

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

Mgr. Ing. Lubomír Prokeš, Ph.D.

Modern analytical chemistry, applied in various fields of human activity, generate a large amount of numerical data (elemental concentrations, signal intensities, mass spectra, and chromatographic profiles), which, without sophisticated statistical analysis and results interpretation, remain fundamentally meaningless.

This habilitation thesis synthesises 18 selected research articles demonstrating how advanced statistical methods and chemometric approaches bridge the critical gap between raw measurement and scientific interpretation. The work encompasses four primary disciplines: archaeometry, geochemistry, environmental chemistry, and mass spectrometry, united by a consistent emphasis on rigorous data analysis and critical awareness of analytical bias.

Archaeological, geochemical, and environmental data usually exhibit specific characteristics (sampling bias, missing data, censored data, compositional nature of data, data with measurement error, etc.) that must always be taken into account when processing and interpreting the results.

Classification and provenance determination. Archaeological and geological materials are inherently multivariate in character, described by chemical or mineralogical compositions requiring advanced statistical approaches. The thesis demonstrates successful application of multivariate and machine learning methods (PCA, LDA, SVM, CART and RF), to classify pottery fragments from Bronze Age settlements, distinguish obsidian artefacts by source provenance using laser microsampling, and identify conflict minerals (coltan) to support EU supply chain regulations. The work emphasises that machine learning approaches substantially outperform classical techniques in handling complex geochemical fingerprints and mitigating compositional constraints.

Spatial geochemical data analysis in archaeology and geochemistry. Chemical mapping of archaeological soils reveals historical activity zones (metalworking, waste disposal) through multivariate analysis of elemental distributions. Using geostatistics, inverse distance weighting interpolation, and compositional data analysis (centered log-ratio transformation), the thesis demonstrates how spatial autocorrelation metrics (local Moran's I, Getis-Ord G, Lee's L) and clustering methods identify functional specialisation in archaeological contexts. In addition, elemental analysis of agricultural and contaminated soils allows the delineation of geochemical anomalies and environmental monitoring.

Elemental bioimaging and chemical mapping. Techniques suitable for hyperspectral and elemental mapping data treatment include principal component analysis or some other

projection methods enable reduction and visualisation. Unsupervised clustering (k-means, fuzzy c-means, etc.) and spatial autocorrelation analysis (LISA) facilitate segmentation of elemental maps into meaningful chemical domains.

Interpretation of mass spectrometry data. The thesis addresses challenges specific to mass spectrometry data, where the number of variables exceeds the number of observations ($p \gg n$). Two complementary workflows are presented:

(a) Treatment of laser desorption-ionisation time-of-flight (LDI-TOF) mass spectra of inorganic materials through open source R-based spectral decomposition, visualisation, and stoichiometric determination;

(b) Classification of biological samples via principal component analysis and artificial neural networks applied to intact cell mass spectrometry fingerprints for quality control and phenotypic discrimination.

A novel R-based liquid biopsy workflow that integrates MALDI mass spectrometry with machine learning demonstrates reproducible biomarker discovery and clinical application.

Comparative evaluation of analytical methods. The thesis critically re-examines how different analytical techniques should be compared. It advocates the replacement of misleading approaches (Pearson correlation, paired t-test) with rigorous statistical frameworks. Bland-Altman plots, regression methods that account for errors in both axes (orthogonal regression, Deming regression, York regression, bivariate least squares regression) and visualisation via principal component analysis can be applied for methods bias assessment. Analysis of means (ANOM) provides robust alternatives to traditional hypothesis testing.

The aim of this thesis is to show that advanced data analysis is not a luxury but a necessity in modern analytical chemistry. The synthesis of established chemometric techniques with emerging computational methods provides practical guidance for researchers navigating the complexities of archaeometric, geochemical, and bioanalytical investigations, ultimately enhancing the reliability, reproducibility, and defensibility of scientific conclusions across disciplinary boundaries.

List of Annotated Papers

[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.; Vitešnicková, 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] Přišťáková, 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%