: Passing and Bablok regression. *Biochemia Medica*, 49–52. https://doi.org/10.11613/BM.2011.010

Bintliff, J., & Degryse, P. (2022). A review of soil geochemistry in archaeology. *Journal of Archaeological Science: Reports*, *43*, 103419. https://doi.org/10.1016/j.jasrep.2022.103419

Bishop, R. L., Canouts, V., Crown, P. L., & Atley, S. P. D. (1990). Sensitivity, precision, and accuracy: Their roles in ceramic compositional data bases. *American Antiquity*, *55*(3), 537–546. https://doi.org/10.2307/281283

Bishop, R. L., Rands, R. L., & Holley, G. R. (1982). Ceramic compositional analysis in archaeological perspective. In *Advances in Archaeological Method and Theory* (pp. 275–330). Elsevier. https://doi.org/10.1016/B978-0-12-003105-4.50012-1

Bitetto, A., Mangone, A., Mininni, R. M., & Giannossa, L. C. (2016). A nonlinear principal component analysis to study archeometric data. *Journal of Chemometrics*, *30*(7), 405–415. https://doi.org/10.1002/cem.2807

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland, J. M., & Altman, D. G. (1995a). Calculating correlation coefficients with repeated observations: Part 2—correlation between subjects. *BMJ*, *310*(6980), 633. https://doi.org/10.1136/bmj.310.6980.633

Bland, J. M., & Altman, D. G. (1995b). Comparing methods of measurement: Why plotting difference against standard method is misleading. *Lancet*, *346*(8982), 1085–1087. https://doi.org/10.1016/S0140-6736(95)91748-9

Bland, J. M., & Altman, D. G. (1995c). Statistics notes: Calculating correlation coefficients with repeated observations: Part 1—correlation within subjects. *BMJ*, *310*(6977), 446. https://doi.org/10.1136/bmj.310.6977.446

Bland, J. M., & Altman, D. G. (1996). Statistics Notes: Measurement error and correlation coefficients. *BMJ*, *313*(7048), 41–42. https://doi.org/10.1136/bmj.313.7048.41

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Bland, J. M., & Altman, D. G. (2007). Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics*, *17*(4), 571–582. https://doi.org/10.1080/10543400701329422

Bodinier, B., Rodrigues, S., Karimi, M., Filippi, S., Chiquet, J., & Chadeau-Hyam, M. (2025). Stability selection and consensus clustering in R : The R package sharp. *Journal of Statistical Software*, *112*(5). https://doi.org/10.18637/jss.v112.i05

Bondu, R., Cloutier, V., Rosa, E., & Roy, M. (2020). An exploratory data analysis approach for assessing the sources and distribution of naturally occurring contaminants (F, Ba, Mn, As) in groundwater from southern Quebec (Canada). *Applied Geochemistry*, *114*, 104500. https://doi.org/10.1016/j.apgeochem.2019.104500

Bonizzoni, L., Kulchytska, O., & Ruschioni, G. (2023). XRF semi-quantitative analysis and multivariate statistics for the classification of obsidian flows in the Mediterranean area. *Applied Sciences*, *13*(6), 3495. https://doi.org/10.3390/app13063495

Bonsall, C., & Kohút, M. (2025). The Carpathian obsidians reconsidered: An application of portable X-ray fluorescence spectrometry (pXRF). *Journal of Archaeological Science: Reports*, *65*, 105198. https://doi.org/10.1016/j.jasrep.2025.105198

Born, R. T. (2024). Stop fooling yourself! (Diagnosing and treating confirmation bias). *Eneuro*, *11*(10), ENEURO.0415-24.2024. https://doi.org/10.1523/ENEURO.0415-24.2024

Boss, J., Mukherjee, B., Ferguson, K. K., Aker, A., Alshawabkeh, A. N., Cordero, J. F., Meeker, J. D., & Kim, S. (2019). Estimating outcome-exposure associations when exposure biomarker detection limits vary across batches. *Epidemiology*, *30*(5), 746–755. https://doi.org/10.1097/EDE.0000000000001052

Bradley, E. L., & Blackwood, L. G. (1989). Comparing paired data: A simultaneous test for means and variances. *American Statistician*, *43*(4), 234–235. https://doi.org/10.1080/00031305.1989.10475665

Brandl, M., Hauzenberger, C., Postl, W., Martinez, M. M., Filzmoser, P., & Trnka, G. (2014). Radiolarite studies at Krems-Wachtberg (Lower Austria): Northern Alpine versus Carpathian lithic resources. *Quaternary International*, *351*, 146–162. https://doi.org/10.1016/j.quaint.2013.01.031

Braswell, G. E., Clark, J. E., Aoyama, K., McKillop, H. I., & Glascock, M. D. (2000). Determining the geological provenance of obsidian artifacts from the Maya region: A test of the efficacy of visual sourcing. *Latin American Antiquity*, *11*(3), 269–282. https://doi.org/10.2307/972178

Brereton, R. G. (2011). One-class classifiers. *Journal of Chemometrics*, *25*(5), 225–246. https://doi.org/10.1002/cem.1397

Brereton, R. G. (2022). Introduction to Bayesian methods. *Journal of Chemometrics*, *36*(1), e3333. https://doi.org/10.1002/cem.3333

Brett, M. T. (2004). When is a correlation between non-independent variables “spurious”? *Oikos*, *105*(3), 647–656. https://doi.org/10.1111/j.0030-1299.2004.12777.x

Brewer, M. (2025). When should absence of evidence be evidence of absence? A case study from paleogeology. *Philosophy of Science*, 1–10. https://doi.org/10.1017/psa.2025.10125

Brückner, A., & Heethoff, M. (2017). A chemo-ecologists’ practical guide to compositional data analysis. *Chemoecology*, *27*(1), 33–46. https://doi.org/10.1007/s00049-016-0227-8

Budsaba, K., Smith, C. E., & Riviere, J. E. (2000). Compass plots: A combination of star plot and analysis of means to visualize significant interactions in complex toxicology studies. *Toxicology Methods*, *10*(4), 313–332. https://doi.org/10.1080/105172300750048764

Burton, J. H., Price, T. D., & Middleton, W. D. (1999). Correlation of bone Ba/Ca and Sr/Ca due to biological purification of calcium. *Journal of Archaeological Science*, *26*(6), 609–616. https://doi.org/10.1006/jasc.1998.0378

Butler, J. C. (1979). Numerical consequences of changing the units in which chemical analyses of igneous rocks are analyzed. *Lithos*, *12*(1), 33–39. https://doi.org/10.1016/0024-4937(79)90060-4

Buxeda I Garrigós, J. (1999). Alteration and contamination of archaeological ceramics: The perturbation problem. *Journal of Archaeological Science*, *26*(3), 295–313. https://doi.org/10.1006/jasc.1998.0390

Carey, R. Neill., Wold, Svante., & Westgard, J. O. (1975). Principal component analysis. Alternative to referee methods in method comparison studies. *Analytical Chemistry*, *47*(11), 1824–1829. https://doi.org/10.1021/ac60361a037

Carkeet, A. (2015). Exact parametric confidence intervals for Bland-Altman limits of agreement. *Optometry and Vision Science*, *92*(3), e71–e80. https://doi.org/10.1097/OPX.0000000000000513

Carkeet, A., & Goh, Y. T. (2018). Confidence and coverage for Bland–Altman limits of agreement and their approximate confidence intervals. *Statistical Methods in Medical Research*, *27*(5), 1559–1574. https://doi.org/10.1177/0962280216665419

Carlosena, A., Andrade, J. M., Kubista, M., & Prada, D. (1995). Procrustes rotation as a way to compare different sampling seasons in soils. *Analytical Chemistry*, *67*(14), 2373–2378. https://doi.org/10.1021/ac00110a008

Carrasco, J. L., & Jover, L. (2003). Estimating the generalized concordance correlation coefficient through variance components. *Biometrics*, *59*(4), 849–858. https://doi.org/10.1111/j.0006-341X.2003.00099.x

Carrasco, J. L., King, T. S., & Chinchilli, V. M. (2009). The Concordance Correlation Coefficient for Repeated Measures Estimated by Variance Components. *Journal of Biopharmaceutical Statistics*, *19*(1), 90–105. https://doi.org/10.1080/10543400802527890

Carrasco, J. L., Phillips, B. R., Puig-Martinez, J., King, T. S., & Chinchilli, V. M. (2013). Estimation of the concordance correlation coefficient for repeated measures using SAS and R. *Computer Methods and Programs in Biomedicine*, *109*(3), 293–304. https://doi.org/10.1016/j.cmpb.2012.09.002

Carroll, R. J., & Ruppert, D. (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *American Statistician*, *50*(1), 1–6. https://doi.org/10.1080/00031305.1996.10473533

Castellanos-García, L. J., Gokhan Elci, S., & Vachet, R. W. (2020). Reconstruction, analysis, and segmentation of LA-ICP-MS imaging data using Python for the identification of sub-organ regions in tissues. *The Analyst*, *145*(10), 3705–3712. https://doi.org/10.1039/C9AN02472G

Cerrillo-Cuenca, E., Sepúlveda, M., Cabello, G., & Bastías, F. (2024). Color-based discrimination of color hues in rock paintings through Gaussian mixture models: A case study from Chomache site (Chile). *Heritage Science*, *12*(1), 169. https://doi.org/10.1186/s40494-024-01239-9

Cerrillo-Cuenca, E., Sepúlveda, M., & Guerrero-Bueno, Z. (2021). Independent component analysis (ICA): A statistical approach to the analysis of superimposed rock paintings. *Journal of Archaeological Science*, *125*, 105269. https://doi.org/10.1016/j.jas.2020.105269

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, *4*(1), 300–307.

Chakraborti, S. (1991). Graphical comparison of means with a standard in the one-way layout. *Journal of Applied Statistics*, *18*(3), 319–329. https://doi.org/10.1080/02664769100000031

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 151–155. https://doi.org/10.1016/S0022-5371(67)80066-5

Chayes, F. (1960). On correlation between variables of constant sum. *Journal of Geophysical Research*, *65*(12), 4185–4193. https://doi.org/10.1029/JZ065i012p04185

Chayes, F. (1962). Numerical correlation and petrographic variation. *The Journal of Geology*, *70*(4), 440–452. https://doi.org/10.1086/626835

Chen, C.-Y., & Chang, Y.-W. (2024). Missing data imputation using classification and regression trees. *PeerJ Computer Science*, *10*, e2119. https://doi.org/10.7717/peerj-cs.2119

Chen, H., Zhang, S., Chen, W., Mei, H., Zhang, J., Mercer, A., Liang, R., & Qu, H. (2015). Uncertainty-aware multidimensional ensemble data visualization and exploration. *IEEE Transactions on Visualization and Computer Graphics*, *21*(9), 1072–1086. https://doi.org/10.1109/TVCG.2015.2410278

Chen, Y., & Lu, L. (2023). The anomaly detector, semi-ssupervised classifier, and supervised classifier based on k-Nearest Neighbors in geochemical anomaly detection: A comparative study. *Mathematical Geosciences*, *55*(7), 1011–1033. https://doi.org/10.1007/s11004-022-10042-w

Chen, Y., Zhao, Q., & Lu, L. (2021). Combining the outputs of various k-nearest neighbor anomaly detectors to form a robust ensemble model for high-dimensional geochemical anomaly detection. *Journal of Geochemical Exploration*, *231*, 106875. https://doi.org/10.1016/j.gexplo.2021.106875

Chew, G., & Walczyk, T. (2012). A Monte Carlo approach for estimating measurement uncertainty using standard spreadsheet software. *Analytical and Bioanalytical Chemistry*, *402*(7), 2463–2469. https://doi.org/10.1007/s00216-011-5698-4

Chicco, D., Campagner, A., Spagnolo, A., Ciucci, D., & Jurman, G. (2025). The silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters. *PeerJ Computer Science*, *11*, e3309. https://doi.org/10.7717/peerj-cs.3309

Christensen, H. S., Borgbjerg, J., Børty, L., & Bøgsted, M. (2020). On Jones et al.'s method for extending Bland-Altman plots to limits of agreement with the mean for multiple observers. *BMC Medical Research Methodology*, *20*(1), 304. https://doi.org/10.1186/s12874-020-01182-w

Chung, C.-J. F. (1993). Estimation of covariance matrix from geochemical data with observations below detection limits. *Mathematical Geology*, *25*(7), 851–865. https://doi.org/10.1007/BF00891047

Coccato, A., & Caggiani, M. C. (2024). An overview of principal components analysis approaches in Raman studies of cultural heritage materials. *Journal of Raman Spectroscopy*, *55*(2), 125–147. https://doi.org/10.1002/jrs.6621

Cogswell, J., Neff, H., & Glascock, M. (1996). The effect of firing temperature on the elemental characterization of pottery. *Journal of Archaeological Science*, *23*(2), 283–287. https://doi.org/10.1006/jasc.1996.0026

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328. https://doi.org/10.1080/01621459.1994.10476871

Crema, E. R. (2025). Statistical modelling in archaeology: Some recent trends and future perspectives. *Journal of Archaeological Science*, *180*, 106295. https://doi.org/10.1016/j.jas.2025.106295

Cro, S., Morris, T. P., Kenward, M. G., & Carpenter, J. R. (2020). Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide. *Statistics in Medicine*, *39*(21), 2815–2842. https://doi.org/10.1002/sim.8569

Curran-Everett, D. (2013). Explorations in statistics: The analysis of ratios and normalized data. *Advances in Physiology Education*, *37*(3), 213–219. https://doi.org/10.1152/advan.00053.2013

Danzer, K., & Singer, R. (1985). Application of pattern recognition methods for the investigation of chemical homogeneity of solids. *Mikrochimica Acta*, *85*(3–4), 219–226. https://doi.org/10.1007/BF01198294

Daumas, M., Lefèvre, P., Martrille, L., Beauthier, J.-P., Louryan, S., & Polet, C. (2023). Method assessment and observer variation in age estimation: A comparative analysis of the Suchey-Brooks and the İşcan methods on an archaeological medieval population. *Journal of Archaeological Science: Reports*, *52*, 104234. https://doi.org/10.1016/j.jasrep.2023.104234

Davis, J. C. (2002). *Statistics and data analysis in geology* (3rd ed). J. Wiley.

De Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. *Applied Soft Computing*, *133*, 109924. https://doi.org/10.1016/j.asoc.2022.109924

De Castilho, M. V. (2004). A comparison of statistical techniques for detecting analytical bias in geoanalysis. *Geostandards and Geoanalytical Research*, *28*(2), 277–290. https://doi.org/10.1111/j.1751-908X.2004.tb00743.x

DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920965119. https://doi.org/10.1177/2515245920965119

Defernez, M., & Kemsley, E. K. (1997). The use and misuse of chemometrics for treating classification problems. *TrAC Trends in Analytical Chemistry*, *16*(4), 216–221. https://doi.org/10.1016/S0165-9936(97)00015-0

Deming, S. N., Palasota, J. A., & Nocerino, J. M. (1993). The geometry of multivariate object preprocessing. *Journal of Chemometrics*, *7*(5), 393–425. https://doi.org/10.1002/cem.1180070506

Deming, W. E. (1964). *Statistical adjustment of data* (Nachdr. d. Ausg. 1943). Dover Publ.

Demir, C., Keskin, S., & Şen, F. (2022). ANOM approach for statistical svaluation of some antioxidant enzyme activities. *Frontiers in Chemistry*, *10*, 894547. https://doi.org/10.3389/fchem.2022.894547

Derde, M. P., & Massart, D. L. (1986). UNEQ: A disjoint modelling technique for pattern recognition based on normal distribution. *Analytica Chimica Acta*, *184*, 33–51. https://doi.org/10.1016/S0003-2670(00)86468-5

Detalle, V., & Bai, X. (2022). The assets of laser-induced breakdown spectroscopy (LIBS) for the future of heritage science. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *191*, 106407. https://doi.org/10.1016/j.sab.2022.106407

Dhanoa, M. S., Sanderson, R., López, S., & France, J. (2016). Bivariate relationships incorporating method comparison: A review of linear regression methods. *CABI Reviews*, 1–15. https://doi.org/10.1079/PAVSNNR201611028

Dickinson, E., & High, K. E. (2022). The use of infrared spectroscopy and chemometrics to investigate deterioration in vegetable tanned leather: Potential applications in heritage science. *Heritage Science*, *10*(1), 65. https://doi.org/10.1186/s40494-022-00690-w

Doerffel, K., Herfurth, G., Liebich, V., & Wendlandt, E. (1991). The shape of CUSUM - an indicator for tendencies in a time series. *Fresenius' Journal of Analytical Chemistry*, *341*(9), 519–523. https://doi.org/10.1007/BF00328488

Doerffel, K., Küchler, L., & Meyer, N. (1990). Treatment of noisy data from distribution analysis using models from time-series analysis. *Fresenius' Journal of Analytical Chemistry*, *337*(7), 802–807. https://doi.org/10.1007/BF00322255

Domthong, U., Parikh, C. R., Kimmel, P. L., & Chinchilli, V. M. (2014). Assessing the agreement of biomarker data in the presence of left-censoring. *BMC Nephrology*, *15*(1), 144. https://doi.org/10.1186/1471-2369-15-144

Donais, M. K., Douglass, L., Ramundt, W. H., Bizzarri, C., & George, D. B. (2023). Handheld laser-induced breakdown spectroscopy for field archaeology: Characterization of Roman wall mortars and Etruscan ceramics. *Applied Spectroscopy Practica*, *1*(1), 27551857231175847. https://doi.org/10.1177/27551857231175847

Donner, A., & Zou, G. (2012). Closed-form confidence intervals for functions of the normal mean and standard deviation. *Statistical Methods in Medical Research*, *21*(4), 347–359. https://doi.org/10.1177/0962280210383082

Duewer, D. L., Kowalski, B. R., & Fasching, J. L. (1976). Improving the reliability of factor analysis of chemical data by utilizing the measured analytical uncertainty. *Analytical Chemistry*, *48*(13), 2002–2010. https://doi.org/10.1021/ac50007a048

Dunlap, W. P., Dietz, J., & Cortina, J. M. (1997). The spurious correlation of ratios that have common variables: A Monte Carlo examination of Pearson's formula. *The Journal of General Psychology*, *124*(2), 182–193. https://doi.org/10.1080/00221309709595516

Durvasula, S., Sharma, S., & Carter, K. (2012). Correcting the t statistic for measurement error. *Marketing Letters*, *23*(3), 671–682. https://doi.org/10.1007/s11002-012-9170-9

Eberl, M., Bell, C. S., Spencer-Smith, J., Raj, M., Sarubbi, A., Johnson, P. S., Rieth, A. E., Chaudhry, U., Estrada Aguila, R., & McBride, M. (2023). Machine learning–based identification of lithic microdebitage. *Advances in Archaeological Practice*, *11*(2), 152–163. https://doi.org/10.1017/aap.2022.35

Eder, F. M., Neelmeijer, C., Pearce, N. J. G., Bichler, M., Sterba, J. H., Ntaflos, T., & Merchel, S. (2013). Volcanic glass under fire – a comparison of three complementary analytical methods. *X-Ray Spectrometry*, *42*(6), 412–422. https://doi.org/10.1002/xrs.2493

Egozcue, J. J. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*(3), 279–300. https://doi.org/10.1023/A:1023818214614

Egozcue, J. J., Gozzi, C., Buccianti, A., & Pawlowsky-Glahn, V. (2024). Exploring geochemical data using compositional techniques: A practical guide. *Journal of Geochemical Exploration*, *258*, 107385. https://doi.org/10.1016/j.gexplo.2024.107385

Eksborg, S. (1981). Evaluation of method-comparison data. *Clinical Chemistry*, *27*(7), 1311–1312. https://doi.org/10.1093/clinchem/27.7.1311

Ellison, S. L. R., & Williams, A. (2013). *Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement* (3rd edn). https://www.eurachem.org/

Engle, M. A., & Chaput, J. (2023). Visualizing high dimensional structures in geochemical datasets using a combined compositional data analysis and Databionic swarm approach. *International Journal of Coal Geology*, *275*, 104303. https://doi.org/10.1016/j.coal.2023.104303

Ericson, J. E., Makishima, A., Mackenzie, J. D., & Berger, R. (1975). Chemical and physical properties of obsidian: A naturally occuring glass. *Journal of Non-Crystalline Solids*, *17*(1), 129–142. https://doi.org/10.1016/0022-3093(75)90120-9

Esmaeili, M., Ahmadi, M., & Kazemi, A. (2020). Kernel-based two-dimensional principal component analysis applied for parameterization in history matching. *Journal of Petroleum Science and Engineering*, *191*, 107134. https://doi.org/10.1016/j.petrol.2020.107134

Esmaeiloghli, S., & Tabatabaei, S. H. (2020a). Comparative analysis of geochemical data processing methods for allocation of anomalies and background. *Geochemistry International*, *58*(4), 472–485. https://doi.org/10.1134/S0016702920040084

Esmaeiloghli, S., & Tabatabaei, S. H. (2020b). Comparative analysis of geochemical data processing methods for allocation of anomalies and background. *Geochemistry International*, *58*(4), 472–485. https://doi.org/10.1134/S0016702920040084

Essack, Y., Hoffman, M., Rensburg, M., Van Wyk, J., Meyer, C., & Erasmus, R. (2009). A comparison of five glucometers in South Africa. *Journal of Endocrinology, Metabolism and Diabetes of South Africa*, *14*(2), 102–105. https://doi.org/10.1080/22201009.2009.10872202

Euser, A. M., Dekker, F. W., & Le Cessie, S. (2008). A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *Journal of Clinical Epidemiology*, *61*(10), 978–982. https://doi.org/10.1016/j.jclinepi.2007.11.003

Faber, N. M., Buydens, L. M. C., & Kateman, G. (1993). Standard errors in the eigenvalues of a cross-product matrix: Theory and applications. *Journal of Chemometrics*, *7*(6), 495–526. https://doi.org/10.1002/cem.1180070605

Farnham, I. M., Singh, A. K., Stetzenbach, K. J., & Johannesson, K. H. (2002). Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemometrics and Intelligent Laboratory Systems*, *60*(1–2), 265–281. https://doi.org/10.1016/S0169-7439(01)00201-5

Farrance, I., & Frenkel, R. (2012). Uncertainty of measurement: A review of the rules for calculating uncertainty components through functional relationships. *Clinical Biochemist Reviews*, *33*(2), 49–75.

Faucheux, L., Resche-Rigon, M., Curis, E., Soumelis, V., & Chevret, S. (2021). Clustering with missing and left-censored data: A simulation study comparing multiple-imputation-based procedures. *Biometrical Journal*, *63*(2), 372–393. https://doi.org/10.1002/bimj.201900366

Feenstra, S. (2006). Use of logarithmic-scale correlation plots to represent contaminant ratios for evaluation of subsurface environmental data. *Environmental Forensics*, *7*(2), 175–185. https://doi.org/10.1080/15275920600667153

Fernández, S., Cotos-Yáñez, T., Roca-Pardiñas, J., & Ordóñez, C. (2018). Geographically weighted principal components analysis to assess diffuse pollution sources of soil heavy metal: Application to rough mountain areas in Northwest Spain. *Geoderma*, *311*, 120–129. https://doi.org/10.1016/j.geoderma.2016.10.012

Feroze, N., & Aslam, M. (2022). Bayesian analysis of left-censored data using Weibull mixture model. *Soft Computing*, *26*(1), 375–394. https://doi.org/10.1007/s00500-021-06506-6

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 665–694. https://doi.org/10.1348/000711010X502733

Filzmoser, P., Gschwandtner, M., & Todorov, V. (2012). Review of sparse methods in regression and classification with application to chemometrics. *Journal of Chemometrics*, *26*(3–4), 42–51. https://doi.org/10.1002/cem.1418

Fish, P. R. (1978). Consistency in archaeological measurement and classification: A pilot study. *American Antiquity*, *43*(1), 86–89. https://doi.org/10.2307/279635

Fonville, J. M., Carter, C. L., Pizarro, L., Steven, R. T., Palmer, A. D., Griffiths, R. L., Lalor, P. F., Lindon, J. C., Nicholson, J. K., Holmes, E., & Bunch, J. (2013). Hyperspectral Visualization of Mass Spectrometry Imaging Data. *Analytical Chemistry*, *85*(3), 1415–1423. https://doi.org/10.1021/ac302330a

Frahm, E. (2014). Characterizing obsidian sources with portable XRF: Accuracy, reproducibility, and field relationships in a case study from Armenia. *Journal of Archaeological Science*, *49*, 105–125. https://doi.org/10.1016/j.jas.2014.05.003

Frahm, E. (2025). Archaeological obsidian sourcing: Looking from the first 60 years to the next. *Journal of Archaeological Science*, *177*, 106200. https://doi.org/10.1016/j.jas.2025.106200

Franceschi, P., & Wehrens, R. (2014). Self-organizing maps: A versatile tool for the automatic analysis of untargeted imaging datasets. *Proteomics*, *14*(7–8), 853–861. https://doi.org/10.1002/pmic.201300308

Francq, B. G., Berger, M., & Boachie, C. (2020). To tolerate or to agree: A tutorial on tolerance intervals in method comparison studies with BivRegBLS R Package. *Statistics in Medicine*, *39*(28), 4334–4349. https://doi.org/10.1002/sim.8709

Francq, B. G., & Govaerts, B. (2016). How to regress and predict in a Bland-Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models: How to regress and predict in a Bland-Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. *Statistics in Medicine*, *35*(14), 2328–2358. https://doi.org/10.1002/sim.6872

Francq, B. G., & Govaerts, B. B. (2014). Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies. *Journal de La Société Française de Statistique*, *155*(1), 23–45.

Freund, K. P. (2013). An assessment of the current applications and future directions of obsidian sourcing studies in archaeological research. *Archaeometry*, *55*(5), 779–793. https://doi.org/10.1111/j.1475-4754.2012.00708.x

Frey, M. E., Petersen, H. C., & Gerke, O. (2020). Nonparametric limits of agreement for small to moderate sample sizes: A simulation study. *Stats*, *3*(3), 343–355. https://doi.org/10.3390/stats3030022

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gäbler, H.-E., Rehder, S., Bahr, A., Melcher, F., & Goldmann, S. (2013). Cassiterite fingerprinting by LA-ICP-MS. *Journal of Analytical Atomic Spectrometry*, *28*(8), 1247. https://doi.org/10.1039/c3ja50106j

Gäbler, H.-E., Schink, W., & Gawronski, T. (2020). Data evaluation for cassiterite and coltan fingerprinting. *Minerals*, *10*(10), 926. https://doi.org/10.3390/min10100926

Gajić-Kvaščev, M. D., Marić-Stojanović, M. D., Jančić-Heinemann, R. M., Kvaščev, G. S., & Andrić, V. D. (2012). Non-destructive characterisation and classification of ceramic artefacts using pEDXRF and statistical pattern recognition. *Chemistry Central Journal*, *6*(1), 102. https://doi.org/10.1186/1752-153X-6-102

Garcia, S., Luengo, J., Sáez, J. A., López, V., & Herrera, F. (2013). A survey of discretization techniques: Taxonomy and empirical analysis in Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, *25*(4), 734–750. https://doi.org/10.1109/TKDE.2012.35

Gemignani, Z. (2014). *Data fluency: Empowering your organization with effective data communication*. Wiley.

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Gerke, O., & Möller, S. (2023). Modeling Bland–Altman limits of agreement with fractional polynomials—An example with the Agatston score for coronary calcification. *Axioms*, *12*(9), 884. https://doi.org/10.3390/axioms12090884

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Ghezelbash, R., Daviran, M., Maghsoudi, A., & Hajihosseinlou, M. (2025). Density based spatial clustering of applications with noise and fuzzy C-means algorithms for unsupervised mineral

prospectivity mapping. *Earth Science Informatics*, *18*(2), 217. https://doi.org/10.1007/s12145-025-01708-0

Ghouse, G., Rehman, A. U., & Bhatti, M. I. (2024). Understanding of causes of spurious associations: Problems and prospects. *Journal of Statistical Theory and Applications*, *23*(1), 44–66. https://doi.org/10.1007/s44199-024-00072-0

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Gille, M., Beaurepaire, P., Dumas, A., Yalamas, T., & Gayton, N. (2024). A Bayesian approach for the consideration of measurement errors. *Procedia CIRP*, *129*, 175–180. https://doi.org/10.1016/j.procir.2024.10.031

Ginau, A., Steiniger, D., Hartmann, R., Hartung, U., Schiestl, R., Altmeyer, M., Seeliger, M., & Wunderlich, J. (2020). What settlements leave behind—pXRF compositional data analysis of archaeological layers from Tell el-Fara'in (Buto, Egypt) using machine learning. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *546*, 109666. https://doi.org/10.1016/j.palaeo.2020.109666

Glascock, M. D. (2020). A systematic approach to geochemical sourcing of obsidian artifacts. *Scientific Culture*, *6*(2), 35–47. https://doi.org/10.5281/ZENODO.3724847

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., & Egozcue, J. J. (2016). It's all relative: Analyzing microbiome data as compositions. *Annals of Epidemiology*, *26*(5), 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Gnaden, D., & Holdaway, S. (2000). Understanding observer variation when recording stone artifacts. *American Antiquity*, *65*(4), 739–747. https://doi.org/10.2307/2694425

Godmer, A., Benzerara, Y., Varon, E., Veziris, N., Druart, K., Mozet, R., Matondo, M., Aubry, A., & Gianetto, Q. G. (2025). MSclassifR: An R package for supervised classification of mass spectra with machine learning methods. *Expert Systems with Applications*, *294*, 128796. https://doi.org/10.1016/j.eswa.2025.128796

Golitko, M., Meierhoff, J., Feinman, G. M., & Williams, P. R. (2012). Complexities of collapse: The evidence of Maya obsidian as revealed by social network graphical analysis. *Antiquity*, *86*(332), 507–523. https://doi.org/10.1017/S0003598X00062906

Goltz, H. H., & Smith, M. L. (2010). Yule-Simpson's paradox in research. *Practical Assessment, Research, and Evaluation*, *15*(1), 15. https://doi.org/10.7275/DGCC-JV81

Gómez-Carracedo, M. P., Andrade, J. M., López-Mahía, P., Muniategui, S., & Prada, D. (2014). A practical comparison of single and multiple imputation methods to handle complex missing data in air quality datasets. *Chemometrics and Intelligent Laboratory Systems*, *134*, 23–33. https://doi.org/10.1016/j.chemolab.2014.02.007

Gonçalves, T. R., Wentzell, P. D., Matsushita, M., & Valderrama, P. (2023). Exploring the scores: Procrustes analysis for comprehensive exploration of multivariate data. *Chemometrics and Intelligent Laboratory Systems*, *238*, 104841. https://doi.org/10.1016/j.chemolab.2023.104841

Gordon, C. C., & Buikstra, J. E. (1981). Soil pH, bone preservation, and sampling bias at mortuary sites. *American Antiquity*, *46*(3), 566–571. https://doi.org/10.2307/280601

Gortler, J., Spinner, T., Streeb, D., Weiskopf, D., & Deussen, O. (2020). Uncertainty-aware principal component analysis. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 822–831. https://doi.org/10.1109/TVCG.2019.2934812

Graham, D. J., & Castner, D. G. (2012). Multivariate analysis of ToF-SIMS data from multicomponent systems: The why, when, and how. *Biointerphases*, *7*(1), 49. https://doi.org/10.1007/s13758-012-0049-3

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for multi-class classification: An overview* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2008.05756

Greenacre, M. (2009). Power transformations in correspondence analysis. *Computational Statistics & Data Analysis*, *53*(8), 3107–3116. https://doi.org/10.1016/j.csda.2008.09.001

Greenacre, M. (2010). Log-ratio analysis is a limiting case of correspondence analysis. *Mathematical Geosciences*, *42*(1), 129–134. https://doi.org/10.1007/s11004-008-9212-2

Greenacre, M. (2021). Compositional data analysis. *Annual Review of Statistics and Its Application*, *8*(1), 271–299. https://doi.org/10.1146/annurev-statistics-042720-124436

Greenacre, M., Grunsky, E., & Bacon-Shone, J. (2021). A comparison of isometric and amalgamation logratio balances in compositional data analysis. *Computers & Geosciences*, *148*, 104621. https://doi.org/10.1016/j.cageo.2020.104621

Greenacre, M., & Wood, J. R. (2024). A comprehensive workflow for compositional data analysis in archaeometry, with code in R. *Archaeological and Anthropological Sciences*, *16*(10), 171. https://doi.org/10.1007/s12520-024-02070-w

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, *879*, 10–23. https://doi.org/10.1016/j.aca.2015.02.012

Grünfeld, K. (2005). Dealing with outliers and censored values in multi-element geochemical data – a visualization approach using XmdvTool. *Applied Geochemistry*, *20*(2), 341–352. https://doi.org/10.1016/j.apgeochem.2004.08.006

Grunsky, E. C., & Caritat, P. D. (2020). State-of-the-art analysis of geochemical data for mineral exploration. *Geochemistry: Exploration, Environment, Analysis*, *20*(2), 217–232. https://doi.org/10.1144/geochem2019-031

Gualtieri, S. (2020). Ceramic raw materials: How to establish the technological suitability of a raw material. *Archaeological and Anthropological Sciences*, *12*(8), 183. https://doi.org/10.1007/s12520-020-01135-w

Gut, Y., Boiret, M., Bultel, L., Renaud, T., Chetouani, A., Hafiane, A., Ginot, Y.-M., & Jennane, R. (2015). Application of chemometric algorithms to MALDI mass spectrometry imaging of pharmaceutical tablets. *Journal of Pharmaceutical and Biomedical Analysis*, *105*, 91–100. https://doi.org/10.1016/j.jpba.2014.11.047

Haber, G., Sampson, J., & Graubard, B. (2021). Bias due to Berkson error: Issues when using predicted values in place of observed covariates. *Biostatistics*, *22*(4), 858–872. https://doi.org/10.1093/biostatistics/kxaa002

Hagele, D., Krake, T., & Weiskopf, D. (2022). Uncertainty-aware multidimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, 1–10. https://doi.org/10.1109/TVCG.2022.3209420

Haig, B. D. (2003). What is a spurious correlation? *Understanding Statistics*, *2*(2), 125–132. https://doi.org/10.1207/S15328031US0202_03

Halfon, Efraim. (1985). Regression method in ecotoxicology: A better formulation using the geometric mean functional regression. *Environmental Science & Technology*, *19*(8), 747–749. https://doi.org/10.1021/es00138a016

Hall, M. M., Veeraraghavan, V. G., Rubin, H., & Winchell, P. G. (1977). The approximation of symmetric X-ray peaks by Pearson type VII distributions. *Journal of Applied Crystallography*, *10*(1), 66–68. https://doi.org/10.1107/S0021889877012849

Hamlett, A., Ryan, L., Serrano-Trespalacios, P., & Wolfinger, R. (2003). Mixed models for assessing correlation in the presence of replication. *Journal of the Air & Waste Management Association*, *53*(4), 442–450. https://doi.org/10.1080/10473289.2003.10466174

Hancock, R. G. V., & Carter, T. (2010). How reliable are our published archaeometric analyses? Effects of analytical techniques through time on the elemental analysis of obsidians. *Journal of Archaeological Science*, *37*(2), 243–250. https://doi.org/10.1016/j.jas.2009.10.004

Hancock, R. G. V., Hancock, K. E., & Hancock, J. K. (2008). Thoughts on the utility of a bivariate-splitting approach to Olmec ceramic data interpretation. *Archaeometry*, *50*(4), 710–726. https://doi.org/10.1111/j.1475-4754.2007.00353.x

Hancock, R. G. V., Michelaki, K., Mahaney, W. C., & Aufreiter, S. (2019). Justification for reassessing elemental analysis data of ceramics, sediments and lithics using rare earth element concentrations and ratios. *Archaeometry*, *61*(6), 1430–1445. https://doi.org/10.1111/arcm.12497

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

Hanselmann, M., Köthe, U., Kirchner, M., Renard, B. Y., Amstalden, E. R., Glunde, K., Heeren, R. M. A., & Hamprecht, F. A. (2009). Toward digital staining using imaging mass spectrometry and random forests. *Journal of Proteome Research*, *8*(7), 3558–3567. https://doi.org/10.1021/pr900253y

Hansen, C., Steinmetz, H., & Block, J. (2022). How to conduct a meta-analysis in eight steps: A practical guide. *Management Review Quarterly*, *72*(1), 1–19. https://doi.org/10.1007/s11301-021-00247-4

Hao, Z., Liu, K., Lian, Q., Song, W., Hou, Z., Zhang, R., Wang, Q., Sun, C., Li, X., & Wang, Z. (2024). Machine learning in laser-induced breakdown spectroscopy: A review. *Frontiers of Physics*, *19*(6), 62501. https://doi.org/10.1007/s11467-024-1427-2

Harmon, R. S., Shughrue, K. M., Remus, J. J., Wise, M. A., East, L. J., & Hark, R. R. (2011). Can the provenance of the conflict minerals columbite and tantalite be ascertained by laser-induced

breakdown spectroscopy? *Analytical and Bioanalytical Chemistry*, *400*(10), 3377–3382. https://doi.org/10.1007/s00216-011-5015-2

Harmon, R. S., Throckmorton, C. S., Haverstock, G., Baron, D., Yohe, R. M., Hark, R. R., & Knott, J. R. (2023). Connecting obsidian artifacts with their sources using multivariate statistical analysis of LIBS spectral signatures. *Minerals*, *13*(10), 1284. https://doi.org/10.3390/min13101284

Harris, P., Brunsdon, C., & Charlton, M. (2011). Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, *25*(10), 1717–1736. https://doi.org/10.1080/13658816.2011.554838

Härtel, B., Jonckheere, R., Krause, J., & Ratschbacher, L. (2022). Spurious age-eU associations in thermochronological data. *Earth and Planetary Science Letters*, *599*, 117870. https://doi.org/10.1016/j.epsl.2022.117870

Harter, H. L. (1963). Percentage points of the ratio of two ranges and power of the associated test. *Biometrika*, *50*(1–2), 187–194. https://doi.org/10.1093/biomet/50.1-2.187

Hartmann, C., & Massart, D. L. (1994). Display methods for visual comparison of the results of two measurement methods. *Journal of AOAC International*, *77*(5), 1318–1325. https://doi.org/10.1093/jaoac/77.5.1318

Haslauer, C. P., Meyer, J. R., Bárdossy, A., & Parker, B. L. (2017). Estimating a representative value and proportion of true zeros for censored analytical data with applications to contaminated site assessment. *Environmental Science & Technology*, *51*(13), 7502–7510. https://doi.org/10.1021/acs.est.6b05385

Hawkins, D. M. (2013). A general variance model in methods comparison. *Journal of Chemometrics*, *27*(11), 414–419. https://doi.org/10.1002/cem.2550

Hawkins, D. M., & Kraker, J. J. (2025). *Precision profile weighted Deming regression for methods comparison* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2508.02888

Hazenfratz Marks, R. (2016). Evaluating data: Uncertainty in ceramic analysis. In A. Hunt (Ed.), *The Oxford Handbook of Archaeological Ceramic Analysis* (pp. 47–57). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199681532.013.4

Hazenfratz, R., Mongeló, G. Z., Munita, C. S., & Neves, E. G. (2024). Comparison of log-ratio and log10 chemical elemental data analysis of Central Amazonian pottery and archaeological implications. *Archaeological and Anthropological Sciences*, *16*(5), 70. https://doi.org/10.1007/s12520-024-01965-y

Healy, J., & McInnes, L. (2024). Uniform manifold approximation and projection. *Nature Reviews Methods Primers*, *4*(1), 82. https://doi.org/10.1038/s43586-024-00363-x

Hegedus, E. J., & Moody, J. (2010). Clinimetrics corner: The many faces of selection bias. *Journal of Manual & Manipulative Therapy*, *18*(2), 69–73. https://doi.org/10.1179/106698110X12640740712699

Hein, A., & Kilikoglou, V. (2017). Compositional variability of archaeological ceramics in the eastern Mediterranean and implications for the design of provenance studies. *Journal of Archaeological Science: Reports*, *16*, 564–572. https://doi.org/10.1016/j.jasrep.2017.03.020

Hein, A., & Kilikoglou, V. (2020). Ceramic raw materials: How to recognize them and locate the supply basins: chemistry. *Archaeological and Anthropological Sciences*, *12*(8), 180. https://doi.org/10.1007/s12520-020-01129-8

Hellton, K. H., & Thoresen, M. (2014). The impact of measurement error on principal component analysis. *Scandinavian Journal of Statistics*, *41*(4), 1051–1063. https://doi.org/10.1111/sjos.12083

Helsel, D. R. (2005). More than obvious: Better methods for interpreting nondetect data. *Environmental Science & Technology*, *39*(20), 419A-423A. https://doi.org/10.1021/es053368a

Helsel, D. R. (2009). Much ado about next to nothing: Incorporating nondetects in science. *The Annals of Occupational Hygiene*. https://doi.org/10.1093/annhyg/mep092

Helsel, D. R. (2012). *Statistics for censored environmental data using Minitab and R* (2nd ed). Wiley.

Heymans, M. W., & Twisk, J. W. R. (2022). Handling missing data in clinical research. *Journal of Clinical Epidemiology*, *151*, 185–188. https://doi.org/10.1016/j.jclinepi.2022.08.016

Hibbert, D. B. (2003). The measurement uncertainty of ratios which share uncertainty components in numerator and denominator. *Accreditation and Quality Assurance*, *8*(5), 195–199. https://doi.org/10.1007/s00769-003-0615-y

Hites, R. A. (2019). Correcting for censored environmental measurements. *Environmental Science & Technology*, *53*(19), 11059–11060. https://doi.org/10.1021/acs.est.9b05042

Holland, P. W., & Rubin, D. B. (1982). On Lord's paradox. *ETS Research Report Series*, *1982*(2). https://doi.org/10.1002/j.2333-8504.1982.tb01321.x

Holmes, D. T., & Buhr, K. A. (2007). Error propagation in calculated ratios. *Clinical Biochemistry*, *40*(9–10), 728–734. https://doi.org/10.1016/j.clinbiochem.2006.12.014

Hornung, R. W., & Reed, L. D. (1990). Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*, *5*(1), 46–51. https://doi.org/10.1080/1047322X.1990.10389587

Horrocks, T., Holden, E.-J., Wedge, D., Wijns, C., & Fiorentini, M. (2019). Geochemical characterisation of rock hydration processes using t-SNE. *Computers & Geosciences*, *124*, 46–57. https://doi.org/10.1016/j.cageo.2018.12.005

Hroch, M. (2025). Exploratory multivariate analysis using R Language for method development in liquid chromatography. *Analytical and Bioanalytical Chemistry*, *417*(6), 1113–1125. https://doi.org/10.1007/s00216-024-05705-y

Hron, K., Coenders, G., Filzmoser, P., Palarea-Albaladejo, J., Faměra, M., & Matys Grygar, T. (2021). Analysing pairwise logratios revisited. *Mathematical Geosciences*, *53*(7), 1643–1666. https://doi.org/10.1007/s11004-021-09938-w

Hron, K., Templ, M., & Filzmoser, P. (2010). Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics & Data Analysis*, *54*(12), 3095–3107. https://doi.org/10.1016/j.csda.2009.11.023

Hutcheon, J. A., Chiolero, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ*, *340*(jun23 2), c2289–c2289. https://doi.org/10.1136/bmj.c2289

Huybrechts, T., Thas, O., Dewulf, J., & Van Langenhove, H. (2002). How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples. *Journal of Chromatography A*, *975*(1), 123–133. https://doi.org/10.1016/S0021-9673(02)01327-4

Hyltoft Petersen, P., Stöckl, D., Blaabjerg, O., Pedersen, B., Birkemose, E., Thienpont, L., Lassen, J. F., & Kjeldsen, J. (1997). Graphical interpretation of analytical data from comparison of a field method with a Reference Method by use of difference plots. *Clinical Chemistry*, *43*(11), 2039–2046. https://doi.org/10.1093/clinchem/43.11.2039

Hyslop, N. P., & White, W. H. (2009). Estimating precision using duplicate measurements. *Journal of the Air & Waste Management Association*, *59*(9), 1032–1039. https://doi.org/10.3155/1047-3289.59.9.1032

Ibáñez, J. J., Ortega, D., Campos, D., Khalidi, L., Méndez, V., & Teira, L. (2016). Developing a complex network model of obsidian exchange in the Neolithic Near East: Linear regressions, ethnographic models and archaeological data. *Paléorient*, *42*(2), 9–32.

Idkowiak, J., Dehairs, J., Schwarzerová, J., Olešová, D., Truong, J. X. M., Kvasnička, A., Eftychiou, M., Cools, R., Spotbeen, X., Jirásko, R., Veseli, V., Giampà, M., De Laat, V., Butler, L. M., Weckwerth, W., Friedecký, D., Demeulemeester, J., Hron, K., Swinnen, J. V., & Holčapek, M. (2025). Best practices and tools in R and Python for statistical processing and visualization of lipidomics and metabolomics data. *Nature Communications*, *16*(1), 8714. https://doi.org/10.1038/s41467-025-63751-1

Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). Cluster validity indices for automatic clustering: A comprehensive review. *Heliyon*, *11*(2), e41953. https://doi.org/10.1016/j.heliyon.2025.e41953

Innes, G. K., Bhondoekhan, F., Lau, B., Gross, A. L., Ng, D. K., & Abraham, A. G. (2021). The measurement error elephant in the room: Challenges and solutions to measurement error in epidemiology. *Epidemiologic Reviews*, *43*(1), 94–105. https://doi.org/10.1093/epirev/mxab011

Isles, P. D. F. (2020). The misuse of ratios in ecological stoichiometry. *Ecology*, *101*(11), e03153. https://doi.org/10.1002/ecy.3153

Jackson, D. A., & Somers, K. M. (1991). The spectre of 'spurious' correlations. *Oecologia*, *86*(1), 147–151. https://doi.org/10.1007/BF00317404

Jackson, J. D., & Dunlevy, J. A. (1988). Orthogonal least squares and the interchangeability of alternative proxy variables in the social sciences. *Statistician*, *37*(1), 7. https://doi.org/10.2307/2348374

Jasiewicz, J., Niedzielski, P., Krueger, M., Hildebrandt-Radke, I., & Michałowski, A. (2021). Elemental variability of prehistoric ceramics from postglacial lowlands and its implications for emerging of pottery traditions – An example from the pre-Roman Iron Age. *Journal of Archaeological Science: Reports*, *39*, 103177. https://doi.org/10.1016/j.jasrep.2021.103177

Jašková, P., Palarea-Albaladejo, J., Hron, K., Lachman, D., Templ, M., & Berland, M. (2025). Orthonormal pairwise logratio selection (OPALS) algorithm for compositional data analysis in high dimensions. *Bioinformatics Advances*, vbaf229. https://doi.org/10.1093/bioadv/vbaf229

Jaumot, J., & Tauler, R. (2015). Potential use of multivariate curve resolution for the analysis of mass spectrometry images. *Analyst*, *140*(3), 837–846. https://doi.org/10.1039/C4AN00801D

Jayalath, K. P., & Ng, H. K. T. (2020). Analysis of means approach in advanced designs. *Applied Stochastic Models in Business and Industry*, *36*(3), 501–520. https://doi.org/10.1002/asmb.2540

Jayalath, K. P., & Turner, J. (2021). Analysis of means (ANOM). Concepts and computations. *Applications and Applied Mathematics: An International Journal*, *16*(1), 75–96.

Ji, H., Zeng, D., Shi, Y., Wu, Y., & Wu, X. (2007). Semi-hierarchical correspondence cluster analysis and regional geochemical pattern recognition. *Journal of Geochemical Exploration*, *93*(2), 109–119. https://doi.org/10.1016/j.gexplo.2006.10.002

Ji, H., Zhu, Y., & Wu, X. (1995). Correspondence cluster analysis and its application in exploration geochemistry. *Journal of Geochemical Exploration*, *55*(1–3), 137–144. https://doi.org/10.1016/0375-6742(95)00025-9

Johnson, C. M. (1977). Factor analysis as a technique for exploring patterned variability in archaeological remains. *Plains Anthropologist*, *22*(78), 38–52. https://doi.org/10.1080/2052546.1977.11908852

Jones, E. A., Deininger, S.-O., Hogendoorn, P. C. W., Deelder, A. M., & McDonnell, L. A. (2012). Imaging mass spectrometry statistical analysis. *Journal of Proteomics*, *75*(16), 4962–4989. https://doi.org/10.1016/j.jprot.2012.06.014

Jones, M., Dobson, A., & O'Brian, S. (2011). A graphical method for assessing agreement with the mean between multiple observers using continuous measures. *International Journal of Epidemiology*, *40*(5), 1308–1313. https://doi.org/10.1093/ije/dyr109

Kahneman, D. (2011). *Thinking, fast and slow* (1st ed). Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds), *Heuristics and Biases* (1st edn, pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kane, M. T., & Mroch, A. A. (2020). Orthogonal regression, the Cleary criterion, and Lord's paradox: Asking the right questions. *ETS Research Report Series*, *2020*(1), 1–24. https://doi.org/10.1002/ets2.12298

Kenney, B. C. (1982). Beware of spurious self-correlations! *Water Resources Research*, *18*(4), 1041–1048. https://doi.org/10.1029/WR018i004p01041

Kerig, T., Hilpert, J., Strohm, S., Berger, D., Denis, S., Gauthier, E., Gibaja, J. F., Mallet, N., Massa, M., Mazzucco, N., Nessel, B., Pelegrin, J., Pétrequin, P., Sabatini, S., Schumacher, T. X., Serbe, B., & Wilkinson, T. (2023). Interlinking research: The *Big Exchange* project. *Antiquity*, *97*(394), e22. https://doi.org/10.15184/aqy.2023.78

Keskin, N., Kunter, B., Celik, H., Kaya, Ö., & Keskin, S. (2021). ANOM approach for the statistical evaluation of organic acid contents of clones of the grape variety 'Kalecik Karası'. *Mitteilungen Klosterneuburg*, *71*, 126–138.

Khan, S. S., & Madden, M. G. (2014). One-class classification: Taxonomy of study and review of techniques. *Knowledge Engineering Review*, *29*(3), 345–374. https://doi.org/10.1017/S026988891300043X

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00513

Kim, J.-H. (1999). Spurious correlation between ratios with a common divisor. *Statistics & Probability Letters*, *44*(4), 383–386. https://doi.org/10.1016/S0167-7152(99)00030-9

Kliegr, T., Bahník, Š., & Fürnkranz, J. (2021). A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence*, *295*, 103458. https://doi.org/10.1016/j.artint.2021.103458

Kohút, M., Westgate, J. A., Pearce, N. J. G., & Bačo, P. (2021). The Carpathian obsidians – Contribution to their FT dating and provenance (Zemplín, Slovakia). *Journal of Archaeological Science: Reports*, *37*, 102861. https://doi.org/10.1016/j.jasrep.2021.102861

Kowalski, B. R., Schatzki, T. F., & Stross, F. H. (1972). Classification of archaeological artifacts by applying pattern recognition to trace element data. *Analytical Chemistry*, *44*(13), 2176–2180. https://doi.org/10.1021/ac60321a002

Kowalski, C. J., Schneiderman, E. D., & Willis, S. M. (1995). PC program for simultaneously testing equality of means and variances for paired bivariate normal data. *International Journal of Bio-Medical Computing*, *38*(3), 239–242. https://doi.org/10.1016/S0020-7101(05)80006-6

Koziol, J. A., Feng, A. C., Yu, J., Griffin, N. M., & Schnitzer, J. E. (2008). Range charts for agreement in measurement comparison studies, with application to replicate mass spectrometry experiments. *Journal of Proteomics & Bioinformatics*, *01*(06), 287–292. https://doi.org/10.4172/jpb.1000036

Kragten, J. (1994). Tutorial review. Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique. *Analyst*, *119*(10), 2161. https://doi.org/10.1039/an9941902161

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, *37*(2), 233–243. https://doi.org/10.1002/aic.690370209

Krishna Reddy, M., Naveen Kumar, B., Ramu, Y., & Vengopala Rao, M. (2011). Bootstrap graphical test for equality of variances. *Electronic Journal of Applied Statistical Analysis*, *4*(2), 184–188. https://doi.org/10.1285/i20705948v4n2p184

Krishnapuram, R., & Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, *1*(2), 98–110. https://doi.org/10.1109/91.227387

Krishnapuram, R., & Keller, J. M. (1996). The possibilistic c-means algorithm: Insights and recommendations. *IEEE Transactions on Fuzzy Systems*, *4*(3), 385–393. https://doi.org/10.1109/91.531779

Krouwer, J. S. (2008). Why Bland–Altman plots should use X, not (Y + X)/2 when X is a reference method. *Statistics in Medicine*, *27*(5), 778–780. https://doi.org/10.1002/sim.3086

Kuwatani, T., Akaho, S., Nakamura, K., & Komai, T. (2025). Geochemical-process extraction and interpretation using matrix factorization: A framework for verifying effectiveness through forward modeling and inversion analysis. *Frontiers in Earth Science*, *13*, 1559321. https://doi.org/10.3389/feart.2025.1559321

Kuzmin, Y. V., Oppenheimer, C., & Renfrew, C. (2020). Global perspectives on obsidian studies in archaeology. *Quaternary International*, *542*, 41–53. https://doi.org/10.1016/j.quaint.2020.02.036

Kvascev, G., Kvascev, M. G., & Djurovic, Z. (2012). Radial basis function network based feature extraction for improvement the procedure of sourcing neolithic ceramics. *11th Symposium on Neural Network Applications in Electrical Engineering*, 95–100. https://doi.org/10.1109/NEUREL.2012.6419973

Larson, D. A. (1992). Analysis of variance with just summary statistics as input. *American Statistician*, *46*(2), 151–152. https://doi.org/10.1080/00031305.1992.10475872

Lee, S.-I. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems*, *3*(4), 369–385. https://doi.org/10.1007/s101090100064

Lenahan, M. J., Bristow, K. L., & De Caritat, P. (2011). Detecting induced correlations in hydrochemistry. *Chemical Geology*, *284*(1–2), 182–192. https://doi.org/10.1016/j.chemgeo.2011.02.018

Leng, L., Zhang, T., Kleinman, L., & Zhu, W. (2007). Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science. *Journal of Physics: Conference Series*, *78*, 012084. https://doi.org/10.1088/1742-6596/78/1/012084

Levine, E., & Domany, E. (2001). Resampling method for unsupervised estimation of cluster validity. *Neural Computation*, *13*(11), 2573–2593. https://doi.org/10.1162/089976601753196030

Liebal, U. W., Phan, A. N. T., Sudhakar, M., Raman, K., & Blank, L. M. (2020). Machine learning applications for mass spectrometry-based metabolomics. *Metabolites*, *10*(6), 243. https://doi.org/10.3390/metabo10060243

Liebich, V. (1990). Characterizing inhomogeneity by modified cusum technique. *Fresenius' Journal of Analytical Chemistry*, *338*(1), 72–73. https://doi.org/10.1007/BF00322789

Liebich, V., Ehrlich, G., Herrmann, U., Siegert, L., & Kluge, W. (1992). Characterization of the chemical homogeneity of solid-state materials by chemometric methods. Multivariate aspects and general recommendations. *Fresenius' Journal of Analytical Chemistry*, *343*(3), 251–258. https://doi.org/10.1007/BF00322370

Liebich, V., Ehrlich, G., Stahlberg, U., & Kluge, W. (1989). Characterization of the chemical homogeneity of solid-state standard materials by chemometric methods. *Fresenius' Zeitschrift Für Analytische Chemie*, *335*(8), 945–953. https://doi.org/10.1007/BF00466387

Linnet, K. (1990). Estimation of the linear relationship between the measurements of two methods with proportional errors. *Statistics in Medicine*, *9*(12), 1463–1473. https://doi.org/10.1002/sim.4780091210

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry*, *39*(3), 424–432. https://doi.org/10.1093/clinchem/39.3.424

Linnet, K. (1999). Limitations of the paired t-test for evaluation of method comparison data. *Clinical Chemistry*, *45*(2), 314–315. https://doi.org/10.1093/clinchem/45.2.314

Lischka, J. J. (1975). Broken K revisited: A short discussion of factor analysis. *American Antiquity*, *40*(2Part1), 220–227. https://doi.org/10.2307/279619

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Liu, B., Guo, K., & Zhang, L. (2016). Kernel principal component analysis in the application of geochemical comprehensive feature extraction. In N. J. Raju (Ed.), *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment* (pp. 15–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-18663-4_3

Liu, H., Yang, J., Ye, M., James, S. C., Tang, Z., Dong, J., & Xing, T. (2021). Using t-distributed stochastic neighbor embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. *Journal of Hydrology*, *597*, 126146. https://doi.org/10.1016/j.jhydrol.2021.126146

Liu, T., Yu, H., & Blair, R. H. (2022). Stability estimation for unsupervised clustering: A review. *WIREs Computational Statistics*, *14*(6), e1575. https://doi.org/10.1002/wics.1575

Liu, Y., & Brown, S. D. (2014). Imputation of left-censored data for cluster analysis. *Journal of Chemometrics*, *28*(3), 148–160. https://doi.org/10.1002/cem.2586

Liu, Y., Lyu, S., Hou, M., & Yin, Q. (2018). The comparison between NMF and ICA in pigment mixture identification of ancient Chinese paintings. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII–3*, 1169–1172. https://doi.org/10.5194/isprs-archives-XLII-3-1169-2018

Liu, Y., & Salvendy, G. (2009). Effects of measurement errors on psychometric measurements in ergonomics studies: Implications for correlations, ANOVA, linear regression, factor analysis, and linear discriminant analysis. *Ergonomics*, *52*(5), 499–511. https://doi.org/10.1080/00140130802392999

Lletí, R., Sarabia, L. A., Ortiz, M. C., Todeschini, R., & Colombini, M. P. (2003). Application of the Kohonen artificial neural network in the identification of proteinaceous binders in samples of panel painting using gas chromatography-mass spectrometry. *Analyst*, *128*(3), 281–286. https://doi.org/10.1039/b212509a

López-García, P. A., & Argote, D. L. (2023). Cluster analysis for the selection of potential discriminatory variables and the identification of subgroups in archaeometry. *Journal of Archaeological Science: Reports*, *49*, 104022. https://doi.org/10.1016/j.jasrep.2023.104022

López-García, P., Argote-Espino, D., & Fačevicová, K. (2018). Statistical processing of compositional data. The case of ceramic samples from the archaeological site of Xalasco, Tlaxcala, Mexico. *Journal of Archaeological Science: Reports*, *19*, 100–114. https://doi.org/10.1016/j.jasrep.2018.02.023

Loun, J., Novák, M., Cempírek, J., Škoda, R., Vašinová Galiová, M., Prokeš, L., Dosbaba, M., & Čopjaková, R. (2018). Geochemistry and secondary alterations of microlite from eluvial deposits in the Numbi mining area, South Kivu, Democratic Republic of the Congo. *The Canadian Mineralogist*, *56*(2), 203–220. https://doi.org/10.3749/canmin.1700091

Lovell, D. R., Chua, X.-Y., & McGrath, A. (2020). Counts: An outstanding challenge for log-ratio analysis of compositional data in the molecular biosciences. *NAR Genomics and Bioinformatics*, *2*(2), lqaa040. https://doi.org/10.1093/nargab/lqaa040

Lubbe, S. (2024). Bootstrapping cluster analysis solutions with the R package ClusBoot. *Austrian Journal of Statistics*, *53*(3), 1–19. https://doi.org/10.17713/ajs.v53i3.1169

Lubbe, S., Filzmoser, P., & Templ, M. (2021). Comparison of zero replacement strategies for compositional data with large numbers of zeros. *Chemometrics and Intelligent Laboratory Systems*, *210*, 104248. https://doi.org/10.1016/j.chemolab.2021.104248

Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., Bernstein, L., & Hartge, P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, *112*(17), 1691–1696. https://doi.org/10.1289/ehp.7199

Ludbrook, J. (1997). Special article comparing methods of measurement. *Clinical and Experimental Pharmacology and Physiology*, *24*(2), 193–203. https://doi.org/10.1111/j.1440-1681.1997.tb01807.x

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: A critical review. *Clinical and Experimental Pharmacology and Physiology*, *29*(7), 527–536. https://doi.org/10.1046/j.1440-1681.2002.03686.x

Ludbrook, J. (2010a). Confidence in Altman–Bland plots: A critical review of the method of differences. *Clinical and Experimental Pharmacology and Physiology*, *37*(2), 143–149. https://doi.org/10.1111/j.1440-1681.2009.05288.x

Ludbrook, J. (2010b). Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clinical and Experimental Pharmacology and Physiology*, *37*(7), 692–699. https://doi.org/10.1111/j.1440-1681.2010.05376.x

Lund, S. P., & Iyer, H. (2017). Likelihood Ratio as weight of forensic evidence: A closer look. *Journal of Research of the National Institute of Standards and Technology*, *122*, 27. https://doi.org/10.6028/jres.122.027

Lyman, R. L., & VanPool, T. L. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity*, *74*(3), 485–504. https://doi.org/10.1017/S0002731600048721

Ma, Q., Yan, A., Hu, Z., Li, Z., & Fan, B. (2000). Principal component analysis and artificial neural networks applied to the classification of Chinese pottery of neolithic age. *Analytica Chimica Acta*, *406*(2), 247–256. https://doi.org/10.1016/S0003-2670(99)00764-3

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, *110*, 63–73. https://doi.org/10.1016/j.jclinepi.2019.02.016

Mahmood, T., Riaz, M., Iqbal, A., & Mulenga, K. (2023). An improved statistical approach to compare means. *AIMS Mathematics*, *8*(2), 4596–4629. https://doi.org/10.3934/math.2023227

Mahon, K. I. (1996). The new "York" regression: Application of an improved statistical method to geochemistry. *International Geology Review*, *38*(4), 293–303. https://doi.org/10.1080/00206819709465336

Mameli, V., Sanna Angotzi, M., Farinini, E., Leardi, R., Lugliè, C., & Cannas, C. (2023). Intra-source provenance study on Monte Arci (Sardinia) obsidian by pXRF: Role of the data acquisition and analysis tools. *Heliyon*, *9*(3), e13958. https://doi.org/10.1016/j.heliyon.2023.e13958

Mansournia, M. A., Waters, R., Nazemipour, M., Bland, M., & Altman, D. G. (2021). Bland-Altman methods for comparing methods of measurement and response to criticisms. *Global Epidemiology*, *3*, 100045. https://doi.org/10.1016/j.gloepi.2020.100045

Maritan, L. (2019). Archaeo-ceramic 2.0: Investigating ancient ceramics using modern technological approaches. *Archaeological and Anthropological Sciences*, *11*(10), 5085–5093. https://doi.org/10.1007/s12520-019-00927-z

Maritan, L. (2020). Ceramic abandonment. How to recognise post-depositional transformations. *Archaeological and Anthropological Sciences*, *12*(8), 199. https://doi.org/10.1007/s12520-020-01141-y

Marsaglia, G. (2006). Ratios of normal variables. *Journal of Statistical Software*, *16*(4). https://doi.org/10.18637/jss.v016.i04

Martínez, À., Riu, J., & Xavier Rius, F. (2001). Multiple analytical method comparison using maximum likelihood principal component analysis and linear regression with errors in both axes. *Analytica Chimica Acta*, *446*(1–2), 145–156. https://doi.org/10.1016/S0003-2670(01)01032-7

Martyna, A., Gäbler, H.-E., Bahr, A., & Zadora, G. (2018). Geochemical wolframite fingerprinting – the likelihood ratio approach for laser ablation ICP-MS data. *Analytical and Bioanalytical Chemistry*, *410*(13), 3073–3091. https://doi.org/10.1007/s00216-018-1007-9

Massart, B., Guo, Q., Questier, F., Massart, D. L., Boucon, C., De Jong, S., & Vandeginste, B. G. M. (2001). Data structures and data transformations for clustering chemical data. *TrAC Trends in Analytical Chemistry*, *20*(1), 35–41. https://doi.org/10.1016/S0167-2940(01)90097-4

McCombie, G., Staab, D., Stoeckli, M., & Knochenmuss, R. (2005). Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Analytical Chemistry*, *77*(19), 6118–6124. https://doi.org/10.1021/ac051081q

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

McKinley, J. M., Hron, K., Grunsky, E. C., Reimann, C., De Caritat, P., Filzmoser, P., Van Den Boogaart, K. G., & Tolosana-Delgado, R. (2016). The single component geochemical map: Fact or fiction? *Journal of Geochemical Exploration*, *162*, 16–28. https://doi.org/10.1016/j.gexplo.2015.12.005

Meier, R. J. (2005). On art and science in curve-fitting vibrational spectra. *Vibrational Spectroscopy*, *39*(2), 266–269. https://doi.org/10.1016/j.vibspec.2005.03.003

Mellinger, M. (1987). Correspondence analysis: The method and its application. *Chemometrics and Intelligent Laboratory Systems*, *2*(1–3), 61–77. https://doi.org/10.1016/0169-7439(87)80086-2

Meloun, M., & Militký, J. (2012). *Statistical data analysis: A practical guide* (Reprinted). WPI, Woodhead Publ. India Pvt. Ltd.

Meyer, J. G. (2021). Deep learning neural network tools for proteomics. *Cell Reports Methods*, *1*(2), 100003. https://doi.org/10.1016/j.crmeth.2021.100003

Michael, N. A., Al Ibrahim, M. A., Scheibe, C., & Craigie, N. (2025). Best practices of utilizing principal component analysis in chemostratigraphic studies. *Applied Geochemistry*, *186*, 106355. https://doi.org/10.1016/j.apgeochem.2025.106355

Michelaki, K., & Hancock, R. G. V. (2011). Chemistry versus data dispersion: Is there a better way to assess and interpret archaeometric data? *Archaeometry*, *53*(6), 1259–1279. https://doi.org/10.1111/j.1475-4754.2011.00590.x

Mienye, I. D., & Swart, T. G. (2025). Deep Autoencoder Neural Networks: A comprehensive review and new perspectives. *Archives of Computational Methods in Engineering*, *32*(7), 3981–4000. https://doi.org/10.1007/s11831-025-10260-5

Miesch, A. T. (1979). Vector analysis of chemical variation in the lavas of Paricutin volcano, Mexico. *Journal of the International Association for Mathematical Geology*, *11*(4), 345–371. https://doi.org/10.1007/BF01029294

Mikolajewicz, N., & Komarova, S. V. (2019). Meta-analytic methodology for basic research: A practical guide. *Frontiers in Physiology*, *10*, 203. https://doi.org/10.3389/fphys.2019.00203

Militký, J., & Meloun, M. (2000). Prostorová autokorelace v analytické chemii. In *Zajištění kvality analytických výsledků: Sborník přednášek ze semináře 28.-30.3.2000 v Komorní Lhotce* (pp. 85–92). 2 Theta.

Milošević, D., Medeiros, A. S., Stojković Piperac, M., Cvijanović, D., Soininen, J., Milosavljević, A., & Predić, B. (2022). The application of uniform manifold approximation and projection (UMAP) for unconstrained ordination and classification of biological indicators in aquatic ecology. *Science of The Total Environment*, *815*, 152365. https://doi.org/10.1016/j.scitotenv.2021.152365

Miriello, D., & De Luca, R. (2025). Enhancing stone provenance studies through software built with language model artificial intelligence (AI): An example of ancient Calabrian quarries (southern Italy). *Archaeometry*, *67*(5), 1283–1308. https://doi.org/10.1111/arcm.13091

Mohammed, M. A., & Holder, R. (2012). Introducing analysis of means to medical statistics. *BMJ Quality & Safety*, *21*(6), 529–532. https://doi.org/10.1136/bmjqs-2011-000477

Moholy-Nagy, H. (2003). Source attribution and the utilization of obsidian in the Maya area. *Latin American Antiquity*, *14*(3), 301–310. https://doi.org/10.2307/3557562

Moholy-Nagy, H., Meierhoff, J., Golitko, M., & Kestle, C. (2013). An analysis of pXRF obsidian source attributions from Tikal, Guatemala. *Latin American Antiquity*, *24*(1), 72–97. https://doi.org/10.7183/1045-6635.24.1.72

Moholy-Nagy, H., & Nelson, F. W. (1990). New data on sources of obsidian artifacts from Tikal, Guatemala. *Ancient Mesoamerica*, *1*(1), 71–80. https://doi.org/10.1017/S0956536100000080

Möller, S., Debrabant, B., Halekoh, U., Petersen, A. K., & Gerke, O. (2021). An extension of the Bland–Altman plot for analyzing the agreement of more than two raters. *Diagnostics*, *11*(1), 54. https://doi.org/10.3390/diagnostics11010054

Montes, R. O. (2024). Frequentist and Bayesian tolerance intervals for setting specification limits for left-censored gamma distributed drug quality attributes. *Pharmaceutical Statistics*, *23*(2), 168–184. https://doi.org/10.1002/pst.2344

Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning*, *52*(1–2), 91–118. https://doi.org/10.1023/A:1023949509487

Moore, J. H., & Tatonetti, N. (2025). Vibe coding: A new paradigm for biomedical software development. *BioData Mining*, *18*(1), 46, s13040-025-00462–00469. https://doi.org/10.1186/s13040-025-00462-9

Moore, P. (2025). Green and the grey: Use of visual attributes in the analysis of New Zealand obsidian artefact assemblages. *Journal of Pacific Archaeology*, *15*(1), 5. https://doi.org/10.70460/jpa.v15i1.374

Moreno, C. (1996). A least-squares-based method for determining the ratio between two measured quantities. *Measurement Science and Technology*, *7*(2), 137–141. https://doi.org/10.1088/0957-0233/7/2/003

Morisaki, K., Izuho, M., & Ferguson, J. R. (2024). Non-Destructive energy dispersive X-ray fluorescence sourcing of glassy andesite and basalt artifacts in Upper Paleolithic southwestern Japan: A preliminary assessment. *Journal of Archaeological Science: Reports*, *57*, 104589. https://doi.org/10.1016/j.jasrep.2024.104589

Moustaki, I., & Papageorgiou, I. (2005). Latent class models for mixed variables with applications in Archaeometry. *Computational Statistics & Data Analysis*, *48*(3), 659–675. https://doi.org/10.1016/j.csda.2004.03.001

Mpaka, Y. W., & Von Der Heyden, B. P. (2024). Enhanced classification of pyrite generations based on mineral chemistry using uniform manifold approximation and projection (UMAP). *Journal of African Earth Sciences*, *218*, 105363. https://doi.org/10.1016/j.jafrearsci.2024.105363

Muhammad, L. N. (2023). Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *Journal of Clinical Investigation*, *133*(11), e171058. https://doi.org/10.1172/JCI171058

Müller, N. S., Hein, A., Georgakopoulou, M., Kilikoglou, V., & Kiriatzi, E. (2018). The effect of inter- and intra-source variation: A comparison between WD-XRF and NAA data from Cretan clay deposits. *Journal of Archaeological Science: Reports*, *21*, 929–937. https://doi.org/10.1016/j.jasrep.2017.09.023

Musumarra, G., & Fichera, M. (1998). Chemometrics and cultural heritage. *Chemometrics and Intelligent Laboratory Systems*, *44*(1–2), 363–372. https://doi.org/10.1016/S0169-7439(98)00069-0

Nawarathna, L. S., & Choudhary, P. K. (2013). Measuring agreement in method comparison studies with heteroscedastic measurements. *Statistics in Medicine*, *32*(29), 5156–5171. https://doi.org/10.1002/sim.5955

Nawarathna, L. S., & Choudhary, P. K. (2015). A heteroscedastic measurement error model for method comparison data with replicate measurements. *Statistics in Medicine*, *34*(7), 1242–1258. https://doi.org/10.1002/sim.6424

Nelson, P. R., Wludyka, P. S., & Copeland, K. A. F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718362

Němec, T. (2015). *Critical analysis of the slope method for estimation of ice-water interfacial energy from ice nucleation experimental data (with reviews)* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.1503.02967

Ng, S., Masarone, S., Watson, D., & Barnes, M. R. (2023). The benefits and pitfalls of machine learning for biomarker discovery. *Cell and Tissue Research*, *394*(1), 17–31. https://doi.org/10.1007/s00441-023-03816-z

Nguyen, T., & Jiang, J. (2011). Simple estimation of hidden correlation in repeated measures. *Statistics in Medicine*, *30*(29), 3403–3415. https://doi.org/10.1002/sim.4366

Nicholls, J. (1988). The statistics of Pearce element diagrams and the Chayes closure problem. *Contributions to Mineralogy and Petrology*, *99*(1), 11–24. https://doi.org/10.1007/BF00399361

Nicholls, J., & Russell, J. K. (2016). Igneous rock associations 20. Pearce element ratio diagrams: Linking geochemical data to magmatic processes. *Geoscience Canada*, *43*(2), 133. https://doi.org/10.12789/geocanj.2016.43.095

Nickerson, C. A. E. (1997). A note on 'A concordance correlation coefficient to evaluate reproducibility'. *Biometrics*, *53*(4), 1503. https://doi.org/10.2307/2533516

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nounou, M. N., Bakshi, B. R., Goel, P. K., & Shen, X. (2002). Bayesian principal component analysis. *Journal of Chemometrics*, *16*(11), 576–595. https://doi.org/10.1002/cem.759

Oliveri, P. (2017). UNEQ – An intuitive probabilistic method for class-modelling approaches. *NIR News*, *28*(3), 4–6. https://doi.org/10.1177/0960336017703246

Oonk, S., & Spijker, J. (2015). A supervised machine-learning approach towards geochemical predictive modelling in archaeology. *Journal of Archaeological Science*, *59*, 80–88. https://doi.org/10.1016/j.jas.2015.04.002

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, *27*(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Ortega, D., Ibañez, J. J., Khalidi, L., Méndez, V., Campos, D., & Teira, L. (2014). Towards a multi-agent-based modelling of obsidian exchange in the Neolithic Near East. *Journal of Archaeological Method and Theory*, *21*(2), 461–485. https://doi.org/10.1007/s10816-013-9196-1

Ortega-Villa, A. M., Liu, D., Ward, M. H., & Albert, P. S. (2021). New insights into modeling exposure measurements below the limit of detection. *Environmental Epidemiology*, *5*(1), e116. https://doi.org/10.1097/EE9.0000000000000116

Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., & Canals, A. (2005). Relative vs. absolute statistical analysis of compositions: A comparative study of surface waters of a Mediterranean river. *Water Research*, *39*(7), 1404–1414. https://doi.org/10.1016/j.watres.2005.01.012

Ownby, M. F., Huntley, D. L., & Peeples, M. A. (2014). A combined approach: Using NAA and petrography to examine ceramic production and exchange in the American southwest. *Journal of Archaeological Science*, *52*, 152–162. https://doi.org/10.1016/j.jas.2014.08.018

Pagnin, L., Brunnbauer, L., Wiesinger, R., Limbeck, A., & Schreiner, M. (2020). Multivariate analysis and laser-induced breakdown spectroscopy (LIBS): A new approach for the spatially resolved classification of modern art materials. *Analytical and Bioanalytical Chemistry*, *412*(13), 3187–3198. https://doi.org/10.1007/s00216-020-02574-z

Paine, R. R., & Harpending, H. C. (1998). Effect of sample bias on paleodemographic fertility estimates. *American Journal of Physical Anthropology*, *105*(2), 231–240. https://doi.org/10.1002/(SICI)1096-8644(199802)105:2%253C231::AID-AJPA9%253E3.0.CO;2-X

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2013). Values below detection limit in compositional chemical data. *Analytica Chimica Acta*, *764*, 32–43. https://doi.org/10.1016/j.aca.2012.12.029

Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2015). zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics and Intelligent Laboratory Systems*, *143*, 85–96. https://doi.org/10.1016/j.chemolab.2015.02.019

Pallavi, Joshi, S., Singh, D., Kaur, M., & Lee, H.-N. (2022). Comprehensive review of orthogonal regression and its applications in different domains. *Archives of Computational Methods in Engineering*, *29*(6), 4027–4047. https://doi.org/10.1007/s11831-022-09728-5

Pallmann, P., & Hothorn, L. A. (2016). Analysis of means: A generalized approach using R. *Journal of Applied Statistics*, *43*(8), 1541–1560. https://doi.org/10.1080/02664763.2015.1117584

Palmer, A. D., Bunch, J., & Styles, I. B. (2015). The use of random projections for the analysis of mass spectrometry imaging data. *Journal of the American Society for Mass Spectrometry*, *26*(2), 315–322. https://doi.org/10.1007/s13361-014-1024-7

Pan, G., & Harris, D. P. (1991). A new multidimensional scaling technique based upon associations of triple objects—Pijk and its application to the analysis of geochemical data. *Mathematical Geology*, *23*(6), 861–886. https://doi.org/10.1007/BF02068781

Pang, J., & Liu, X. (2023). Evaluation of missing data imputation methods for human osteometric measurements. *American Journal of Biological Anthropology*, *181*(4), 666–676. https://doi.org/10.1002/ajpa.24787

Pankowska, P., Oberski, D., Garnier-Villarreal, M., & Pavlopoulos, D. (2025). The effect of measurement error on clustering. *Quality & Quantity*, *59*(5), 4825–4860. https://doi.org/10.1007/s11135-025-02177-9

Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and avoiding bias in research: *Plastic and Reconstructive Surgery*, *126*(2), 619–625. https://doi.org/10.1097/PRS.0b013e3181de24bc

Papageorgiou, I. (2020). Ceramic investigation: How to perform statistical analyses. *Archaeological and Anthropological Sciences*, *12*(9), 210. https://doi.org/10.1007/s12520-020-01142-x

Papageorgiou, I., Baxter, M. J., & Cau, M. A. (2001). Model-based cluster analysis of artefact compositional data. *Archaeometry*, *43*(4), 571–588. https://doi.org/10.1111/1475-4754.00037

Papageorgiou, I., & Liritzis, I. (2007). Multivariate mixture of normals with unknown number of components: An application to cluster Neolithic ceramics from Aegean and Asia Minor using portable XRF. *Archaeometry*, *49*(4), 795–813. https://doi.org/10.1111/j.1475-4754.2007.00336.x

Parish, R., & Durham, E. (2015). The problems with visual identification of Dover and Fort Payne chert. *Southeastern Archaeology*, *34*(1), 71–83. https://doi.org/10.1179/2168472314Y.0000000003

Park, Y. W., & Klabjan, D. (2018). Three iteratively reweighted least squares algorithms for L 1-norm principal component analysis. *Knowledge and Information Systems*, *54*(3), 541–565. https://doi.org/10.1007/s10115-017-1069-6

Pashkova, G. V., Statkus, M. A., Mukhamedova, M. M., Finkelshtein, A. L., Abdrashitova, I. V., Belozerova, O. Yu., Chubarov, V. M., Amosova, A. A., Maltsev, A. S., Demonterova, E. I., & Shergin, D. L. (2023). A workflow for uncertainty assessment in elemental analysis of archaeological ceramics: A case study of Neolithic coarse pottery from Eastern Siberia. *Heritage*, *6*(5), 4434–4450. https://doi.org/10.3390/heritage6050234

Passing, H., & Bablok, W. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *Journal of Clinical Chemistry and Clinical Biochemistry*, *21*(11), 709–720. https://doi.org/10.1515/cclm.1983.21.11.709

Paul, A., & De Boves Harrington, P. (2021). Chemometric applications in metabolomic studies using chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry*, *135*, 116165. https://doi.org/10.1016/j.trac.2020.116165

Paul, B., Woodhead, J. D., Paton, C., Hergt, J. M., Hellstrom, J., & Norris, C. A. (2014). Towards a method for quantitative LA-ICP-MS imaging of multi-phase assemblages: Mineral identification and analysis correction procedures. *Geostandards and Geoanalytical Research*, *38*(3), 253–263. https://doi.org/10.1111/j.1751-908X.2014.00270.x

Pawlowicz, L. M., & Downum, C. E. (2021). Applications of deep learning to decorated ceramic typology and classification: A case study using Tusayan White Ware from Northeast Arizona. *Journal of Archaeological Science*, *130*, 105375. https://doi.org/10.1016/j.jas.2021.105375

Pawlowsky-Glahn, V., & Egozcue, J. J. (2020). Compositional data in geostatistics: A log-ratio based framework to analyze regionalized compositions. *Mathematical Geosciences*, *52*(8), 1067–1084. https://doi.org/10.1007/s11004-020-09873-2

Pearl, J. (2016). Lord's paradox revisited – (Oh Lord! Kumbaya!). *Journal of Causal Inference*, *4*(2), 20160021. https://doi.org/10.1515/jci-2016-0021

Peisach, M., Jacobson, L., Boulle, G. J., Gihwala, D., & Underhill, L. G. (1982). Multivariate analysis of trace elements determined in archaeological materials and its use for characterisation. *Journal of Radioanalytical Chemistry*, *69*(1–2), 349–364. https://doi.org/10.1007/BF02515934

Perera, P., Oza, P., & Patel, V. M. (2021). *One-class classification: A survey* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2101.03064

Pesenson, M. Z., Suram, S. K., & Gregoire, J. M. (2015). Statistical analysis and interpolation of compositional data in materials science. *ACS Combinatorial Science*, *17*(2), 130–136. https://doi.org/10.1021/co5001458

Pesonen, M., Pesonen, H., & Nevalainen, J. (2015). Covariance matrix estimation for left-censored data. *Computational Statistics & Data Analysis*, *92*, 13–25. https://doi.org/10.1016/j.csda.2015.06.005

Pham-Gia, T., Turkkan, N., & Marchand, E. (2006). Density of the ratio of two normal random variables and applications. *Communications in Statistics - Theory and Methods*, *35*(9), 1569–1591. https://doi.org/10.1080/03610920600683689

Pierce, D. E. (2015). Visual and geochemical analyses of obsidian source use at San Felipe Aztatán, Mexico. *Journal of Anthropological Archaeology*, *40*, 266–279. https://doi.org/10.1016/j.jaa.2015.09.002

Pingitore, N. E., & Engle, M. A. (2022). Compositional closure—Its origin lies not in mathematics but rather in nature Itself. *Minerals*, *12*(1), 74. https://doi.org/10.3390/min12010074

Pîrnău, R. G., Patriche, C. V., Roşca, B., Vasiliniuc, I., Vornicu, N., & Stanc, S. (2020). Soil spatial patterns analysis at the ancient city of Ibida (Dobrogea, SE Romania), via portable X-ray fluorescence spectrometry and multivariate statistical methods. *CATENA*, *189*, 104506. https://doi.org/10.1016/j.catena.2020.104506

Pollman, C. D., & Axelrad, D. M. (2014). Mercury bioaccumulation factors and spurious correlations. *Science of The Total Environment*, *496*, vi–xii. https://doi.org/10.1016/j.scitotenv.2014.07.050

Pollock, M. A., Jefferson, S. G., Kane, J. W., Lomax, K., MacKinnon, G., & Winnard, C. B. (1992). Method Comparison—A Different Approach. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, *29*(5), 556–560. https://doi.org/10.1177/000456329202900512

Portier, K. (2001). Statistical issues in assessing anthropogenic background for arsenic. *Environmental Forensics*, *2*(2), 155–160. https://doi.org/10.1006/enfo.2001.0051

Prasad, M., Postma, G., Franceschi, P., Buydens, L. M. C., & Jansen, J. J. (2022). Evaluation and comparison of unsupervised methods for the extraction of spatial patterns from mass spectrometry imaging data (MSI). *Scientific Reports*, *12*(1), 15687. https://doi.org/10.1038/s41598-022-19365-4

Praus, P. (2019). Robust multivariate analysis of compositional data of treated wastewaters. *Environmental Earth Sciences*, *78*(8), 248. https://doi.org/10.1007/s12665-019-8248-6

Preece, D. A. (1982). t Is for trouble (and textbooks): A critique of some examples of the paired-samples t-test. *The Statistician*, *31*(2), 169. https://doi.org/10.2307/2987888

Přichystal, A. (2018). State of arts: The Carpathian obsidians in the Czech Republic. *Archeometriai Műhely*, *15*(3), 231–240.

Prokeš, L. (2012). Prvková analýza dolnověstonické populace Na Pískách. In *Antropologická analýza raně středověké populace z Dolních Věstonic—Na Pískách* (pp. 121–132). Moravské zemské muzeum.

Prokeš, L., & Hegrová, J. (2006). Chemická analýza kosterních pozůstatků příslušníků knížecí rodiny Dietrichsteinů. In *Dietrichsteinové z Mikulova. Výsledky antropologického výzkumu vybraných příslušníků rodu* (pp. 177–197). Masaryk University.

Prokeš, L., Hegrová, J., Průšová, B., Baroň, M., Hablovičová, B., Sochor, J., & Ličbinský, R. (2025). Impact of traffic intensity and vehicular emissions on heavy metal content in vineyard soils, grapes, and wine: A comparative study of two vineyards in South Moravia (Czech Republic). *Environmental Geochemistry and Health*, *47*(6), 216. https://doi.org/10.1007/s10653-025-02530-9

Prokeš, L., Petřík, J., Beran, V., Vlach, M., Humpola, D., Šmerda, J., Hlavica, M., & Sova, E. (2013). Možnosti statistické a prostorové analýzy hodnot půdních fosfátů na příkladě sekundárně narušených hrobů z Hodonic a Kyjova. In O. Chvojka (Ed.), *Archeologická prospekce a nedestruktivní archeologie v Jihočeském kraji, kraji Vysočina, Jihomoravském kraji a v Dolním Rakousku: Sborník z konference, Jindřichův Hradec 6.3. - 7.3. 2013* (pp. 229–236). Jihočeské muzeum v Českých Budějovicích.

Prokeš, L., Vašinová Galiová, M., Hušková, S., Vaculovič, T., Hrdlička, A., Mason, A. Z., Neff, H., Přichystal, A., & Kanický, V. (2015). Laser microsampling and multivariate methods in provenance studies of obsidian artefacts. *Chemical Papers*, *69*(6). https://doi.org/10.1515/chempap-2015-0019

Prokeš, L., Vystrčilová, M., Shbat, A., Kanický, V., & Otruba, V. (2004). Chemická analýza kosterních pozůstatků z františkánského kláštera v Uherském Hradišti (Příspěvek k identifikaci pozůstatků biskupa Jana Filipce). In *Sborník k sedmdesátinám RNDr. Emanuela Opravila, CSc.* (Vol. 5, pp. 259–269).

Proschan, M. A., & Leifer, E. S. (2006). Comparison of two or more measurement techniques to a standard. *Contemporary Clinical Trials*, *27*(5), 472–482. https://doi.org/10.1016/j.cct.2006.02.008

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, *2*, 927312. https://doi.org/10.3389/fbinf.2022.927312

Purtill, M. P. (2023). More than just Clovis: The broad impact of sampling bias on archaeological site distributions. *Journal of Archaeological Method and Theory*, *30*(3), 1002–1027. https://doi.org/10.1007/s10816-022-09581-7

Qi, J., Zhang, T., Tang, H., & Li, H. (2018). Rapid classification of archaeological ceramics via laser-induced breakdown spectroscopy coupled with random forest. *Spectrochimica Acta Part B: Atomic Spectroscopy*, *149*, 288–293. https://doi.org/10.1016/j.sab.2018.09.006

Rácz, A., Héberger, K., Rajkó, R., & Elek, J. (2013). Classification of Hungarian medieval silver coins using x-ray fluorescent spectroscopy and multivariate data analysis. *Heritage Science*, *1*(1), 2. https://doi.org/10.1186/2050-7445-1-2

Rajalahti, T., & Kvalheim, O. M. (2011). Multivariate data analysis in pharmaceutics: A tutorial review. *International Journal of Pharmaceutics*, *417*(1–2), 280–290. https://doi.org/10.1016/j.ijpharm.2011.02.019

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the influence of normalization/transformation process on the accuracy of supervised classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Ramil, A., López, A. J., & Yáñez, A. (2008). Application of artificial neural networks for the rapid classification of archaeological ceramics by means of laser induced breakdown spectroscopy (LIBS). *Applied Physics A*, *92*(1), 197–202. https://doi.org/10.1007/s00339-008-4481-7

Rao, C. V. (2005). Analysis of means—A review. *Journal of Quality Technology*, *37*(4), 308–315. https://doi.org/10.1080/00224065.2005.11980334

Rao, J. N. K., & Beegle, L. D. (1967). A Monte Carlo study of some ratio estimators. *Sankhyā: The Indian Journal of Statistics, Series B*, *29*(1), 47–56.

Reimann, C., Filzmoser, P., Fabian, K., Hron, K., Birke, M., Demetriades, A., Dinelli, E., & Ladenberger, A. (2012). The concept of compositional data analysis in practice—Total major element concentrations in agricultural and grazing land soils of Europe. *Science of The Total Environment*, *426*, 196–210. https://doi.org/10.1016/j.scitotenv.2012.02.032

Reimann, C., Filzmoser, P., & Garrett, R. G. (2002). Factor analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry*, *17*(3), 185–206. https://doi.org/10.1016/S0883-2927(01)00066-X

Reimann, C., Filzmoser, P., Garrett, R. G., & Dutter, R. (2008). *Statistical data analysis explained: Applied environmental statistics with R*. John Wiley & Sons.

Remolá, J. A., Larrechi, M. S., & Rius, F. X. (1993). Chemometric characterization of 5th century A.D. amphora-producing centres in the Mediterranean. *Talanta*, *40*(11), 1749–1757. https://doi.org/10.1016/0039-9140(93)80093-7

Remolà, J. A., Lozano, J., Ruisánchez, I., Larrechi, M. S., Rius, F. X., & Zupan, J. (1996). New chemometric tools to study the origin of amphorae produced in the Roman Empire. *TrAC Trends in Analytical Chemistry*, *15*(3), 137–151. https://doi.org/10.1016/0165-9936(95)00091-7

Rice, P. M., & Saffer, M. E. (1982). Cluster analysis of mixed-level data: Pottery provenience as an example. *Journal of Archaeological Science*, *9*(4), 395–409. https://doi.org/10.1016/0305-4403(82)90045-0

Rietjens, M. (1995). Reduction of error propagation due to normalization: Effect of error propagation and closure on spurious correlations. *Analytica Chimica Acta*, *316*(2), 205–215. https://doi.org/10.1016/0003-2670(95)00355-4

Riggs, D. S., Guarnieri, J. A., & Addelman, S. (1978). Fitting straight lines when both variables are subject to error. *Life Sciences*, *22*(13–15), 1305–1360. https://doi.org/10.1016/0024-3205(78)90098-X

Ringrose, T. J. (2012). Bootstrap confidence regions for correspondence analysis. *Journal of Statistical Computation and Simulation*, *82*(10), 1397–1413. https://doi.org/10.1080/00949655.2011.579968

Rittner, M., & Müller, W. (2012). 2D mapping of LA-ICPMS trace element distributions using R. *Computers & Geosciences*, *42*, 152–161. https://doi.org/10.1016/j.cageo.2011.07.016

Riu, J., & Giussani, B. (2024). Analytical chemistry meets art: The transformative role of chemometrics in cultural heritage preservation. *Chemometrics and Intelligent Laboratory Systems*, *247*, 105095. https://doi.org/10.1016/j.chemolab.2024.105095

Riu, J., & Rius, F. X. (1996). Assessing the accuracy of analytical methods using linear regression with errors in both axes. *Analytical Chemistry*, *68*(11), 1851–1857. https://doi.org/10.1021/ac951217s

Riu, J., & Rius, F. X. (1997). Method comparison using regression with uncertainties in both axes. *TrAC Trends in Analytical Chemistry*, *16*(4), 211–216. https://doi.org/10.1016/S0165-9936(97)00014-9

Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV- coefficient. *Applied Statistics*, *25*(3), 257. https://doi.org/10.2307/2347233

Robles-Berumen, H., Zafra, A., & Ventura, S. (2024). A survey of genetic algorithms for clustering: Taxonomy and empirical analysis. *Swarm and Evolutionary Computation*, *91*, 101720. https://doi.org/10.1016/j.swevo.2024.101720

Rodríguez, J. G., Tueros, I., Borja, A., Belzunce, M. J., Franco, J., Solaun, O., Valencia, V., & Zuazo, A. (2006). Maximum likelihood mixture estimation to determine metal background values in estuarine and coastal sediments within the European Water Framework Directive. *Science of The Total Environment*, *370*(2–3), 278–293. https://doi.org/10.1016/j.scitotenv.2006.08.035

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Romanenko, S. V., Stromberg, A. G., & Pushkareva, T. N. (2006). Modeling of analytical peaks: Peaks properties and basic peak functions. *Analytica Chimica Acta*, *580*(1), 99–106. https://doi.org/10.1016/j.aca.2006.07.050

Roumeliotis, S., Abd ElHafeez, S., Jager, K. J., Dekker, F. W., Stel, V. S., Pitino, A., Zoccali, C., & Tripepi, G. (2021). Be careful with ecological associations. *Nephrology*, *26*(6), 501–505. https://doi.org/10.1111/nep.13861

Roy, A. (2006). Estimating correlation coefficient between two variables with repeated observations using mixed effects model. *Biometrical Journal*, *48*(2), 286–301. https://doi.org/10.1002/bimj.200510192

Roy, T. (1994). Fitting a straight line when both variables are subject to error: Pharmaceutical applications. *Journal of Pharmaceutical and Biomedical Analysis*, *12*(10), 1265–1269. https://doi.org/10.1016/0731-7085(94)00057-3

Ruschioni, G., Malchiodi, D., Zanaboni, A. M., & Bonizzoni, L. (2023). Supervised learning algorithms as a tool for archaeology: Classification of ceramic samples described by chemical element concentrations. *Journal of Archaeological Science: Reports*, *49*, 103995. https://doi.org/10.1016/j.jasrep.2023.103995

Saccenti, E., Hendriks, M. H. W. B., & Smilde, A. K. (2020). Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models. *Scientific Reports*, *10*(1), 438. https://doi.org/10.1038/s41598-019-57247-4

Sadeghi, M., Casey, P., Carranza, E. J. M., & Lynch, E. P. (2024). Principal components analysis and K-means clustering of till geochemical data: Mapping and targeting of prospective areas for lithium exploration in Västernorrland Region, Sweden. *Ore Geology Reviews*, *167*, 106002. https://doi.org/10.1016/j.oregeorev.2024.106002

Sadler, W. A. (2019a). Methods comparison biases due to differing uncertainties and data censoring. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, *56*(5), 608–612. https://doi.org/10.1177/0004563219859920

Sadler, W. A. (2019b). Using the variance function to generalize Bland–Altman analysis. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, *56*(2), 198–203. https://doi.org/10.1177/0004563218806560

Sanchez-Pinto, L. N., Venable, L. R., Fahrenbach, J., & Churpek, M. M. (2018). Comparison of variable selection methods for clinical predictive modeling. *International Journal of Medical Informatics*, *116*, 10–17. https://doi.org/10.1016/j.ijmedinf.2018.05.006

Saremi, M., Hezarkhani, A., Mirzabozorg, S. A. A. S., DehghanNiri, R., Shirazy, A., & Shirazi, A. (2025). Unsupervised anomaly detection for mineral prospectivity mapping using Isolation Forest and Extended Isolation Forest algorithms. *Minerals*, *15*(4), 411. https://doi.org/10.3390/min15040411

Saunders, S. R., & Hoppa, R. D. (1993). Growth deficit in survivors and non-survivors: Biological mortality bias in subadult skeletal samples. *American Journal of Physical Anthropology*, *36*(S17), 127–151. https://doi.org/10.1002/ajpa.1330360608

Savu-Krohn, C., Rantitsch, G., Auer, P., Melcher, F., & Graupner, T. (2011). Geochemical fingerprinting of coltan ores by machine learning on uneven datasets. *Natural Resources Research*, *20*(3), 177–191. https://doi.org/10.1007/s11053-011-9142-4

Sayago, A., & Asuero, A. G. (2004). Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances. *Critical Reviews in Analytical Chemistry*, *34*(3–4), 133–146. https://doi.org/10.1080/10408340490888599

Sayago, A., Boccio, M., & Asuero, A. (2004). Fitting straight lines with replicated observations by linear regression: The least squares postulates. *Critical Reviews in Analytical Chemistry*, *34*(1), 39–50. https://doi.org/10.1080/10408340490273744

Saylor, J. E., Sundell, K. E., & Sharman, G. R. (2019). Characterizing sediment sources by non-negative matrix factorization of detrital geochronological data. *Earth and Planetary Science Letters*, *512*, 46–58. https://doi.org/10.1016/j.epsl.2019.01.044

Schmitt, B., Souidi, Z., Duquesnoy, F., & Donzé, F.-V. (2023). From RGB camera to hyperspectral imaging: A breakthrough in Neolithic rock painting analysis. *Heritage Science*, *11*(1), 91. https://doi.org/10.1186/s40494-023-00940-5

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319. https://doi.org/10.1162/089976698300017467

Scholze, H., Stephanowitz, H., Hoffmann, E., & Skole, J. (1994). Element analysis of inhomogeneous samples by laser ablation ICP mass spectrometry. *Fresenius' Journal of Analytical Chemistry*, *350*(4–5), 247–252. https://doi.org/10.1007/BF00322477

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21. https://doi.org/10.1145/3068335

Schuermans, M., Markovsky, I., Wentzell, P. D., & Van Huffel, S. (2005). On the equivalence between total least squares and maximum likelihood PCA. *Analytica Chimica Acta*, *544*(1–2), 254–267. https://doi.org/10.1016/j.aca.2004.12.059

Schuessler, K. (1974). Analysis of ratio variables: Opportunities and pitfalls. *American Journal of Sociology*, *80*(2), 379–396. https://doi.org/10.1086/225805

Scott, E. M., Cook, G. T., & Naysmith, P. (2007). Error and uncertainty in radiocarbon measurements. *Radiocarbon*, *49*(2), 427–440. https://doi.org/10.1017/S0033822200042351

Seccaroni, C., Volante, N., Rosada, A., Ambrosone, L., Bufalo, G., & Avino, P. (2008). Identification of provenance of obsidian samples analyzing elemental composition by INAA. *Journal of Radioanalytical and Nuclear Chemistry*, *278*(2), 277–282. https://doi.org/10.1007/s10967-008-9502-7

Seliya, N., Abdollah Zadeh, A., & Khoshgoftaar, T. M. (2021). A literature review on one-class classification and its potential applications in big data. *Journal of Big Data*, *8*(1), 122. https://doi.org/10.1186/s40537-021-00514-x

Sevrukov, A. B., Bland, J. M., & Kondos, G. T. (2005). Serial electron beam CT measurements of coronary artery calcium: Has your patient's calcium score actually changed? *American Journal of Roentgenology*, *185*(6), 1546–1553. https://doi.org/10.2214/AJR.04.1589

Shahrestani, S., & Carranza, E. J. M. (2024). Effectiveness of LOF, iForest and OCSVM in detecting anomalies in stream sediment geochemical data. *Geochemistry: Exploration, Environment, Analysis*, *24*(3), geochem2024-009. https://doi.org/10.1144/geochem2024-009

Shahrestani, S., Cohen, D. R., & Mokhtari, A. R. (2024). A comparison of PCA and ICA in geochemical pattern recognition of soil data: The case of Cyprus. *Journal of Geochemical Exploration*, *264*, 107539. https://doi.org/10.1016/j.gexplo.2024.107539

Shahrestani, S., & Sanislav, I. (2025). How does dimensionality influence outlier detection effectiveness in multivariate geochemical data? Insights from LOF and IF methods. *Earth Science Informatics*, *18*(1), 27. https://doi.org/10.1007/s12145-024-01611-0

Shamraeva, M. A., Visvikis, T., Zoidis, S., Anthony, I. G. M., & Van Nuffel, S. (2024). The application of a random forest classifier to ToF-SIMS imaging data. *Journal of the American Society for Mass Spectrometry*, *35*(12), 2801–2814. https://doi.org/10.1021/jasms.4c00324

Shan, G., Zhang, H., & Jiang, T. (2020). Correlation coefficients for a study with repeated measures. *Computational and Mathematical Methods in Medicine*, *2020*, 1–11. https://doi.org/10.1155/2020/7398324

Sheikh, R. H., Raghuwanshi, M. M., & Jaiswal, A. N. (2008). Genetic algorithm based clustering: A survey. *2008 First International Conference on Emerging Trends in Engineering and Technology*, 314–319. https://doi.org/10.1109/ICETET.2008.48

Shennan, S. (1988). *Quantifying Archaeology*. Elsevier. https://doi.org/10.1016/C2009-0-21716-8

Shieh, G. (2018). The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement. *BMC Medical Research Methodology*, *18*(1), 45. https://doi.org/10.1186/s12874-018-0505-y

Shoari, N., & Dubé, J.-S. (2016). An investigation of the impact of left-censored soil contamination data on the uncertainty of descriptive statistical parameters. *Environmental Toxicology and Chemistry*, *35*(10), 2623–2631. https://doi.org/10.1002/etc.3420

Shoari, N., & Dubé, J.-S. (2017). Toward improved analysis of concentration data: Embracing nondetects. *Environmental Toxicology and Chemistry*, *37*(3), 643–656. https://doi.org/10.1002/etc.4046

Shoemaker, J. D., Lynch, R. E., Hoffmann, J. W., & Sly, W. S. (1992). Misidentification of propionic acid as ethylene glycol in a patient with methylmalonic acidemia. *Journal of Pediatrics*, *120*(3), 417–421. https://doi.org/10.1016/S0022-3476(05)80909-6

Shoukri, M. M. (2024). Cautionary Remarks when testing agreement between two raters for continuous scale measurements: A tutorial in clinical epidemiology with implementation using R. *Open Journal of Epidemiology*, *14*(01), 56–74. https://doi.org/10.4236/ojepi.2024.141005

Singer, R., & Danzer, K. (1984). Homogenitätsuntersuchungen von Festkörpern mit Hilfe linearer Regressionsmodelle. *Zeitschrift Für Chemie*, *24*(9), 339–341. https://doi.org/10.1002/zfch.19840240918

Singh, A., & Nocerino, J. (2002). Robust estimation of mean and variance using environmental data sets with below detection limit observations. *Chemometrics and Intelligent Laboratory Systems*, *60*(1–2), 69–86. https://doi.org/10.1016/S0169-7439(01)00186-1

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Sircombe, K. N., & Hazelton, M. L. (2004). Comparison of detrital zircon age distributions by kernel functional estimation. *Sedimentary Geology*, *171*(1–4), 91–111. https://doi.org/10.1016/j.sedgeo.2004.05.012

Skala, W. (1979). Some effects of the constant-sum problem in geochemistry. *Chemical Geology*, *27*(1–2), 1–9. https://doi.org/10.1016/0009-2541(79)90099-8

Smets, T., Waelkens, E., & De Moor, B. (2020). Prioritization of m/z values in mass spectrometry imaging profiles obtained using uniform manifold approximation and projection for dimensionality reduction. *Analytical Chemistry*, *92*(7), 5240–5248. https://doi.org/10.1021/acs.analchem.9b05764

Snoep, J. D., Morabia, A., Hernández-Díaz, S., Hernán, M. A., & Vandenbroucke, J. P. (2014). Commentary: A structural approach to Berkson's fallacy and a guide to a history of opinions about it. *International Journal of Epidemiology*, *43*(2), 515–521. https://doi.org/10.1093/ije/dyu026

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, *11*(2), 33–40. https://doi.org/10.2307/1217208

Solheim, W. G. (1960). The use of sherd weights and counts in the handling of archaeological data. *Current Anthropology*, *1*(4), 325–329. https://doi.org/10.1086/200116

Song, Y., Chen, X., Li, Y., Fan, Y., & Collins, A. L. (2022). Quantifying the provenance of dune sediments in the Taklimakan Desert using machine learning, multidimensional scaling and sediment source fingerprinting. *CATENA*, *210*, 105902. https://doi.org/10.1016/j.catena.2021.105902

Srinivas, J., Naveen Kumar, B., & Krishna Reddy, M. (2013). Comparison of ratio estimators of population mean. *International Journal of Mathematical Sciences, Technology and Humanities*, *108*, 1158–1163.

Srivastava, D., Dubey, C. P., Banerji, U. S., & Joshi, K. B. (2024). Geochemical trends in sedimentary environments using PCA approach. *Journal of Earth System Science*, *133*(3), 122. https://doi.org/10.1007/s12040-024-02306-2

Stanimirova, I. (2013). Practical approaches to principal component analysis for simultaneously dealing with missing and censored elements in chemical data. *Analytica Chimica Acta*, *796*, 27–37. https://doi.org/10.1016/j.aca.2013.08.026

Stanley, C. R. (2003). Estimating sampling errors for major and trace elements in geological materials using a propagation of variance approach. *Geochemistry: Exploration, Environment, Analysis*, *3*(2), 169–178. https://doi.org/10.1144/1467-7873/03-008

Stanley, C. R. (2006). Numerical transformation of geochemical data: 2. Stabilizing measurement error to facilitate data interpretation. *Geochemistry: Exploration, Environment, Analysis*, *6*(1), 79–96. https://doi.org/10.1144/1467-7873/05-079

Stanley, C. R. (2019). Closure in compositional data analysis: Is it always such a hazard? *Atlantic Geology*, *55*(1), 204–205. https://doi.org/10.4138/atlgeol.2019.005

Stanley, C. R. (2020). Molar element ratio analysis of lithogeochemical data: A toolbox for use in mineral exploration and mining. *Geochemistry: Exploration, Environment, Analysis*, *20*(2), 233–256. https://doi.org/10.1144/geochem2019-033

Stanley, C. R., & Lawie, D. (2007). Average relative error in geochemical determinations: Clarification, calculation, and a plea for consistency. *Exploration and Mining Geology*, *16*(3–4), 267–275. https://doi.org/10.2113/gsemg.16.3-4.267

Stanley, C. R., O'Driscoll, N., & Ranjan, P. (2010). Determining the magnitude of true analytical error in geochemical analysis. *Geochemistry: Exploration, Environment, Analysis*, *10*(4), 355–364. https://doi.org/10.1144/1467-7873/09-226

Stöckl, D., Dewitte, K., & Thienpont, L. M. (1998). Validity of linear regression in method comparison studies: Is it limited by the statistical model or the quality of the analytical input data? *Clinical Chemistry*, *44*(11), 2340–2346. https://doi.org/10.1093/clinchem/44.11.2340

Strani, L., Cocchi, M., Tanzilli, D., Biancolillo, A., Marini, F., & Vitale, R. (2025). One class classification (class modelling): State of the art and perspectives. *TrAC Trends in Analytical Chemistry*, *183*, 118117. https://doi.org/10.1016/j.trac.2024.118117

Streiner, D. V. L. (2001). Regression toward the mean: Its etiology, diagnosis, and treatment. *Canadian Journal of Psychiatry*, *46*(1), 72–76. https://doi.org/10.1177/070674370104600111

Stroth, L., Otto, R., Daniels, J. T., & Braswell, G. E. (2019). Statistical artifacts: Critical approaches to the analysis of obsidian artifacts by portable X-ray fluorescence. *Journal of Archaeological Science: Reports*, *24*, 738–747. https://doi.org/10.1016/j.jasrep.2019.02.023

Sun, H., Liu, M., Li, L., Yan, L., Zhou, Y., & Feng, X. (2020). A new classification method of ancient Chinese ceramics based on machine learning and component analysis. *Ceramics International*, *46*(6), 8104–8110. https://doi.org/10.1016/j.ceramint.2019.12.037

Sun, J., & Xia, Y. (2024). Pretreating and normalizing metabolomics data for statistical analysis. *Genes & Diseases*, *11*(3), 100979. https://doi.org/10.1016/j.gendis.2023.04.018

Suzuki, Y., Kitayama, I., Harimoto, M., Akiyama, H., & Tsutsumi, T. (2025). Probabilistic estimation of the dietary exposure to Cadmium in Japan in 2019-2021 using two-dimensional Monte Carlo simulation after Bayesian handling of left-censored data. *Food Chemistry Advances*, *7*, 101013. https://doi.org/10.1016/j.focha.2025.101013

Suzuki, Y., Tanaka, N., & Akiyama, H. (2020). Attempt of Bayesian estimation from left-censored data using the Markov Chain Monte Carlo method: Exploring Cr(VI) concentrations in mineral water products. *Food Safety*, *8*(4), 67–89. https://doi.org/10.14252/foodsafetyfscj.D-20-00007

Syvilay, D., Bousquet, B., Chapoulie, R., Orange, M., & Le Bourdonnec, F.-X. (2019). Advanced statistical analysis of LIBS spectra for the sourcing of obsidian samples. *Journal of Analytical Atomic Spectrometry*, *34*(5), 867–873. https://doi.org/10.1039/C8JA00340H

Szpak, P., Metcalfe, J. Z., & Macdonald, R. A. (2017). Best practices for calibrating and reporting stable isotope measurements in archaeology. *Journal of Archaeological Science: Reports*, *13*, 609–616. https://doi.org/10.1016/j.jasrep.2017.05.007

Szymańska, E. (2018). Modern data science for analytical chemical data – A comprehensive review. *Analytica Chimica Acta*, *1028*, 1–10. https://doi.org/10.1016/j.aca.2018.05.038

Taffé, P. (2020). Assessing bias, precision, and agreement in method comparison studies. *Statistical Methods in Medical Research*, *29*(3), 778–796. https://doi.org/10.1177/0962280219844535

Taffé, P., Halfon, P., & Halfon, M. (2020). A new statistical methodology overcame the defects of the Bland–Altman method. *Journal of Clinical Epidemiology*, *124*, 1–7. https://doi.org/10.1016/j.jclinepi.2020.03.018

Taffé, P., Zuppinger, C., Burger, G. M., & Nusslé, S. G. (2022). The Bland-Altman method should not be used when one of the two measurement methods has negligible measurement errors. *PLOS ONE*, *17*(12), e0278915. https://doi.org/10.1371/journal.pone.0278915

Teil, H. (1975). Correspondence factor analysis: An outline of its method. *Journal of the International Association for Mathematical Geology*, *7*(1), 3–12. https://doi.org/10.1007/BF02080630

Tellinghuisen, J. (2020). Least squares methods for treating problems with uncertainty in x and y. *Analytical Chemistry*, *92*(16), 10863–10871. https://doi.org/10.1021/acs.analchem.0c02178

Templ, M., Alfons, A., & Filzmoser, P. (2012). Exploring incomplete data using visualization techniques. *Advances in Data Analysis and Classification*, *6*(1), 29–47. https://doi.org/10.1007/s11634-011-0102-y

Templ, M., Filzmoser, P., & Reimann, C. (2008). Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry*, *23*(8), 2198–2213. https://doi.org/10.1016/j.apgeochem.2008.03.004

Terrill, D. F., Jarochowska, E., Henderson, C. M., Shirley, B., & Bremer, O. (2022). Sr/Ca and Ba/Ca ratios support trophic partitioning within a Silurian conodont community from Gotland, Sweden. *Paleobiology*, *48*(4), 601–621. https://doi.org/10.1017/pab.2022.9

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Debono, V. B., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., & Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. *BMC Medical Research Methodology*, *13*(1), 92. https://doi.org/10.1186/1471-2288-13-92

Theodorsson, E. (2017). Uncertainty in measurement and total error. *Clinics in Laboratory Medicine*, *37*(1), 15–34. https://doi.org/10.1016/j.cll.2016.09.002

Thompson, M., & Howarth, R. J. (1976). Duplicate analysis in geochemical practice. Part I. Theoretical approach and estimation of analytical reproducibility. *The Analyst*, *101*(1206), 690. https://doi.org/10.1039/an9760100690

Thompson, W. C., & Scurich, N. (2018). When does absence of evidence constitute evidence of absence? *Forensic Science International*, *291*, e18–e19. https://doi.org/10.1016/j.forsciint.2018.08.040

Thorpe, O. W., Warren, S. E., & Nandris, J. G. (1984). The distribution and provenance of archaeological obsidian in Central and Eastern Europe. *Journal of Archaeological Science*, *11*(3), 183–212. https://doi.org/10.1016/0305-4403(84)90001-3

Tierney, N., & Cook, D. (2023). Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *Journal of Statistical Software*, *105*(7). https://doi.org/10.18637/jss.v105.i07

Timmerman, J. G., Beinat, E., Termeer, K., & Cofino, W. (2010). Analyzing the Data-Rich-But-Information-Poor syndrome in Dutch water management in historical perspective. *Environmental Management*, *45*(5), 1231–1242. https://doi.org/10.1007/s00267-010-9459-5

Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *61*(3), 611–622. https://doi.org/10.1111/1467-9868.00196

Tripathi, V. S. (1979). Factor analysis in geochemical exploration. *Journal of Geochemical Exploration*, *11*(3), 263–275. https://doi.org/10.1016/0375-6742(79)90004-9

Troiano, M., Nobile, E., Grignaffini, F., Mangini, F., Mastrogiuseppe, M., Conati Barbaro, C., & Frezza, F. (2024). A comparative analysis of machine learning algorithms for identifying cultural and technological groups in archaeological datasets through clustering analysis of homogeneous data. *Electronics*, *13*(14), 2752. https://doi.org/10.3390/electronics13142752

Trygg, J., Holmes, E., & Lundstedt, T. (2007). Chemometrics in metabonomics. *Journal of Proteome Research*, *6*(2), 469–479. https://doi.org/10.1021/pr060594q

Tseng, C. Y., Salguero, J. A., Breidenbach, J. D., Solomon, E., Sanders, C. K., Harvey, T., Thornhill, M. G., Palmisano, S. J., Sasiene, Z. J., Blackwell, B. R., McBride, E. M., Luchini, K. A., LeBrun, E. S., Alvarez, M., Mach, P. M., Rivera, E. S., & Glaros, T. G. (2025). Evaluation of normalization strategies for mass spectrometry-based multi-omics datasets. *Metabolomics*, *21*(4), 98. https://doi.org/10.1007/s11306-025-02297-1

Tu, Y., Bælum, V., & Gilthorpe, M. S. (2005). The problem of analysing the relationship between change and initial value in oral health research. *European Journal of Oral Sciences*, *113*(4), 271–278. https://doi.org/10.1111/j.1600-0722.2005.00228.x

Tu, Y.-K., Clerehugh, V., & Gilthorpe, M. S. (2004). Ratio variables in regression analysis can give rise to spurious results: Illustration from two studies in periodontology. *Journal of Dentistry*, *32*(2), 143–151. https://doi.org/10.1016/j.jdent.2003.09.004

Tu, Y.-K., Gunnell, D., & Gilthorpe, M. S. (2008). Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon – the reversal paradox. *Emerging Themes in Epidemiology*, *5*(1), 2. https://doi.org/10.1186/1742-7622-5-2

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110. https://doi.org/10.1037/h0031322

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Twomey, P. J. (2006). How to use difference plots in quantitative method comparison studies. *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, *43*(2), 124–129. https://doi.org/10.1258/000456306776021616

Üçer, B. (2011). Analyzing dependence structure of thyroid hormones: A copula approach. *Turkish Journal of Medical Sciences*. https://doi.org/10.3906/sag-1007-959

Underhill, L. G., & Peisach, M. (1985). Correspondence analysis and its application in multielemental trace analysis. *Journal of Trace and Microprobe Techniques*, *3*(1 and 2), 41–65.

Valenchon, F. (1982). The use of correspondence analysis in geochemistry. *Journal of the International Association for Mathematical Geology*, *14*(4), 331–342. https://doi.org/10.1007/BF01032594

Valletta, E., Kučera, L., Prokeš, L., Amato, F., Pivetta, T., Hampl, A., Havel, J., & Vaňhara, P. (2016). Multivariate calibration approach for quantitative determination of cell-line cross contamination by intact cell mass spectrometry and artificial neural networks. *PLOS ONE*, *11*(1), e0147414. https://doi.org/10.1371/journal.pone.0147414

Van Der Weijden, C. H. (2002). Pitfalls of normalization of marine geochemical data using a common divisor. *Marine Geology*, *184*(3–4), 167–187. https://doi.org/10.1016/S0025-3227(01)00297-3

Van Es, A., Wiarda, W., Hordijk, M., Alberink, I., & Vergeer, P. (2017). Implementation and assessment of a likelihood ratio approach for the evaluation of LA-ICP-MS evidence in forensic glass analysis. *Science & Justice*, *57*(3), 181–192. https://doi.org/10.1016/j.scijus.2017.03.002

Van Kempen, G. M. P., & Van Vliet, L. J. (2000). Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry*, *39*(4), 300–305. https://doi.org/10.1002/(SICI)1097-0320(20000401)39:4%253C300::AID-CYTO8%253E3.0.CO;2-O

Varmuza, K., Engrand, C., Filzmoser, P., Hilchenbach, M., Kissel, J., Krüger, H., Silén, J., & Trieloff, M. (2011). Random projection for dimensionality reduction—Applied to time-of-flight secondary ion mass spectrometry data. *Analytica Chimica Acta*, *705*(1–2), 48–55. https://doi.org/10.1016/j.aca.2011.03.031

Verbeeck, N., Caprioli, R. M., & Van De Plas, R. (2020). Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. *Mass Spectrometry Reviews*, *39*(3), 245–291. https://doi.org/10.1002/mas.21602

Vermeesch, P., Lipp, A. G., Hatzenbühler, D., Caracciolo, L., & Chew, D. (2023). Multidimensional Scaling of varietal data in sedimentary provenance analysis. *Journal of Geophysical Research: Earth Surface*, *128*(3), e2022JF006992. https://doi.org/10.1029/2022JF006992

Vermeesch, P., Resentini, A., & Garzanti, E. (2016). An R package for statistical provenance analysis. *Sedimentary Geology*, *336*, 14–25. https://doi.org/10.1016/j.sedgeo.2016.01.009

Vierra, R. K., & Carlson, D. L. (1981). Factor analysis, random data, and patterned results. *American Antiquity*, *46*(2), 272–283. https://doi.org/10.2307/280208

Vignon, M. (2015). Extracting environmental histories from sclerochronological structures—Recursive partitioning as a mean to explore multi-elemental composition of fish otolith. *Ecological Informatics*, *30*, 159–169. https://doi.org/10.1016/j.ecoinf.2015.10.002

Vock, M. (2016). Intervals for the assessment of measurement agreement: Similarities, differences, and consequences of incorrect interpretations. *Biometrical Journal*, *58*(3), 489–501. https://doi.org/10.1002/bimj.201400234

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly*, *74*(2), 223–243. https://doi.org/10.1093/poq/nfq007

Walczak, B., & Massart, D. L. (2001a). Dealing with missing data. *Chemometrics and Intelligent Laboratory Systems*, *58*(1), 15–27. https://doi.org/10.1016/S0169-7439(01)00131-9

Walczak, B., & Massart, D. L. (2001b). Dealing with missing data: Part II. *Chemometrics and Intelligent Laboratory Systems*, *58*(1), 29–42. https://doi.org/10.1016/S0169-7439(01)00132-0

Walker, P. L., Johnson, J. R., & Lambert, P. M. (1988). Age and sex biases in the preservation of human skeletal remains. *American Journal of Physical Anthropology*, *76*(2), 183–188. https://doi.org/10.1002/ajpa.1330760206

Wallach, E. (2019). Inference from absence: The case of archaeology. *Palgrave Communications*, *5*(1), 94. https://doi.org/10.1057/s41599-019-0307-9

Walsh, C., Stallard-Olivera, E., & Fierer, N. (2024). Nine (not so simple) steps: A practical guide to using machine learning in microbial ecology. *mBio*, *15*(2), e02050-23. https://doi.org/10.1128/mbio.02050-23

Wang, C., Zhao, K.-D., Chen, J., & Ma, X. (2022). Examining fingerprint trace elements in cassiterite: Implications for primary tin deposit exploration. *Ore Geology Reviews*, *149*, 105082. https://doi.org/10.1016/j.oregeorev.2022.105082

Wang, W.-L., Castro, L. M., Lachos, V. H., & Lin, T.-I. (2019). Model-based clustering of censored data via mixtures of factor analyzers. *Computational Statistics & Data Analysis*, *140*, 104–121. https://doi.org/10.1016/j.csda.2019.06.001

Ward, R. C., Loftis, J. C., & McBride, G. B. (1986). The "data-rich but information-poor" syndrome in water quality monitoring. *Environmental Management*, *10*(3), 291–297. https://doi.org/10.1007/BF01867251

Watkins, P., Stockham, K., Stewart, S., & Gardner, G. (2021). Contemporary chemical lean determination used in the Australian meat processing industry: A method comparison. *Meat Science*, *171*, 108289. https://doi.org/10.1016/j.meatsci.2020.108289

Wentzell, P. D. (2013). Measurement errors in multivariate chemical data. *Journal of the Brazilian Chemical Society*. https://doi.org/10.5935/0103-5053.20130293

Wentzell, P. D., & Lohnes, M. T. (1999). Maximum likelihood principal component analysis with correlated measurement errors: Theoretical and practical considerations. *Chemometrics and Intelligent Laboratory Systems*, *45*(1–2), 65–85. https://doi.org/10.1016/S0169-7439(98)00090-2

Wentzell, P. D., Wicks, C. C., Braga, J. W. B., Soares, L. F., Pastore, T. C. M., Coradin, V. T. R., & Davrieux, F. (2018). Implications of measurement error structure on the visualization of multivariate chemical data: Hazards and alternatives. *Canadian Journal of Chemistry*, *96*(7), 738–748. https://doi.org/10.1139/cjc-2017-0730

Westgard, J. O., & Hunt, M. R. (1973). Use and interpretation of common statistical tests in method-comparison studies. *Clinical Chemistry*, *19*(1), 49–57. https://doi.org/10.1093/clinchem/19.1.49

Westreich, D. (2012). Berkson's bias, selection bias, and missing data. *Epidemiology*, *23*(1), 159–164. https://doi.org/10.1097/EDE.0b013e31823b6296

Williams, F. L., Belcher, R. L., & Armelagos, G. J. (2005). Forensic misclassification of ancient Nubian crania: Implications for assumptions about human variation¨. *Current Anthropology*, *46*(2), 340–346. https://doi.org/10.1086/428792

Williams, M. R., Lamont, B. B., & He, T. (2022). Dealing with 'the spectre of 'spurious' correlations': Hazards in comparing ratios and other derived variables with a randomization test to determine if a biological interpretation is justified. *Oikos*, *2022*(3), e08575. https://doi.org/10.1111/oik.08575

Williamson, J. H. (1968). Least-squares fitting of a straight line. *Canadian Journal of Physics*, *46*(16), 1845–1847. https://doi.org/10.1139/p68-523

Wilson, A. L. (1978). Elemental analysis of pottery in the study of its provenance: A review. *Journal of Archaeological Science*, *5*(3), 219–236. https://doi.org/10.1016/0305-4403(78)90041-9

Winzar, H. (2015). The ecological fallacy: How to spot one and tips on how to use one to your advantage. *Australasian Marketing Journal*, *23*(1), 86–92. https://doi.org/10.1016/j.ausmj.2014.12.002

Wissler, A., Blevins, K. E., & Buikstra, J. E. (2022). Missing data in bioarchaeology II: A test of ordinal and continuous data imputation. *American Journal of Biological Anthropology*, *179*(3), 349–364. https://doi.org/10.1002/ajpa.24614

Wludyka, P. S., & Nelson, P. R. (1997). Analysis of Means type tests for variances using subsampling and jackknifing. *American Journal of Mathematical and Management Sciences*, *17*(1–2), 31–60. https://doi.org/10.1080/01966324.1997.10737429

Wood, J. R., & Greenacre, M. (2021). Making the most of expert knowledge to analyse archaeological data: A case study on Parthian and Sasanian glazed pottery. *Archaeological and Anthropological Sciences*, *13*(7), 110. https://doi.org/10.1007/s12520-021-01341-0

Wu, C., & Yu, J. Z. (2018). Evaluation of linear regression techniques for atmospheric applications: The importance of appropriate weighting. *Atmospheric Measurement Techniques*, *11*(2), 1233–1250. https://doi.org/10.5194/amt-11-1233-2018

Wu, R., Chen, J., Zhao, J., Chen, J., & Chen, S. (2020). Identifying geochemical anomalies associated with gold mineralization using factor analysis and spectrum–area multifractal model in Laowan district, Qinling-Dabie Metallogenic Belt, Central China. *Minerals*, *10*(3), 229. https://doi.org/10.3390/min10030229

Wu, W., Massart, D. L., & De Jong, S. (1997). The kernel PCA algorithms for wide data. Part I: Theory and algorithms. *Chemometrics and Intelligent Laboratory Systems*, *36*(2), 165–172. https://doi.org/10.1016/S0169-7439(97)00010-5

Yang, J., & Cheng, Q. (2015a). A comparative study of independent component analysis with principal component analysis in geological objects identification, Part I: Simulations. *Journal of Geochemical Exploration*, *149*, 127–135. https://doi.org/10.1016/j.gexplo.2014.11.013

Yang, J., & Cheng, Q. (2015b). A comparative study of independent component analysis with principal component analysis in geological objects identification. Part II: A case study of Pinghe District, Fujian, China. *Journal of Geochemical Exploration*, *149*, 136–146. https://doi.org/10.1016/j.gexplo.2014.11.014

Yigit, S., & Mendes, M. (2017). ANOM technique for evaluating practical significance of observed difference among treatment groups. *International Journal of Agricultural Science Research*, *6*(1), 001–007.

York, D. (1966). Least-squares fitting of a straight line. *Canadian Journal of Physics*, *44*(5), 1079–1086. https://doi.org/10.1139/p66-090

York, D. (1968). Least squares fitting of a straight line with correlated errors. *Earth and Planetary Science Letters*, *5*, 320–324. https://doi.org/10.1016/S0012-821X(68)80059-7

Yrarrazaval, S., Cartajena, I., Borrero, L., & Salazar, D. (2024). Identifying non-anthropogenic accumulation in zooarchaeological assemblages using naive Bayesian classifier: A trace-oriented actualistic taphonomic approach in the hyperarid coasts of the Atacama desert. *Quaternary Science Advances*, *13*, 100143. https://doi.org/10.1016/j.qsa.2023.100143

Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T. C., & Xu, X. (2022). Beyond t test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron*, *110*(1), 21–35. https://doi.org/10.1016/j.neuron.2021.10.030

Yuan, Y., Cave, M., & Zhang, C. (2018). Using local Moran's I to identify contamination hotspots of rare earth elements in urban soils of London. *Applied Geochemistry*, *88*, 167–178. https://doi.org/10.1016/j.apgeochem.2017.07.011

Zaki, R., Bulgiba, A., Ismail, R., & Ismail, N. A. (2012). Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. *PLoS ONE*, *7*(5), e37908. https://doi.org/10.1371/journal.pone.0037908

Zauner, B., Petrinec, B., Bituh, T., Ceci, S., Volarić, N., Včev, A., Vukoja, A., & Babić, D. (2025). Theory of random measurement errors: Concepts, uses, and misuses. *Archives of Industrial Hygiene and Toxicology*, *76*(2), 79–86. https://doi.org/10.2478/aiht-2025-76-3977

Zec, S., Soriani, N., Comoretto, R., & Baldi, I. (2017). High agreement and high prevalence: The paradox of Cohen's kappa. *Open Nursing Journal*, *11*(1), 211–218. https://doi.org/10.2174/1874434601711010211

Zekri, H., Mokhtari, A. R., & Cohen, D. R. (2019). Geochemical pattern recognition through matrix decomposition. *Ore Geology Reviews*, *104*, 670–685. https://doi.org/10.1016/j.oregeorev.2018.11.026

Zhang, C., Zhou, L., Liu, F., Huang, J., & Peng, J. (2023). Application of deep learning in laser-induced breakdown spectroscopy: A review. *Artificial Intelligence Review*, *56*(S2), 2789–2823. https://doi.org/10.1007/s10462-023-10590-5

Zhang, Q., Liu, Y., & Fang, H. (2024). Manifold learning-based UMAP method for geochemical anomaly identification. *Geochemistry*, *84*(4), 126157. https://doi.org/10.1016/j.chemer.2024.126157

Zhang, S., Carranza, E. J. M., Fu, C., Zhang, W., & Qin, X. (2024). Interpretable machine learning for geochemical anomaly delineation in the Yuanbo Nang district, Gansu province, China. *Minerals*, *14*(5), 500. https://doi.org/10.3390/min14050500

Zhang, S., Carranza, E. J. M., Xiao, K., Chen, Z., Li, N., Wei, H., Xiang, J., Sun, L., & Xu, Y. (2021). Geochemically constrained prospectivity mapping aided by unsupervised cluster analysis. *Natural Resources Research*, *30*(3), 1955–1975. https://doi.org/10.1007/s11053-021-09865-x

Zhang, S., Xiao, K., Carranza, E. J. M., Yang, F., & Zhao, Z. (2019). Integration of auto-encoder network with density-based spatial clustering for geochemical anomaly detection for mineral exploration. *Computers & Geosciences*, *130*, 43–56. https://doi.org/10.1016/j.cageo.2019.05.011

Zhang, W., Claesen, M., Moerman, T., Groseclose, M. R., Waelkens, E., De Moor, B., & Verbeeck, N. (2021). Spatially aware clustering of ion images in mass spectrometry imaging data using deep learning. *Analytical and Bioanalytical Chemistry*, *413*(10), 2803–2819. https://doi.org/10.1007/s00216-021-03179-w

Zhang, Z. (2015). Missing data exploration: Highlighting graphical presentation of missing pattern. *Annals of Translational Medicine*, *3*(22), 356.

Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, *4*(1), 6. https://doi.org/10.3978/j.issn.2305-5839.2015.12.38

Zhang, Z., Lennox, W. C., & Panu, U. S. (2004). Effect of percent non-detects on estimation bias in censored distributions. *Journal of Hydrology*, *297*(1–4), 74–94. https://doi.org/10.1016/j.jhydrol.2004.04.017

Zhao, Y., & Frey, H. C. (2006). Uncertainty for data with non-detects: Air toxic emissions from combustion. *Human and Ecological Risk Assessment: An International Journal*, *12*(6), 1171–1191. https://doi.org/10.1080/10807030600977178

Zuanetti, D. A., Müller, P., Zhu, Y., Yang, S., & Ji, Y. (2019). Bayesian nonparametric clustering for large data sets. *Statistics and Computing*, *29*(2), 203–215. https://doi.org/10.1007/s11222-018-9803-9

Zurer, P. (1991). Mother cleared of ethylene glycol poisoning. *Chemical & Engineering News Archive*, *69*(39), 7–8. https://doi.org/10.1021/cen-v069n039.p007a

Appendix A List of Abbreviations

ACF	Autocorrelation function
AI	Artificial intelligence
AIC	Akaike information criterion
alr	Additive logratio
ANCOVA	Analysis of covariance
ANN	Artificial neural networks
ANOM	Analysis of means
ANOVA	Analysis of variance
ATR	Attenuated total reflectance
ATS	Akritas-Theil-Sen
AUC	Area under curve
BA	Bland-Altman
BIC	Bayes information criterion
BLS	Bivariate least squares
BP	Before present
CA	Correspondence analysis
CART	Classification and regression trees
CCC	Concordance correlation coefficient
clr	Centered logratio
CNN	Convolutional neural networks
CoDa	Compositional data
CRM	Certified reference material
CUSUM	Cumulative sum
DA	Discriminant analysis
DBSCAN	Density-cased spatial clustering of applications with noise
DR	Deming regression
DRC	Democratic Republic of Congo
ED	Energy dispersion
EDA	Exploratory data analysis
EM	Expectation maximisation
FA	Factor analysis
FDA	Flexible discriminant analysis
FFT	Fast Fourier transformation
FMI	Fraction of missing information
FTIR	Fourier transform infrared spectrometry
GA	Genetic algorithm
GC	Gas chromatography
glmnet	Generalised linear models with Elastic Net regularisation
GLS	Generalised least squares
GMM	Gaussian mixture model
ICA	Independent component analysis

IDW	Inverse distance weighting
ilr	Isometric logratio
k-NN	k-nearest neighbour
KDE	Kernel density estimator
KM	Kaplan-Meier estimation
kPCA	Kernel principal component analysis
ICP	Inductively coupled plasma
IF	Isolation forest
LA	Laser ablation
LDA	Linear discriminant analysis
LDI	Laser desorption ionisation
LIBS	Laser induced breakdown spectroscopy
LISA	Local indicators of spatial association
LLM	Large language model
LoA	Limit of acceptance
LOD	Limit of detection
LOF	Local outlier factor
LOO	Leave-one-out
LR	Logistic regression
LVQ	Learning vector quantisation
MALDI	Matrix assisted laser desorption ionisation
MAR	Missing at random
MCAR	Missing completely at random
MCR-ALS	Multivariate curve resolution - alternating least squares
MDA	Gaussian mixture discriminant analysis
MDS	Multidimensional scaling
MLE	Maximum likelihood estimation
MNAR	Missing not at random
MS	Mass spectrometry
NAA	Neutron activation analysis
NB	Naive Bayes
NMF	Non-negative matrix factorization
OCSVM	One-class support vector machine
OES	Optical emission spectrometry
OLS	Ordinary least squares
OPTICS	Ordering points to identify the clustering structure
PACF	Partial autocorrelation function
PBR	Passing – Bablok regression
PCA	Principal component analysis
PCM	Possibilistic c-means
PLS	Partial least squares
pwlr	Pairwise logratio
pXRF	Portable X-ray fluorescence

QDA	Quadratic discriminant analysis
RF	Random forest
ROC	Receiver operating characteristic
ROS	Regression on order statistics
SD	Standard deviation
SIMCA	Soft independent modelling of class analogy
SOM	Self-organising maps
SPCA	Sparse principal component analysis
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor embedding
TLS	Total least squares
TOF	Time-of-flight
TV	Total variance
UMAP	Uniform manifold approximation and projection
XRF	X-ray fluorescence

Appendix A. List of Author's Publications without Annotated Papers

(Listed in WOS and Scopus)

Jandera, P.; Komers, D.; Anděl, L.; Prokeš, L. Fitting competitive adsorption isotherms to the distribution data in normal phase systems with binary mobile phases. *Journal of Chromatography* A, 1999, 831 (2), 131-148. DOI:10.1016/S0021-9673(98)00874-7
Document Type: Article, IF = 2.5; JCR Category + Category Quartile: N/A; AIS = N/A

Janoušková, E.; Krbůšková, M., Řehůřková, I.; Klímová, M.; Prokeš, L.; Ruprich, J. Determination of chlordane in foods by gas chromatography. *Food Chemistry*, 2005, 93 (1), 161-169. DOI:10.1016/j.foodchem.2004.11.011
Document Type: Article, IF = 1.8; JCR Category + Category Quartile: CHEMISTRY, APPLIED Q1 + FOOD SCIENCE & TECHNOLOGY Q1; AIS = N/A

Prokeš, L.; Hložek, M. Identification of some adhesives and wood pyrolysis products of archaeological origin by direct inlet mass spectrometry. *Chemia Analityczna*, 2007, 52 (4), 701-714.
Document Type: Article, IF = 0.5; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q4; AIS = 0.124

Weidlich, T.; Prokeš, L., Růžička, A., Padělková, Z. Condensation of aromatic aldehydes with N,N-dimethylacetamide in presence of dialkyl carbonates as dehydrating agents. *Monatshefte für Chemie*, 2010, 141 (2), 205-211. DOI:10.1007/s00706-009-0248-x
Document Type: Article, IF = 1.4; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q2; AIS = 0.305

Galiová, M.; Kaiser, J.; Fortes, F.J.; Novotný, K.; Malina, R.; Prokeš, L.; Hrdlička, A.; Vaculovič, T.; Nývltová Fišáková, M.; Svoboda, J.; Kanický, V.; Laserna, J.J. Multielemental analysis of prehistoric animal teeth by laser-induced breakdown spectroscopy and laser ablation inductively coupled plasma mass spectrometry. *Applied Optics*, 2010, 49 (13), C191-C199. DOI:10.1364/AO.49.00C191
Document Type: Article, IF = 1.7; JCR Category + Category Quartile: OPTICS Q2; AIS = 0.496

Hrdlička, A.; Prokeš, L.; Staňková, A.; Novotný, K.; Vitešníková, A.; Kanický, V.; Otruba, V.; Kaiser, J.; Novotný, J.; Malina, R.; Páleníková, K. Development of a remote laser-induced breakdown spectroscopy system for investigation of calcified tissue samples. *Applied Optics*, 2010, 49 (13), C16-C20. DOI:10.1364/AO.49.000C16
Document Type: Article, IF = 1.7; JCR Category + Category Quartile: OPTICS Q2; AIS = 0.496

Weidlich, T.; Krejčová, A.; Prokeš, L. Study of dehalogenation of halogenoanilines using Raney Al-Ni alloy in aqueous medium at room temperature. *Monatshefte für Chemie*, 2010, 141 (9), 1015-1020. DOI:10.1007/s00706-010-0362-9
Document Type: Article, IF = 1.4; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q2; AIS = 0.305

Weidlich, T.; Prokeš, L. Facile dehalogenation of halogenated anilines and their derivatives using Al-Ni alloy in alkaline aqueous solution. *Central European Journal of Chemistry*, 2011, 9 (4), 590-597. DOI:10.2478/s11532-011-0033-7
Document Type: Article, IF = 1.1; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.258

Weidlich, T.; Krejčová, A.; Prokeš, L. Hydrodebromination of 2,4,6-tribromophenol in aqueous solution using Devarda's alloy. *Monatshefte für Chemie*, 2013, 144 (2), 155-162. DOI:10.1007/s00706-012-0870-x
Document Type: Article, IF = 1.4; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.268

Vašinová Galiová, M.; Nývltová Fišáková, M.; Kynický, J.; Prokeš, L.; Neff, H.; Mason, A.Z.; Gadas, P.; Košler, J.; Kanický, V. Elemental mapping in fossil tooth root section of Ursus arctos by laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). *Talanta*, 2013, 105, 235-243. DOI:10.1016/j.talanta.2012.12.037
Document Type: Article, IF = 3.5; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q1; AIS = 0.776

Hrdlička, A.; Prokeš, L.; Vasinová Galiová, M.; Novotný, K.; Vitešníková, A.; Helešicová, T.; Kanický, V. Provenance study of volcanic glass using 266-1064 nm orthogonal double pulse laser induced breakdown spectroscopy. *Chemical Papers*, 2013, 67 (5), 546-555. DOI:10.2478/s11696-013-0332-x
Document Type: Article, IF = 1.2; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.194

Weidlich, T.; Prokeš, L.; Pospíšilová, D. Debromination of 2,4,6-tribromophenol coupled with biodegradation. *Central European Journal of Chemistry*, 2013, 11 (6), 979-987. DOI:10.2478/s11532-013-0231-6
Document Type: Article, IF = 1.3; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q3; AIS = 0.263

Švihľová, K.; Prokeš, L.; Skácelová, D.; Peña-Méndez, E.M.; Havel, J. Laser ablation synthesis of new gold tellurides using tellurium and nanogold as precursors. Laser desorption ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2013, 27 (14), 1600-1606. DOI:10.1002/rcm.6613
Document Type: Article, IF = 2.6; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q2 + CHEMISTRY, ANALYTICAL Q2 + SPECTROSCOPY Q2; AIS = 0.756

Prokeš, L.; Peña-Méndez, E.M.; Conde, J.E.; Panyala, N.R.; Alberti, M.; Havel, J. Laser ablation synthesis of new gold arsenides using nano-gold and arsenic as precursors. Laser desorption ionisation time-of-flight mass spectrometry and spectrophotometry. *Rapid Communications in Mass Spectrometry*, 2014, 28 (6), 577-586. DOI:10.1002/rcm.6815
Document Type: Article, IF = 2.3; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q3 + CHEMISTRY, ANALYTICAL Q2 + SPECTROSCOPY Q2; AIS = 0.701

Vašinová Galiová, M.; Čopjaková, R.; Škoda, R.; Štěpánková, K.; Vaňková, M.; Kuta, J.; Prokeš, L.; Kynický, J.; Kanický, V. 2D elemental mapping of sections of human kidney stones using laser ablation inductively-coupled plasma-mass spectrometry: Possibilities and limitations. *Spectrochimica Acta Part B - Atomic Spectroscopy*, 2014, 100, 105-115. DOI:10.1016/j.sab.2014.08.024
Document Type: Article, IF = 3.2; JCR Category + Category Quartile: SPECTROSCOPY Q1; AIS = 0.695

Vašinová Galiová, M.; Štepánková, K.; Čopjaková, R.; Kuta, J.; Prokeš, L.; Kynický, J., Kanický, V. Preparation and testing of phosphate, oxalate and uric acid matrix-matched standards for accurate quantification of 2D elemental distribution in kidney stone sections using 213 nm nanosecond laser ablation inductively coupled plasma mass spectrometry. *Journal of Analytical Atomic Spectrometry*, 2015, 30 (6), 1356-1368. DOI:10.1039/c4ja00347k
Document Type: Article, IF = 3.2; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q1 + SPECTROSCOPY Q1; AIS = 0.734

Šútorová, K.; Hawlová, P.; Prokeš, L.; Němec, P.; Boidin, R.; Havel, J. Laser desorption ionization time-of-flight mass spectrometry of Ge-As-Te chalcogenides. *Rapid Communications in Mass Spectrometry*, 2015, 29 (5), 408-414. DOI:10.1002/rcm.7120
Document Type: Article, IF = 2.2; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q3 + CHEMISTRY, ANALYTICAL Q2 + SPECTROSCOPY Q2; AIS = 0.659

Štepánová, V.; Prokeš, L.; Slavíček, P.; Alberti, M.; Havel, J. Laser ablation generation of clusters from As-Te mixtures, As-Te glass nano-layers and from Au-As-Te nano-composites. Quadrupole ion trap time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2015, 29 (11), 1000-1008. DOI:10.1002/rcm.7193
Document Type: Article, IF = 2.2; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q3 + CHEMISTRY, ANALYTICAL Q2 + SPECTROSCOPY Q2; AIS = 0.659

Šútorová, K.; Prokeš, L.; Nazabal, V.; Bouška, M.; Havel, J.; Němec, P. Laser desorption ionisation time-of-flight mass spectrometry of chalcogenide glasses from $(GeSe_2)_{100-x}(Sb_2Se_3)_x$ system. *Journal of the American Ceramic Society*, 2015, 98 (12), 4107-4110. DOI:10.1111/jace.13857
Document Type: Article, IF = 2.8; JCR Category + Category Quartile: MATERIALS SCIENCE, CERAMICS Q1; AIS = 0.663

Burgert, P.; Přichystal, A.; Prokeš, L.; Petřík, J.; Hušková, S. The origin of obsidian in prehistoric Bohemia. *Archeologické rozhledy*, 2016, 68 (2), 224-234. https://archeologickerozhledy.cz/index.php/ar/article/view/154/146
Document Type: Article, IF = N/A; JCR Category + Category Quartile: N/A; AIS = N/A

Prokeš, L.; Kubáček, P.; Peña-Méndez, E.M.; Amato, F.; Conde, J.E.; Alberti, M.; Havel, J. Laser ablation synthesis of gold selenides by using a mass spectrometer as a synthesizer: Laser desorption ionization time of flight mass spectrometry. *Chemistry - A European Journal*, 2016, 22 (32), 11261-11268. DOI:10.1002/chem.201600160
Document Type: Article, IF = 5.3; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q1; AIS = 1.337

Šútorová, K.; Prokeš, L.; Nazabal, V.; Baudet, E.; Havel, J.; Němec, P. Laser desorption ionization time-of-flight mass spectrometry of glasses and amorphous films from Ge-As-Se system. *Journal of the American Ceramic Society*, 2016, 99 (11), 3594-3599. DOI:10.1111/jace.14366
Document Type: Article, IF = 2.8; JCR Category + Category Quartile: MATERIALS SCIENCE, CERAMICS Q1; AIS = 0.663

Mawale, R.M.; Ausekar, M.V.; Prokeš, L.; Nazabal, V.; Baudet, E.; Halenkovič, T.; Bouška, M.; Alberti, M.; Němec, P.; Havel, J. Laser desorption ionization of As_2Ch_3 (Ch = S, Se, and Te) chalcogenides using quadrupole ion trap time-of-flight mass spectrometry: A comparative study. *Journal of the American Society for Mass Spectrometry*, 2017, 28 (12), 2569-2579. DOI:10.1007/s13361-017-1785-x
Document Type: Article, IF = 2.9; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q2 + BIOCHEMISTRY & MOLECULAR BIOLOGY Q3; AIS = 0.791

Kubáček, P.; Prokeš, L.; Pamreddy, A.; Peña-Méndez, E.M.; Conde, J.E.; Alberti, M.; Havel, J. Laser ablation synthesis of arsenic-phosphide As_mP_n clusters from As-P mixtures. Laser desorption ionisation with quadrupole ion trap time-of-flight mass spectrometry: The mass spectrometer as a synthesizer. *Rapid Communications in Mass Spectrometry*, 2018, 32 (10), 789-800. DOI:10.1002/rcm.8106
Document Type: Article, IF = 2.0; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q3 + CHEMISTRY, ANALYTICAL Q3 + SPECTROSCOPY Q2; AIS = 0.521

Petřík, J.; Sosna, D.; Prokeš, L.; Štefanisko, D.; Galeta, P. Shape matters: Assessing regional variation of Bell Beaker projectile points in Central Europe using geometric morphometrics. *Archaeological and Anthropological Sciences*, 2018, 10 (4), 893-904. DOI:10.1007/s12520-016-0423-z
Document Type: Article, IF = 2.0; JCR Category + Category Quartile: ANTHROPOLOGY Q1 + GEOSCIENCES, MULTIDISCIPLINARY Q3; AIS = 0.647

Huang, F.; Prokeš, L.; Havel, J. Laser ablation generation of antimony selenide clusters: Laser desorption ionization (LDI) quadrupole ion trap time of flight mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2019, 30 (4), 634-638. DOI:10.1007/s13361-018-2119-3
Document Type: Article, IF = 3.3; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q2 + BIOCHEMISTRY & MOLECULAR BIOLOGY Q2; AIS = 0.757

Mawale, R.; Halenkovič, T.; Bouška, M.; Gutwirth, J. Nazabal, V.; Takáts, V.; Csík, A.; Havel, J.; Prokeš, L.; Němec, P. Laser desorption ionization time-of-flight mass spectrometry of Ge_xSe_{1-x} chalcogenide glasses, their thin films, and Ge:Se mixtures. *Journal of Non-Crystalline Solids*, 2019, 509, 65-73. DOI:10.1016/j.jnoncrysol.2018.12.020
Document Type: Article, IF = 2.9; JCR Category + Category Quartile: MATERIALS SCIENCE, CERAMICS Q1 + MATERIALS SCIENCE, MULTIDISCIPLINARY Q2; AIS = 0.440

Pečinka, L.; Prokeš, L.; Havel, J. Gallium selenide clusters generated via laser desorption ionisation quadrupole ion trap time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2019, 33 (7), 719-726. DOI:10.1002/rcm.8403
Document Type: Article, IF = 2.3; JCR Category + Category Quartile: BIOCHEMICAL RESEARCH METHODS Q3 + CHEMISTRY, ANALYTICAL Q2 + SPECTROSCOPY Q2; AIS = 0.565

Mawale, R.; Halenkovič, T.; Bouška, M.; Gutwirth, J.; Nazabal, V.; Bora, P.L.; Pečinka, L.; Prokeš, L.; Havel, J.; Němec, P. Mass spectrometric investigation of amorphous Ga-Sb-Se thin films. *Scientific Reports*, 2019, 9, 10213. DOI:10.1038/s41598-019-46767-8
Document Type: Article, IF = 4.0; JCR Category + Category Quartile: MULTIDISCIPLINARY SCIENCES Q1; AIS = 1.261

Dvořáková, H.; Čech, J.; Stupavská, M.; Prokeš, L.; Jurmanová, J.; Buršíková, V.; Ráheľ, J.; Sťahel, P. Fast surface hydrophilization via atmospheric pressure plasma polymerization for biological and technical applications. *Polymers*, 2019, 11 (10), 1613. DOI:10.3390/polym11101613
Document Type: Article, IF = 3.4; JCR Category + Category Quartile: POLYMER SCIENCE Q1; AIS = 0.545

Huang, F.; Prokeš, L.; Němec, P.; Nazabal, V.; Havel, J. Comparison of clusters produced from Sb_2Se_3 homemade polycrystalline material, thin films, and commercial polycrystalline bulk using laser desorption ionization with time of flight quadrupole ion trap mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 2019, 30 (12), 2756-2761. DOI:10.1007/s13361-019-02346-9
Document Type: Article, IF = 3.3; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q2 + BIOCHEMISTRY & MOLECULAR BIOLOGY Q2; AIS = 0.757

Sťahel, P.; Mazánková, V.; Tomečková, K.; Matoušková, P.; Brablec, A.; Prokeš, L.; Jurmanová, J.; Buršíková, V.; Přibyl, R.; Lehocký, M.; Humpolíček, P.; Ozaltin, K.; Trunec, D. Atmospheric pressure plasma polymerized oxazoline-based thin films-antibacterial properties and cytocompatibility performance. *Polymers*, 2019, 11 (12), 2069. DOI:10.3390/polym11122069
Document Type: Article, IF = 3.4; JCR Category + Category Quartile: POLYMER SCIENCE Q1; AIS = 0.545

Prokeš, L.; Jarůšková, Z.; Petřík, J.; Frączek, M., Kalicki, T. Origin of a silver Stollhof-type disc excavated at Vanovice (South Moravia). *Praehistorische Zeitschrift*, 2020, 95 (1), 112-127. DOI:10.1515/pz-2020-0007
Document Type: Article, IF = 0.6; JCR Category + Category Quartile: ANTHROPOLOGY Q4; AIS = 0.389

Brtnický, M.; Pecina, V.; Vašinová Galiová, M.; Prokeš, L.; Zvěřina, O.; Juřička, D.; Klimánek, M.; Kynický, J. The impact of tourism on extremely visited volcanic island: Link between environmental pollution and transportation modes. *Chemosphere*, 2020, 249, 126118. DOI:10.1016/j.chemosphere.2020.126118
Document Type: Article, IF = 7.1; JCR Category + Category Quartile: ENVIRONMENTAL SCIENCES Q1; AIS = 1.023

Mazánková, V.; Sťahel, P.; Matoušková, P.; Brablec, A.; Čech, J.; Prokeš, L.; Buršíková, V.; Stupavská, M.; Lehocký, M.; Ozaltin, K.; Humpolíček, P.; Trunec, D. Atmospheric pressure plasma polymerized 2-ethyl-2-oxazoline based thin films for biomedical purposes. *Polymers*, 2020, 12 (11), 2679. DOI:10.3390/polym12112679
Document Type: Article, IF = 4.3; JCR Category + Category Quartile: POLYMER SCIENCE Q1; AIS = 0.597

Čech, J.; Sťahel, P.; Ráheľ, J.; Prokeš, L.; Rudolf, P.; Maršálková, E.; Maršálek, B. Mass production of plasma activated water: Case studies of its biocidal effect on Algae and Cyanobacteria. *Water*, 2020, 12 (11), 3167. DOI:10.3390/w12113167
Document Type: Article, IF = 3.1; JCR Category + Category Quartile: ENVIRONMENTAL SCIENCES Q2 + WATER RESOURCES Q2; AIS = 0.499

Prokeš, L.; Gorylová, M.; Čermák Šraitrová, K.; Nazabal, V.; Havel, J.; Němec, P. Arsenic-doped SnSe thin films prepared by pulsed laser deposition. *ACS Omega*, 2021, 6 (27), 17483-17491. DOI:10.1021/acsomega.1c01892
Document Type: Article, IF = 4.1; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q2; AIS = 0.630

Hegrová, J.; Prokeš, L.; Anděl, P.; Bucková, M.; Jandová, V.; Effenberger, K.; Ličbinský, R. Monitoring of the impact of road salting on spruce forest ecosystem in the vicinity of the highway D1 in the Bohemian-Moravian Highlands, Czech Republic. *Environmental Science and Pollution Research*, 2022, 29 (8), 11232-11242. DOI:10.1007/s11356-021-16468-9
Document Type: Article, IF = 5.8; JCR Category + Category Quartile: ENVIRONMENTAL SCIENCES Q1; AIS = 0.651

Procházková, M.; Killinger, M.; Prokeš, L.; Klepárník, K. Miniaturized bioluminescence technology for single-cell quantification of caspase-3/7. *Journal of Pharmaceutical and Biomedical Analysis*, 2022, 209, 114512. DOI 10.1016/j.jpba.2021.114512
Document Type: Article, IF = 3.4; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q2 + PHARMACOLOGY & PHARMACY Q2; AIS = 0.503

Calábková, G.; Chlachula, J.; Ivanov, M.; Hložková, M.; Czerniawska, J.; Vašinová Galiová, M.; Prokeš, L.; Gadas, P. Microbial degradation of Pleistocene permafrost-sealed fossil mammal remains. *Quaternary Research*, 2023, 111, 84-106. DOI:10.1017/qua.2022.28
Document Type: Article, IF = 2.3; JCR Category + Category Quartile: GEOGRAPHY, PHYSICAL Q3 + GEOSCIENCES, MULTIDISCIPLINARY Q3; AIS = 0.926

Sťahel, P.; Mazánková, V.; Podzemná, D.; Podzemná, E.; Pizúrová, V.; Jurmanová, J.; Prokeš, L.; Lehocký, M.; Ozaltin, K.; Pištěková, H.; Trunec, D. Antibacterial thin films deposited from propane-butane mixture in atmospheric pressure discharge. *International Journal of Molecular Sciences*, 2023, 24 (2), 1706. DOI:10.3390/ijms24021706
Document Type: Article, IF = 5.6; JCR Category + Category Quartile: BIOCHEMISTRY & MOLECULAR BIOLOGY Q2 + CHEMISTRY, MULTIDISCIPLINARY Q1; AIS = 1.030

Sťahel, P.; Mazánková, V.; Prokeš, L.; Buršíková, V.; Stupavská, M.; Lehocký, M.; Pištěková, H.; Ozaltin, K.; Trunec, D. Comparison of plasma-polymerized thin films deposited from 2-methyl-2-oxazoline and 2-ethyl-2-oxazoline: I Film properties. *International Journal of Molecular Sciences*, 2023, 24, 17455. DOI:10.3390/ijms242417455
Document Type: Article, IF = 5.6; JCR Category + Category Quartile: BIOCHEMISTRY & MOLECULAR BIOLOGY Q2 + CHEMISTRY, MULTIDISCIPLINARY Q1; AIS = 1.030

Bouška, M.; Milasheuskaya, Y.; Šlouf, M.; Knotek, P.; Pechev, S.; Prokeš, L.; Pečinka, L.; Havel, J.; Novák, M.; Jambor, R.; Němec, P. Low-temperature synthesis of GeTe nanoparticles. *Chemistry - A European Journal*, 2024, 30, e202402319. DOI:10.1002/chem.202402319
Document Type: Article, IF = 3.7; JCR Category + Category Quartile: CHEMISTRY, MULTIDISCIPLINARY Q2; AIS = 0.762

Čech, J.; Sťahel, P.; Prokeš, L.; Trunec, D.; Horňák, R.; Rudolf, P.; Maršálek, B.; Maršálková, E.; Lukeš, P.; Lavrikova, A.; Machala, Z. CaviPlasma: parametric study of discharge parameters of high-throughput water plasma treatment technology in glow-like discharge regime. *Plasma Sources Science & Technology*, 2024, 33, 115005. DOI:10.1088/1361-6595/ad7e4e
Document Type: Article, IF = 3.3; JCR Category + Category Quartile: PHYSICS, FLUIDS & PLASMAS Q1; AIS = 0.584

Michalicová, R.; Pecina, V.; Hegrová, J.; Brtnický, M.; Svoboda, J.; Prokeš, L.; Baltazár, T.; Ličbinský, R. Seasonal variation of arsenic in PM_{10} and PM_x in an urban park: The influence of vegetation-related biomethylation on the distribution of its organic species and air quality. *Chemosphere*, 2024, 362, 142721. DOI: 10.1016/j.chemosphere.2024.142721
Document Type: Article, IF = 8.1; JCR Category + Category Quartile: ENVIRONMENTAL SCIENCES Q1; AIS = 1.217

Bouška, M.; Gutwirth, J.; Bečvář, K.; Kucek, V.; Šlang, S.; Janíček, P.,; Prokeš, L.; Havel, J.; Nazabal, V.; Němec, P. Sc-doped GeTe thin films prepared by radio-frequency magnetron sputtering. *Scientific Reports*, 2025, 15, 627. DOI:10.1038/s41598-024-84963-3
Document Type: Article, IF = 3.9; JCR Category + Category Quartile: MULTIDISCIPLINARY SCIENCES Q1; AIS = 1.029

Novotný, K.; Krempl, I.; Afonin, I.; Prokeš, L.; Vaňhara, P.; Havel, J. Laser-induced plasma spectroscopy in pulsed laser ablation synthesis of nanoparticles in liquid. *Talanta*, 2025, 290, 127767, DOI:10.1016/j.talanta.2025.127767
Document Type: Article, IF = 6.1; JCR Category + Category Quartile: CHEMISTRY, ANALYTICAL Q1; AIS = 0.773

Prokeš, L.; Hegrová, J.; Průšová, B.; Baroň, M.; Hablovičová, B.; Sochor, J.; Ličbinský, R. Impact of traffic intensity and vehicular emissions on heavy metal content in vineyard soils, grapes, and wine: A comparative study of two vineyards in South Moravia (Czech Republic). *Environmental Geochemistry and Health*, 2025, 47, 216. DOI:10.1007/s10653-025-02530-9
Document Type: Article, IF = 3.8; JCR Category + Category Quartile: ENGINEERING, ENVIRONMENTAL Q3 + ENVIRONMENTAL SCIENCES Q2 + PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH Q1 + WATER RESOURCES Q2; AIS = 0.618

Horňák, R.; Čech, J.; Sťahel, P.; Prokeš, L.; Trunec, D.; Rudolf, P.; Maršálek, B. Spatial mapping of OH radicals produced by electric discharge in hydrodynamic cavitation cloud. *Journal of Physical Chemistry Letters*, 2025, 16, 6279-6285. DOI:10.1021/acs.jpclett.5c00979
Document Type: Article, IF = 4.6; JCR Category + Category Quartile: CHEMISTRY, PHYSICAL Q2 + MATERIALS SCIENCE, MULTIDISCIPLINARY Q2 + NANOSCIENCE & NANOTECHNOLOGY Q2 + PHYSICS, ATOMIC, MOLECULAR & CHEMICAL Q1; AIS = 1.176

Čech J.; Sťahel P.; Prokeš L.; Trunec D.; Horňák R.; Rudolf P.; Maršálek B.; Maršálková E.; Lukeš P. Glow discharge in water cavitation cloud with improved efficiency for hydrogen peroxide production. *Plasma Sources Science & Technology*, 2025, 34, 065009. DOI:10.1088/1361-6595/addf79
Document Type: Article, IF = 3.3; JCR Category + Category Quartile: PHYSICS, FLUIDS & PLASMAS Q1; AIS = 0.584

Papp, P; Mazánková, V.; Moravský, L.; Blaško, J.; Sťahel, P.; Prokeš, L.; Horňák, R.; Lehocký, M.; Pištěková, H,; Trunec, D. Comparison of plasma polymerized thin films deposited from 2-methyl-2-oxazoline and 2-ethyl-2-oxazoline: II Analysis of deposition process. *International Journal of Molecular Sciences*, 2025, 26, 8541. DOI: 10.3390/ijms26178641
Document Type: Article, IF = 5.6; JCR Category + Category Quartile: BIOCHEMISTRY & MOLECULAR BIOLOGY Q2 + CHEMISTRY, MULTIDISCIPLINARY Q1; AIS = 1.030

Appendix B. Attached Commented Articles