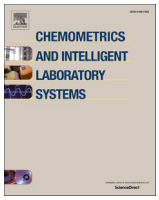
Chemometrics and Intelligent Laboratory Systems 247 (2024) 105095

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



Analytical chemistry meets art: The transformative role of chemometrics in cultural heritage preservation

Jordi Riu [a,1], Barbara Giussani [b,1,*]

[a] Universitat Rovira i Virgili, Department of Analytical Chemistry and Organic Chemistry, Carrer Marcel·lí Domingo 1, 43007, Tarragona, Spain
[b] Dipartimento di Scienza e Alta Tecnologia, Università degli Studi dell'Insubria, Via Valleggio, 9, 22100, Como, Italy

ARTICLE INFO

Keywords:
Cultural heritage
Chemometrics
Exploratory methods
Multivariate regression
Classification

ABSTRACT

This review intends to illustrate the fruitful collaboration between chemometrics and cultural heritage science. It showcases past achievements, aiming to inspire numerous cultural heritage researchers who have yet to incorporate multivariate techniques in analysing their research findings. The goal is to practically discuss applications with examples extracted from the literature. These range from the most basic early applications, influenced by the limited laboratory equipment available at those times, to the more contemporary research endeavours. Recently, numerous new analytical instrumentations have gained widespread adoption, some of which were previously inconceivable for field use or analysing micro-samples, characteristics needed to work with cultural heritage. Chemometrics serves as the binding element in this process. It handles the multivariate data generated by analytical instruments, even from multiple instruments used to characterize the same sample and yields easily interpretable graphs that encapsulate all the information considered simultaneously. This review will delve into the challenges of achieving commendable results and constructing effective descriptive or predictive models. Additionally, it will offer essential theoretical insights crucial for a comprehensive grasp of the fundamental algorithms employed. We have explored fundamental qualitative and quantitative models capable of addressing most issues encountered in studying a historical or artistic artifact. Additionally, focus was directed towards data pre-processing, at times essential for enhancing model outcomes.

1. Introduction

The term "chemometrics" was coined in 1971 by Svante Wold [1]. In 1974 he defined it as "the art of extracting chemically relevant information from data produced in chemical experiments" [2]. Chemometrics was then defined as "an art." Twenty years later, Wold himself described chemometrics as "how to get chemically relevant information out of measured chemical data, how to represent and display this information, and how to get such information into data." In 1990, Geladi and Esbensen provided fascinating insights into the origins of this field [3,4].

Chemometrics has consistently been associated with analytical chemistry, involving the retrieval of pertinent insights from chemical data. In 1981 A. Borman defined it as a new direction in analytical chemistry [5], while in 1998 Wold and Sjöström celebrated the success of the discipline [6]. Several authors described the evolution of the connection between chemometrics and analytical chemistry in their articles [7–12]. Back in 1988, an article in Nature highlighted two newly established journals, namely "Journal of Chemometrics" and "Chemometrics and Intelligent Laboratory System", dedicated to compiling both theoretical and practical articles on the application of chemometric techniques in the field of analytical chemistry [13]. They still are the two reference journals for finding the state-of-the-art in research within this discipline, even though nowadays, articles on chemometrics can be found in all journals specialized in analytical chemistry and in many journals of related areas. An excellent historical introduction to the link between chemometrics and analytical chemistry can be found in the recent articles by Brereton and co-authors [14,15]: the introductory part delves into the historical roots of the discipline, while the main body of the articles explores its role in enhancing different aspects of optimizing an analytical procedure. It is worth mentioning the international conference known as "Chemometrics in Analytical Chemistry – CAC", which serves as a global gathering for scholars worldwide who are involved in the field of chemometrics within analytical chemistry.

* Corresponding author.
E-mail address: barbara.giussani@uninsubria.it (B. Giussani).
[1] Authors have contributed equally.

https://doi.org/10.1016/j.chemolab.2024.105095
Received 13 December 2023; Received in revised form 22 February 2024; Accepted 23 February 2024
Available online 28 February 2024

Returning to definition of chemometrics according to Wold, we genuinely lack certainty regarding whether chemometrics is an artistic endeavour or a strictly methodical and objective approach to data analysis, a way to solve chemical problems using experimental data efficiently and economically [16]. Nevertheless, it undeniably shares a connection with the realm of art and, more broadly, cultural heritage [17,18].

This association is cultivated primarily through the discipline of analytical chemistry because analytical chemistry is able to provide data from the analysis of artworks: over the years, some articles have chronicled the evolution of the connection between analytical chemistry and cultural heritage [19,20]. A recent article published by Luque de Castro and Jurado-Lopez describes precisely how analytical chemistry enters many areas of research in the field of cultural heritage in studies with a strong multidisciplinary character [21]. They state that reviewing previous and current publications in this field has brought to their attention the limited involvement of analytical chemists in works dealing with cultural heritage, despite the crucial role of analysis in cultural heritage research: individuals from different professions handle the responsibilities typically associated with analytical chemists.

The purpose of this article is to encourage scientists dealing with cultural heritage to use chemometric tools for their data analysis. Most likely, they possess chemical data (and potentially other types such as physical, structural, or descriptive data) suitable for processing with chemometric methods rather than relying solely on univariate statistics. These data are more aptly processed or yield greater benefits when utilizing chemometric techniques. We would like to demonstrate the advantages of using chemometric techniques to process data collected on cultural heritage and how to do it in practice. Drawing inspiration from previously published works, we will provide the reader with suggestions on best practices for utilizing chemometric techniques when handling cultural heritage data. We will concentrate on the most straightforward techniques, elucidating their benefits and the information they can provide to researchers working on cultural heritage in an easily understandable manner. These basic methods often yield the most favourable results and facilitate straightforward interpretation of the findings.

2. Analytical techniques for cultural heritage

As a first step, we would like to clarify what is meant by chemometrics and multivariate analysis. Chemometrics is a specialized field that applies, among others, multivariate analysis techniques to data from chemical analysis. It involves the application of multivariate statistical methods to extract meaningful information from complex chemical data sets, making it a powerful tool for chemical analysis and research. Multivariate analysis is thus a statistical and mathematical approach used to analyse data sets that involve multiple variables or measurements, that often can be correlated.

Multivariate analysis is capable of handling correlated variables by considering the interconnections and dependencies that exist among these variables. In recent years, most of the data obtained from the application of analytical chemistry techniques to cultural heritage, is of multivariate nature. Therefore, the significant benefits of employing chemometric methods in the study of cultural heritage become evident. To begin with, on one hand it is crucial to acknowledge that numerous analytical methods applied to the study of cultural heritage generate extensive sets of interconnected variables, e.g., the case of spectroscopic techniques. On the other hand, multiple techniques are frequently employed in the analysis of cultural artifacts, and while they often provide complementary findings, this is not always the case. The concurrent analysis of all the data generated from these diverse methods can significantly assist researchers in gaining a comprehensive understanding of their subject of investigation.

As previously noted, spectroscopy stands as one of the most extensively employed analytical methods within the field of cultural heritage research. It possesses the potential for non-destructive and portable application, and its capacity to furnish insights into the molecular composition of a sample varies depending on the wavelength range and analytical mode used [22,23]. For the analysis of cultural heritage, the mid-infrared range (4000-400 cm^{-1}) is frequently used [24–26]. Recently, the predominant application of FTIR (Fourier-transform infrared spectroscopy) in heritage science has been the precise molecular identification and localization of both organic and inorganic constituents in micro-samples. Nonetheless, with the advancements in optical materials and components due to technological progress over the past decade, including the development of micromachining technologies such as MEMS (micro-electromechanical systems) and MOEMS (micro-optoelectromechanical systems), FTIR has also emerged as a noteworthy method for non-invasive surface analysis of artworks using fully portable instrumentation [27,28]. Apart from the traditional single-point detection, the use of two-dimensional mapping and imaging is particularly advantageous for visually representing the chemical composition of cultural objects with multi-layered structures [29]. Reflectance imaging spectroscopy within the visible and near-infrared (NIR) spectral range (400–2500 nm) is also experiencing a growing presence and significance within cultural heritage research [30–33], while imaging methods commonly employed in the field of medicine have started being utilized for the preservation and study of cultural heritage [34]. Fig. 1 show the picture of a portable FTIR used for non-invasive investigations of the surface of painted Japanese artworks [35].

Raman spectroscopy has established a specialized role within the areas of diagnosing and preserving cultural heritage [36]. In the context of a wide range of materials, including minerals, gemstones, rocks, patinas, corrosion products, glass, pottery, mortars, dyes, binders, resins, paper, parchment, inks, and human remains, Raman micro spectroscopy and SERS (surface-enhanced Raman spectroscopy) provide solutions to archaeometric and conservation-related inquiries, facilitating in situ investigations [37–39].

Spectroscopic techniques frequently used in the study of metallic compounds in cultural heritage are XRF (X-ray fluorescence) [40] and LIBS (laser-induced breakdown spectroscopy) [41–43]. These techniques are frequently applied without using the whole spectra that the techniques may provide, only using the specific information of the elementary composition that can be derived from the application of these techniques. In contrast to other application areas, the inherent micro-destructive quality of the LIBS technique has somewhat restricted its adoption in cultural heritage. Additionally, the laser ablation method, when coupled with ICP-MS (inductively coupled plasma mass spectrometry) [44], presents another micro destructive approach often utilized for metal analysis, surpassing the requirement for destructive analyses such as atomic spectrometry [45].

Currently, neutron-based techniques are widely employed in neutron-beam facilities worldwide to examine historical and culturally valuable objects. These methods are non-invasive and non-destructive, making them well-suited for gaining structural insights into artifacts. They can provide information about composition, environmental effects, inclusions, structure, manufacturing techniques, and elemental composition, collectively forming a unique profile of the characteristics of an object through neutron-matter interactions [46,47].

As previously stated, spectroscopic applications frequently involve multivariate data analysis, as we will explore further in subsequent sections of this article. However, there are signals that inherently possess multivariate properties but are frequently analysed using univariate techniques instead. A notable example of this occurs in chromatographic techniques, which are also applied to cultural heritage studies [48]. In the majority of cases, researchers opt to calculate the area of individual peaks of interest rather than utilizing the entire chromatogram. A similar situation applies to data generated by electroanalytical sensors used in the cultural heritage field [49] however, in some cases, data processing involves the utilization of chemometric methods [50].

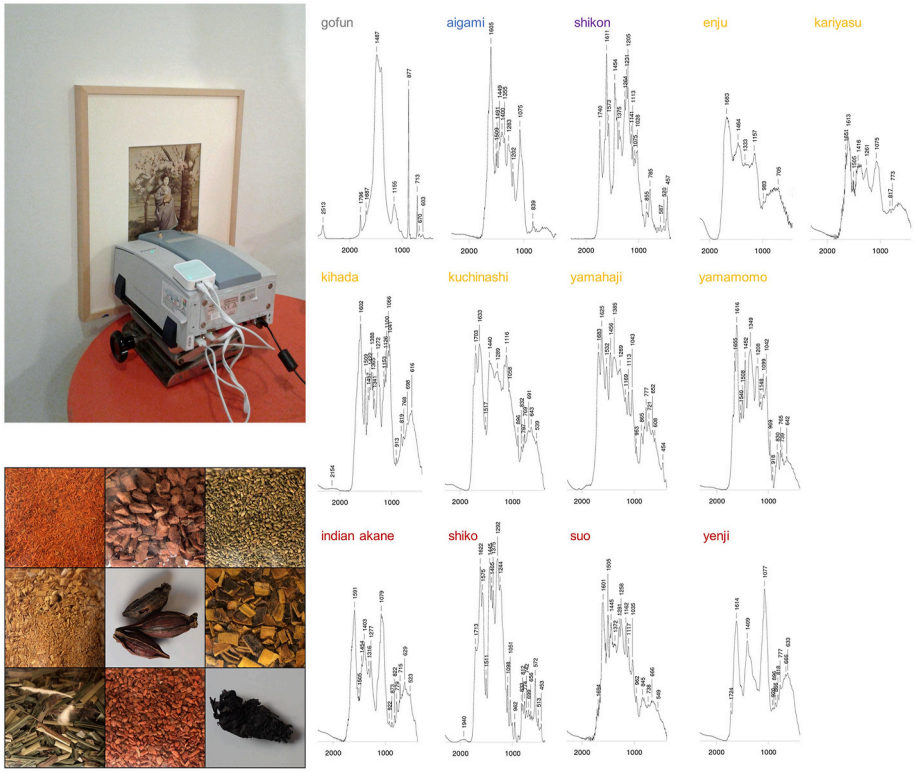


Fig. 1. Photographs of the Alpha Bruker FTIR portable spectrophotometer analysing a photograph and of some of the raw materials used to extract Japanese traditional pigments (left, bottom). FTIR spectra in pseudo-absorbance of the reference traditional Japanese pigments prepared in laboratory (right), presented as pseudo-absorbance (arbitrary units, y-axis) vs. wavenumber (cm−1, x-axis). Reproduced with permission from Ref. [35].

Thermogravimetric (TG-DTG) curves were also have also been employed in archaeometric contexts [51].

3. Chemometric applications in cultural heritage

3.1. Multivariate data matrices

As previously mentioned, many insights derived from artworks are of multivariate nature because numerous chemical measurements are inherently multivariate. But this is also a result of the ongoing development of analytical instrumentation, cost reduction, and, as a result, increased accessibility.

Examples of data of this type are for instance the evaluation of paint cross-sections by FTIR in fragments collected from a mural temple in Nepal [52] or the analysis of nine metal oxides by XRF in antique amphora [53]. In the example of the evaluation of paint cross-sections, the number of variables is 1331, corresponding to 1331 wavenumbers between 6000 and 675 cm^{-1}. In the example of the analysis of metal oxides in antique amphora, the number of variables is nine, corresponding to the number of analysed metal oxides. Although these two datasets are both multivariate, as more than one measurement was performed for each sample (absorbances at different wavelengths or the analysis of different metals), a clear distinction can be made between them. In the first case, with FTIR spectra, the absorbance of the i*th* wavelength is not independent from the absorbance of the immediately previous and following wavelengths. In the second case, the concentration of one specific element is independent from the concentration of the other analysed elements. It is worth mentioning that autocorrelated data (as the example of FTIR measurements) tend to have a significantly higher number of variables compared to independent data: the number of wavelengths or wavenumbers in a spectrum is typically much greater than the number of chemically analysed elements (or physical characteristics of the sample). Both categories of datasets are suitable to treatment through chemometric methods. However, as we will observe in section 2.5, they may need distinct preprocessing steps prior to calculation.

A crucial question is what we want to know from the data. We might want to ascertain whether all the analysed metals provide unique information, or if, for instance, the information from one metal is replicated by another one. For instance, we may aim to determine if samples form clusters or if it is feasible to predict a certain property of a sample based on its IR spectrum. Hence, we may have either qualitative or quantitative needs, and as described in the upcoming chapters, chemometrics can prove to be valuable in all these scenarios.

3.2. Data overview

Are there groups among the samples analysed - e.g., do all the previously mentioned amphora in section 2.1 come from the same production centre, or are there different production centres? Which are the most important variables when different metal oxides are analysed? Is there any metal oxide that provides non-significant information in the analysis of amphora, so that this or these metal oxides do not need to be considered in future analysis? These questions may be solved by plotting the values of the different variables, so that each sample has a single point in the plot defined by the different variables, and then checking for groups in the plotted samples. But when dealing with multivariate data all these operations are not simple. In the two examples in section 2.1 one should be able to visualize a plot in nine-dimensions (the nine metal oxides analysed in antique amphora) or in 1331-dimensions (the 1331 wavenumbers in the paint cross-sections from a mural temple), or in a rather large number of two-dimensional graphs to cover the different dimensions. Fortunately, there are techniques that help the researcher to visualize multivariate data and to draw conclusions regarding similarities or non-similarities between samples and correlations between variables.

3.2.1. Principal component analysis

The most widely used technique for dimensionality reduction, and that has been applied in many fields, is principal component analysis, PCA [54]. It is a multivariate exploratory tool that focuses on the data representation and interpretation, and that is extensively described in many reference textbooks [55–58]. PCA aims to transform data from a high-dimensional space (the space defined by the original variables of the data set) into a lower-dimensional space while preserving all the relevant information. It condenses information contained in large amounts of initial variables into a few parameters, called principal components (PCs), which capture the similarities among the original samples and covariances between variables. PCs are orthogonal directions in the data space, each ordered by the amount of variance they explain. The first PC retains the most variance of the data, the second PC the most variance not explained by the first, and so on.

These components serve as a new basis for projecting the data onto this reduced-dimensional space, so that plots in 2D or 3D (plots using the two or three first PCs) preserving most of the original information can be obtained. The transformed data in the new coordinate system of the PCs reveals the essential patterns and relationships among the observations. Visualizing data in this reduced space often simplifies analysis and interpretation.

The data matrix (**X**), of dimensions I x J (where I is the number of samples analysed and J is the number of variables used in the analysis) is decomposed by the PCA algorithm as follows:

$$\mathbf{X} = \mathbf{T}\mathbf{V}' + \mathbf{E} \qquad \text{Eq. 1}$$

where:

- **T** is the matrix of scores (I x A). They are the coordinates of the initial data points in the new space defined by the PCs. The scores plots show patterns or groups (if presents) among the samples.
- **V** (or **P**) is the matrix of loadings (J x A). They are the coefficients that describe the contribution of each original variable to a specific PC. They reveal how strongly each variable influences the PCs (and therefore indicate the importance of the original variables in differentiating the samples) and covariances between variables.
- **E** is the error matrix, of the same dimensions as **X.**

The number of principal components, A, is selected so that they describe the structured part of data (typically the user will only select the top first PCs that retain most of the explained variance). Fig. 2 shows the geometrical representation of PCA for a data set of 20 samples characterized by two variables.

When the variables are independent, PCA is primarily employed to determine the presence of covariance among variables and whether these covariances have an impact on the grouping or patterns observed in the samples. When the variables are autocorrelated, as in the case of spectra, PCA is used to significantly reduce data dimensionality while retaining most of the information. PCA can thus reveal components (combinations of wavelengths) that are important in explaining variability in the data and facilitating the visualization of hidden patterns related to spectral features.

The application of PCA to results from analytical techniques using independent variables has been used for 40 years. As examples, 27 metals were analysed in medieval Bulgarian glasses by neutron activation analysis (NAA) to evaluate the technological level of glass production in medieval Bulgaria [60], or 10 metal oxides in 50 samples of glass-making objects from the Roman area were analysed by XRF (using only the composition of the elements and not the whole spectra in the data analysis) and PCA was applied to make hypotheses concerning the origin of the samples [61]. The advancements in analytical instrumentation during that period enabled the introduction of new rapid and relatively affordable techniques like ICP-OES (inductively coupled plasma optical emission spectroscopy). These methods were utilized

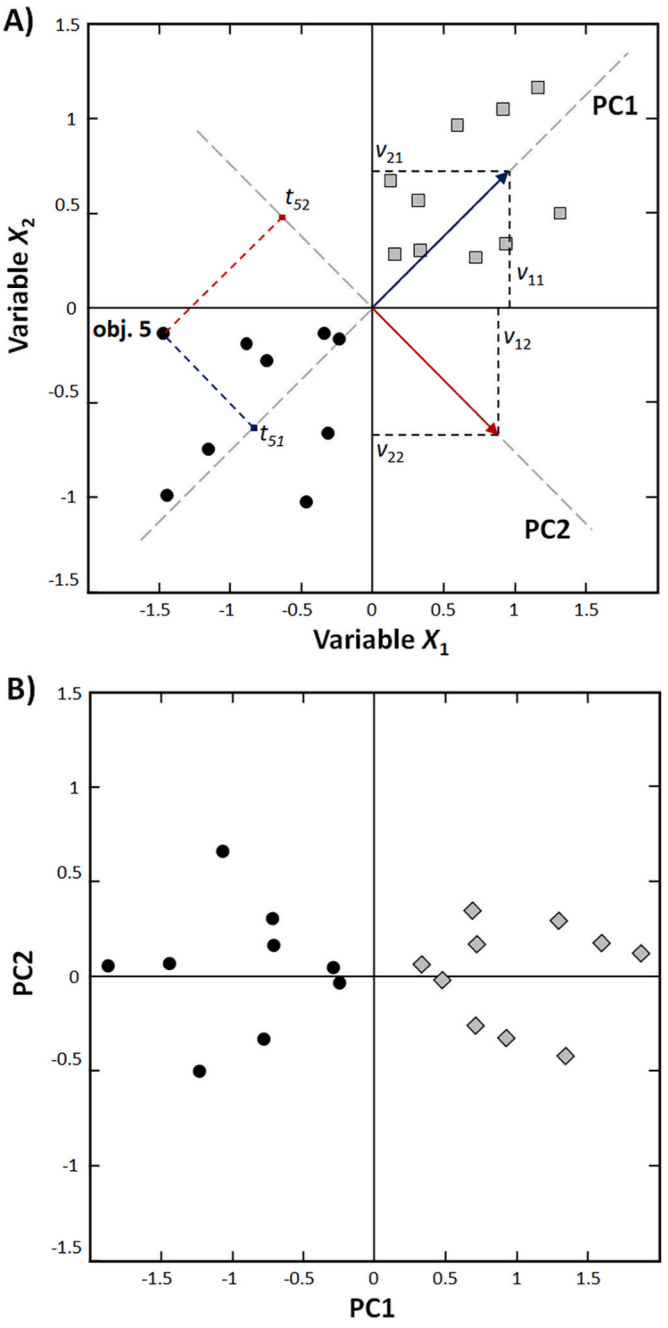


Fig. 2. PCA example with 20 samples (black circles and grey squares) characterized by two variables. (A) The samples are plotted in the space of the original variables x_1 and x_2. The blue and red lines represent the directions of PC1 and PC2 axes, respectively. The coordinates of the blue arrow are v_{11} and v_{21}, the loading values of variables x_1 and x_2 on the first PC, respectively. The coordinates of the red arrow are the v_{12} and v_{22}, the loading values of variable x_1 and x_2 on the second PC, respectively. The score values are the orthogonal projection of the sample coordinates on the PC axes. As an example the scores of sample 5 are shown: t_{51} (PC1 score) and t_{52} (PC2 score). (B) The 20 samples represented in PC space: PC1 versus PC2. Reproduced with permission from Ref. [59]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

alongside others to identify compositional groups of Roman pottery [62] and to detect counterfeit components in ancient marbles [63]. ICP-MS was successfully employed to determine dietary habits through the analysis of trace elements on ancient remains, allowing the classification of adult individuals according to the historical period in which they

lived [64]. LIBS coupled with PCA allowed to distinguish surface contaminations and retouches from the original pictorial materials in small fragments of the oil paintings "Emmaus Dinner" by Bellini (17th century) and "Annunciation" by Brughi (18th century) through the study of stratigraphic variations of the chemical composition [65]. The emergence of portable instrumentation in recent years also gained importance in the analysis of artworks since this instrumentation allows for in situ analysis without the need for the artwork to be taken to the laboratory. In this way, a portable XRF analyser was used for the analysis of major and trace elements (using again only the composition of the elements and not the whole spectra in the data analysis) in a provenance study of marble quarries of Delos island (Greece), finding which of the analysed artifacts came from a Delian quarry and which ones were not of local origin [66].

Regarding results from analytical techniques using autocorrelated variables applied to cultural heritage, the main group of applications comprises different kinds of spectroscopies (e.g. FTIR, NIR, Raman spectroscopy) but also other non-spectroscopic techniques such as chromatographic techniques or mass spectrometry fall into this category. Spectroscopic techniques, which have the advantage of being non-destructive techniques (an invaluable characteristic when working with artworks), have been widely used in recent years in this field. To give some examples, Marengo et al. [67] used ATR-FTIR (attenuated total reflectance Fourier-transform infrared) and statistical process control principles for monitoring the conservation state of artworks. The technique was applied to some canvas painted with mixtures of three organic pigments to identify the starting of the degradations and to provide insights about the chemical alterations induced by the UV exposure. ATR-FTIR was also applied to evaluate the effect of different paint cross-section preparation methods applied to fragments collected from a mural temple (15th century) in Nepal [52]. ATR in combination with mid-FTIR was used in paper samples to study paper restoration processes [68]. Two hydrogels with different cleaning capacities were applied to papers from the 18th century, showing a greater efficiency than traditional methods. FTIR was also applied to identify organic binding media in a variety of samples including wallpapers and polychromed alabaster artworks [69]. The proposed method simplifies notably the interpretation of the obtained spectra, being able to obtain three different sources of information from the same sample. μATR-FTIR chemical mapping coupled with chemometric analysis were applied to three paint samples characterised by different types of coatings/treatments. An efficient spatial stratigraphic localization of compounds of interest was achieved, as well as a differentiation between heterogeneous mixtures characterised by similar spectral features, impossible to identify by a univariate approach. Particularly interesting is the use of the PCs score maps for result interpretation [70] (Fig. 3).

NIR spectroscopy was applied to the characterization of paint cross-sections from a Renaissance wooden painting and from a 15th century wall paintings of the Basilica di San Frediano, Lucca (Italy) [71]. The method was able to provide information for the characterisation of both inorganic and organic compounds within complex paint stratigraphies. Finally, Raman spectroscopy was used to diagnose the impact and conservation state of Pompeian walls exposed to diverse environments [72].

The increase of computer power also enabled the use of high-intensive computing needed in techniques such as hyperspectral imaging [73,74], which combines spectroscopy with imaging capability. Hyperspectral imaging collects and processes information across the electromagnetic spectrum to obtain the spectrum for each pixel in an image. Due to the particular complexity of data obtained from

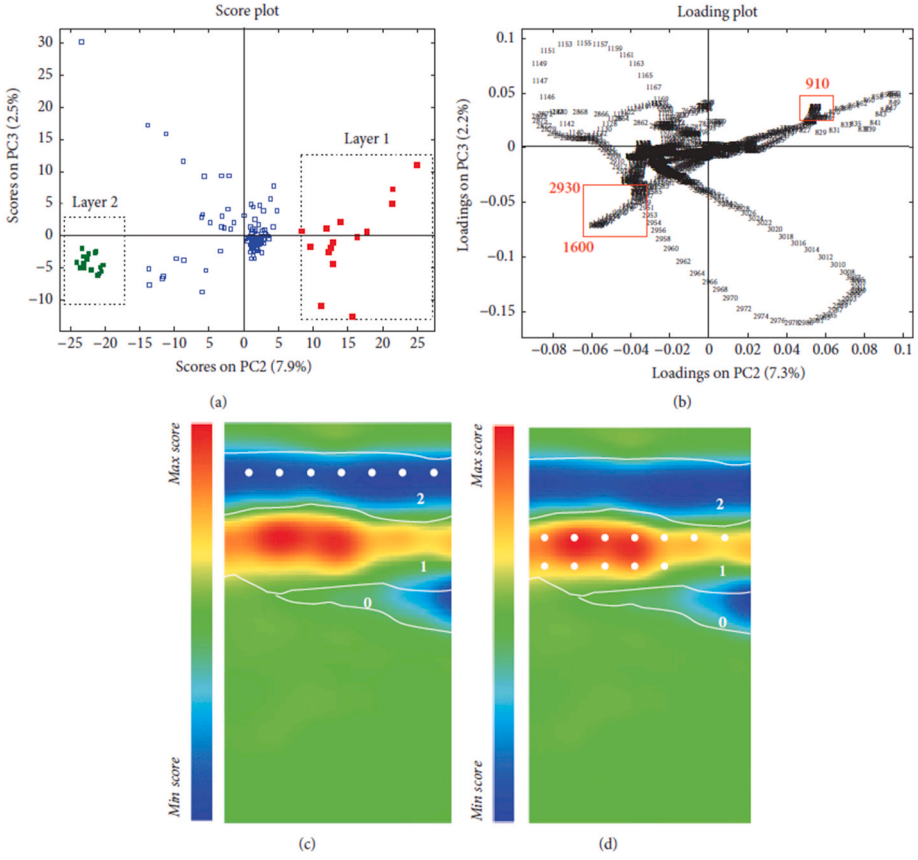


Fig. 3. a) score plot of one of the examined samples. Clusters highlighted in red indicate the objects points localised within one of the examined layers (white dots in (c)) and green for points localised within another layer (white dots in (d)). b) loading plot of the same sample. c) and d) PC2 score map. Reproduced with permission from Ref. [70]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

hyperspectral techniques, specific pre-processing methods have to be applied before data analysis. Image compression, selection of the region of interest or dead pixels [75,76] must be considered as well as the pre-processing techniques of classical spectroscopy. Hyperspectral imaging has been applied in recent years to different artworks. A multivariate approach for processing X-ray fluorescence spectral and hyperspectral data from non-invasive in situ analyses on painted surfaces was developed by Sciutto et al. [77]. This technique was applied to data obtained from the analysis of a Renaissance panel painting in Palazzo Ducale of Urbino (Italy), allowing for a fast interpretation of results that may be useful to support the definition of the sampling points. FTIR hyperspectral imaging was applied to the analysis of degradation products in a patina from the arch of Septimius Severus in the Roman Forum [78] (Fig. 4a shows the PCA differentiation among the different compounds in the scores plot and Fig. 4b shows the most important variables in the loading plot) and to the 16th century Neptune fountain in Bologna (Italy) [79]. In the former case, the method allowed to the discrimination and characterization of complex mixtures through the exact localization of similar compounds and the identification of the heterogeneous mixtures in the studied sample. In the latter case, the method was proposed for the characterization of coatings applied as protective agents, allowing to reduce the time required for data processing by maximizing the degree of automation of the procedure. Fig. 4a shows one of the issues frequently encountered in the visualization of data with chemometric models, attempting the representation of a pseudo-3D figure (probably to increase the percentage of the explained variance) on a 2D support, thereby somewhat complicating the visibility of which PC separates the different groups.

Portable instrumentation has also been used with analytical techniques using autocorrelated variables. For instance, XRF spectroscopy using portable instrumentation has been used to differentiate pigments in artworks [80]. Portable instrumentation was compared with benchtop equipment, showing the possibility to clearly discriminate important pigments in the PCA scores plot. The same technique also using portable instrumentation was applied to the classification of icons from a monastery in Serbia and of modern paints from the beginning of the 20th century [81]. Another miniaturized shortwave infrared (SWIR) spectrometer was used for the analysis of cultural heritage samples (archaeological bones, cinematographic films and bronze patinas) [31]. The results allowed to differentiate the materials used as a support for cinematographic films or to differentiate between corrosion products on bronze sculptures, which is important for assessing the state of conservation of the artwork. A portable FTIR spectrometer working in external reflection mode was used to study historical silk samples coming from traditional Japanese samurai armours (15th-20th century), allowing in a non-invasive way to detect if the samples had been originally subjected to degumming [82].

Finally, chromatographic techniques were also used in different artworks. Direct exposure electron ionization mass spectrometry (DE-MS) and gas chromatography combined with mass spectrometry (GC-MS) were used to analyse organic residues contained in an Egyptian censer (5th–7th century AD) [83]. The same research group also applied GC-MS and HPLC-MS (the combination of high-performance liquid chromatography – HPLC - and mass spectrometry - MS) to determine the influence of the relative humidity on the oxidation and hydrolysis of fresh and aged oil paints [84], in order to understand the evolution of the composition of modern oil paints during ageing under the influence of environmental risks.

3.2.2. Cluster analysis methods

Another frequently used tool for exploratory purposes is cluster analysis [55,56]. While PCA is a data technique used to reduce the dimensionality of data and identify important relationships or patterns looking for differences and variations among the samples, cluster analysis methods focus on grouping samples into clusters based on their similarity without considering, at least at first, dimensionality reduction. These methods therefore assess how closely samples resemble each other based on distance metrics: samples that are more similar are placed into the same cluster or group, while those that are dissimilar are assigned to different clusters. The similarity between two samples of multivariate nature having a total of M variables is typically calculated using the distance (d_{ij}) between all pairs of samples, where the most used distance is the Euclidean distance:

$$d_{ij} = \sqrt{\sum_{m=1}^{M} \left(x_{im} - x_{jm}\right)^2} \quad \text{Eq. 2}$$

where x_{im} and x_{jm} are respectively the values of the mth variable for the ith and jth sample. The similarity between a pair of samples is calculated from this distance, and follows equation (3):

$$s_{ij} = 1 - \frac{d_{ij}}{d_{max}} \quad \text{Eq. 3}$$

being $d_{\max}$ the highest distance between all pair of samples. In this way all the similarity values range in the interval [0,1]. Other distances than the Euclidean distance in equation (2) (e.g. Mahalanobis distance, Manhattan distance and others [85]) can be used to calculate the distance between the pairs of samples (to delve deeper into the topic of distances, research articles and books of Prof. Todeschini research group are recommended, for instance Ref. [86]). While using PCA one can obtain insights into groups of samples and the significance of variables, with clustering techniques one only gathers information about similarities and dissimilarities between samples. Probably because of this reason, cluster analysis is typically used in combination with PCA. At times, when dealing with numerous input variables, cluster analysis is

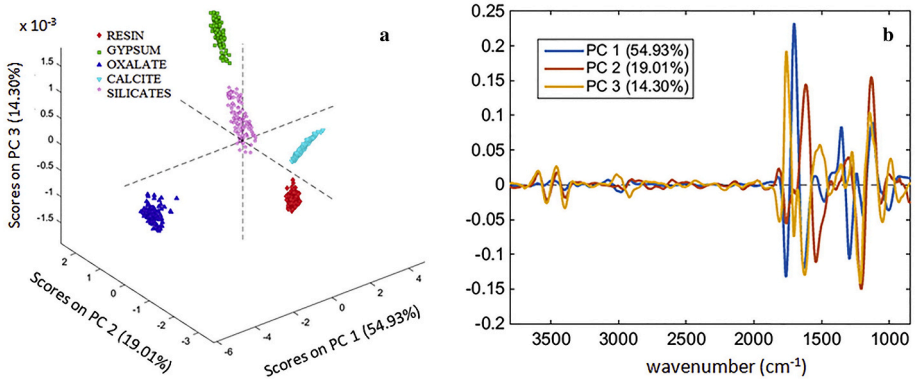


Fig. 4. a) 3D score plot of PCA with the first three PCs; b) loading plot of the PCA models of the analysis of the Arch of Septimius Severus in the Roman Forum (3rd century AD). Reproduced with permission from Ref. [78].

often carried out using principal components, which have been pre-computed from the original data matrix and account for approximately 95% of the explained variance.

Cluster analysis methods can be categorized into two groups: partitioning and hierarchical methods. Partitioning methods focus on the task of segmenting a large dataset containing diverse objects into *k* clusters, where *k* can either be predetermined or estimated through an exploratory process or even determined by the algorithm iteratively. An example of this approach is the *k*-means technique [87]. In contrast, hierarchical cluster methods build a tree-like structure, known as a dendrogram, which graphically represents the hierarchical relationships between clusters [88]. The most used hierarchical clustering is the agglomerative clustering approach, which initiate with each data point as a separate cluster and gradually merge clusters based on their similarity until a stopping criterion is met. Hierarchical clustering does not require specifying the number of clusters in advance and can provide insights into structures in the data. It offers a more flexible approach but can be computationally more intensive due to the construction of the hierarchy. Virtually all the cited references using hierarchical clustering use the agglomerative clustering approach.

As with PCA, clustering analysis has been applying to cultural heritage for almost 40 years. The most used clustering method applied to analytical techniques that use independent variables is hierarchical clustering. These techniques have been mainly applied to very ancient artworks, and to name some examples covering all these years, several elements from European medieval stained glasses were analysed by AAS (atomic absorption spectroscopy), FE (flame photometry) and OES (optical emission spectroscopy) and dendrograms (among other chemometric techniques such as PCA) helped to give some insight into the similarities between different groups of European mediaeval stained glasses [89]. INAA (instrumental neutron activation analysis) and ICP-AES (inductively coupled plasma atomic emission spectroscopy) were used to determine the chemical constituents of ancient roman Samian pottery finding that different workshops were sharing the same clay [90]. 14 elements in antique ceramics were determined by ICP-OES and AES to define groups of different pieces of pottery [91], and INAA was used to analyse different elements in medieval pottery and hierarchical cluster analysis and other techniques such as PCA helped to determine their classification and provenance [92]. A Graph Theory-based blind clustering technique was employed, analysing the entire LIBS spectrum without singling out individual atomic lines, in the study of limestone Nuragic statues from Mont'e Prama site (Sardinia, Italy). This method assesses spectrum similarity by computing their correlation function, visually demonstrating it as node distance within a graph (Fig. 5). Greater correlation results in closer node proximity in the graph (see Ref. [93] and references therein).

Clustering analysis has also been applied in combination with portable instrumentation to cultural heritage. In the reviewed cases, portable instrumentation used only the composition of the elements and not the whole spectra in the data analysis, and therefore variables are considered to be independent. A portable XRF instrument was used to analyse pottery samples from an archaeological site in Mexico, showing a high degree of discrimination between groups of samples [94]. A different portable XRF instrument was used in the analysis of the surface layers of different artworks belonging to the museum collection and storeroom of the Royal Palace in Caserta (Italy), where dendrograms and the *k*-means technique were used to find many different novelties in the analysed artworks [95].

Regarding results that use clustering techniques in data from analytical techniques using autocorrelated variables applied to cultural heritage, different spectroscopic techniques using the whole measured spectra have been used. Probably XRF is the most used one, but several other ones have also been applied to cultural heritage pieces. A benchtop FTIR was used to analyse waterlogged archaeological wood from an ancient Chinese shipwreck [96], where hierarchical cluster analysis was used to classify the recovered samples into four different groups. A different study consisted on the NIR measurements for authenticating stamps of 12 seals on a Chinese traditional painting [97]. Hierarchical cluster analysis was also used to find the time period of some unknown seal stamps.

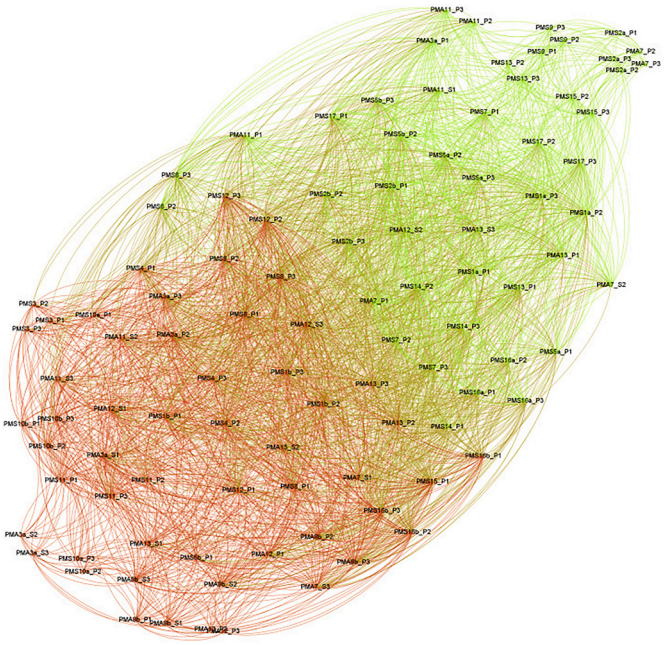


Fig. 5. a) Graphical representation of the graph built using the data of LIBS analysis on the Nuragic statue samples from the archaeological site of Mont'e Prama. Two different (partially overlapping) clusters are highlighted in green and red. Reproduced with permission from Ref. [93]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The whole spectra from portable instrumentation were also used with clustering techniques. A portable XRF spectrometer and a IR camera were used in the study of two Renaissance paintings from the Museo e Real Bosco di Capodimonte in Naples (Italy) [98]. Based on the results obtained from clustering using *k*-means in conjunction with other techniques like PCA, researchers could propose the formulation of the preparatory layer and identify the pigments used. Another on-site study using portable instrumentation consisted on measuring with a portable Raman instrument the stained glass windows in the upper chapel of the Sainte-Chapelle (Paris) [99]. Although the instrument was portable, it was not a commercial one but an assemble of the different parts of a Raman spectrometer. The Raman signature of the glasses, together with different chemometric approaches (PCA and hierarchical cluster analysis), made it possible to obtain information about the relative age of weathered glass.

3.3. Regression modelling

Regression modelling aims to find the relationship between one (or several) dependent variables or **Y**-variables and the predictor variables (independent variables or **X**-variables). Most of the cases deal with only one dependent variable that typically corresponds to concentration values. The simplest form of regression is univariate regression [100] in which, in the context of this paper, the only instrumental signal which is the independent variable **x** (e.g. the intensity of the XRF line for a specific element) is related to the concentration of that specific element (dependent variable **y**) following equation (4):

$$\mathbf{y} = b_0 + b_1\mathbf{x} + \mathbf{e} \qquad \text{Eq. 4}$$

being b_0 and b_1 respectively the intercept and the slope of the regression line and **e** the error term. The criterion that is typically used to find the

coefficients b_0 and b_1 is the least squares criterion [57].

As we have commented on the previous sections, most of the data derived from artworks are of multivariate nature because most of the chemical measurements are inherently multivariate, and therefore the regression model is expressed as [101,102]:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad \text{Eq. 5}$$

being **B** the matrix of the regression coefficients and **E** represents the error term in the model. Most of the regression methods that are normally used assume a linear dependence between **X** and **Y**.

The least squares criterion used to find the coefficients in equation (4) may be generalized to the multivariate case [103], finding therefore as many regression coefficients as the number of variables plus one (the intercept term). This regression technique is referred to as multiple linear regression (MLR), but it is rarely utilized today. This is primarily due to its limited effectiveness when dealing with correlated variables and the requirement for more samples than variables, which can be challenging to fulfil, especially with modern instrumentation capable of generating a large number of variables. To overcome these drawbacks, the regression methods most frequently used are methods based on the reduction of the dimensionality of the original space into a lower-dimensional space while preserving all the relevant information. A first approach in this direction is the principal component regression (PCR) method. PCR works by first applying PCA to the independent variables to find the associated principal components. The regression is then performed between the principal components and the independent variable. However, partial least squares (PLS) [104] takes a step further. It not only reduces data dimensionality like PCR but also considers the relationship between the independent variables and the dependent variable, ensuring that PLS captures not only the variance in the data but also the covariance with the dependent variable. This makes PLS more effective when the variables are highly interrelated and significantly improves the regression model compared to PCR, as it seeks components that are directly relevant to the prediction task. These components are therefore different from the 'principal components' found in PCA or PCR and are called 'factors' or 'latent variables'.

While PLS and PCR are regression methods that assume linearity between dependent and independent variables, there are regression techniques designed to handle non-linearity, but these techniques are much less used in regression using chemical data from cultural heritage. Some of these methods, such as support vector regression (SVR) can capture complex non-linear patterns by mapping data into a higher-dimensional space, or techniques like decision trees, random forests, and neural networks are capable of handling non-linear relationships in data, making them suitable choices when linear assumptions do not hold [105–109].

Two key parameters are pivotal to ensure the validity of a multivariate calibration model: the root mean squared error (RMSE) and the coefficient of determination (R^2) between the predicted outcomes of the model and the reference results (the values from the **Y** matrix), although other parameters are also used [110]. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n}} \quad \text{Eq. 6}$$

where y_i is the ith sample of the **y** vector, $\widehat{y_i}$ is the predicted value of the ith sample using the regression model, and n is the total number of samples. The RMSE can refer to the calibration data, in which case it is written as RMSEC (root mean squared error of calibration) and is a measure indicating the quality of the model in terms of its ability to fit the calibration data, or to the validation data, thus samples not used in the construction of the model [111]. Validation data can be obtained using cross-validation (CV), which involves different iterations with the calibration data where, at each step, a group of samples (or a single sample) is left out, a multivariate model is built with the remaining samples, and the parameter of interest for that group of samples is predicted. In this way one can compute the RMSECV (root mean squared error of cross-validation). Validation can also be performed using an external data set not used in the construction of the model, and RMSEP (root mean squared error of prediction) can be calculated in this way. These validation measures collectively ensure the robustness and the predictive capability of the multivariate calibration model, offering a comprehensive assessment of its performance and generalization to new samples. Selecting the appropriate number of factors for a PLS model is central to prevent underfitting or overfitting. Underfitting arises when the model lacks the complexity to grasp the inherent patterns within the data, resulting in an incomplete representation of the connections between input variables and the target variable, thus causing inadequate predictive performance. Overfitting occurs when a model is too complex, capturing random data fluctuations instead of true patterns. An overfitted model excels on training data but struggles with new data since it essentially memorizes the training set without fully grasping the genuine relationships. In other words, random variability has been included in the model, causing it to fail in predicting new samples. It is crucial to strike the right balance between the chosen number of factors and the prediction error of the model (to delve deeper into the topic, we recommend these readings [111–115]), and very recently a new joint parameter (J-Score) for the selection of the correct number of factors and the best pre-processing technique in spectroscopic data has been proposed [116]. It is important to note that in a multivariate regression model, the prediction error relies on the error of the reference method, which supplies the values for the **Y**-matrix during calibration. Understanding the figures of merit of the reference method is crucial to comprehend the potential constraints of the regression model constructed from the same data.

Regression modelling is typically used in combination with explorative techniques, especially PCA. Regression methods applied to cultural heritage using chemical data have been applied in numerous studies, and in the vast majority of the cases, PLS was the regression method used. Most of the reviewed applications correspond to cases that use data from analytical techniques with autocorrelated variables (i.e., spectra). Among these applications, the vast majority come from studies in the 21st century, making regression modelling a more recent field of application to cultural heritage compared to other chemometric techniques. Starting with data from analytical techniques that use independent variables, particle-induced X-ray emission (PIXE) was used to analyse ancient papyrus and to make a PLS model between the elemental data from the PIXE analysis and the brightness from pixels in the papyrus, being able to predict missing characters of the text [117]. AAS and ICP-OES were used in the analysis of historical mortars, with the PLS model being able to correctly predict the binder/aggregate ration in historical mortars from churches in Milan (Italy) [118].

Focussing on some applications that use regression modelling in data from analytical techniques using autocorrelated variables, IR spectroscopies are the most used instrumental techniques. ATR-IR with the help of experimental design was applied to the determination of the superficial pigments composition of a painting, being able to predict with three different PLS models (one model for each specific pigment) the concentration of the pigments in different mixtures [119], and different parameters of historical paper (such as the age [120], different mechanical properties [121] and gelatine [122]) were predicted using PLS models and data from infrared spectroscopy (FTIR and NIR) using commercial equipment and even developing a NIR prototype instrument [121]. Fig. 6 shows the NIR prototype instrument used for the determination of mechanical properties of historical paper [121]. Fig. 7 shows the modelled results of the content of gelatine in historic papers determined with a PLS model using FTIR data versus the measured reference gelatine values obtained using a chromatographic method [122].

In a similar way, FTNIR spectra were used to determine the tung oil

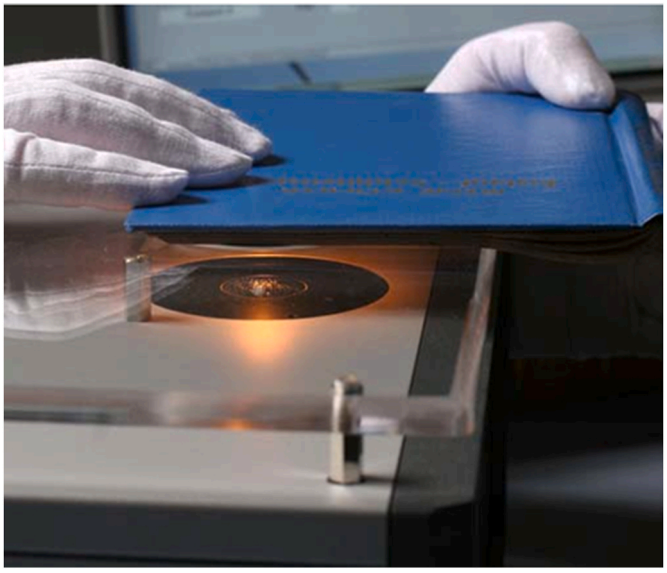


Fig. 6. The NIR instrument prototype. The table made of Perspex is positioned at a fixed distance from the measurement aperture, so that there is no direct contact between an item and the measurement opening. Reproduced with permission from Ref. [121].

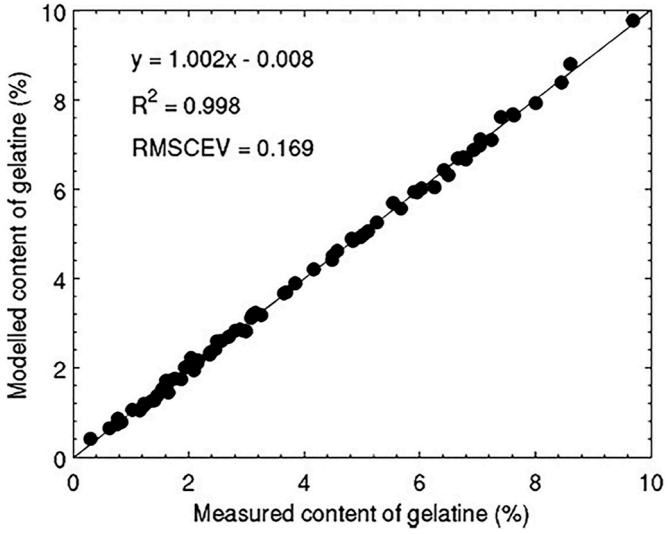


Fig. 7. Predicted versus measured results of percentage of gelatine in historic papers calculated with a PLS model. FTIR data were used for the construction of the model. Reproduced with permission from Ref. [122].

content in antique Chinese lacquerware (Fig. 8) with a PLS model, since the proportions of tung oil in various artworks are still unclear and subject to dispute [123].

Not only IR spectra have been used in this field, and for instance spectra from UV–vis reflectance spectroscopy was used to build a PLS model for the determination of Tyrian purple (a dye used in the antiquity) in archaeological textile fragments dating from the Roman period, relating the spectra with the reference results obtained using HPLC [124]. In a similar way, LIBS was used in the analysis of antique bronze coins from the Roman empire to find the content of copper in the coins (copper was used because it is a major constituent of the coins; concentrations of other elements were small as compared to copper) using a PLS model [125]. Electron probe X-ray microanalysis and micro XRF were used to build different PLS models to determine six elements from 16 to 17th century archaeological glass samples being able to determine the concentrations of the major oxides in glass with adequate errors [126].

Portable instrumentation has also been applied in the recent years in combination with regression modelling, taking advantage of the flexibility to conduct on-site analyses and reducing the need for sample transportation. To review some applications in the field, for instance a portable ATR-FTIR spectrometer in combination with PLS regression was used to monitor the degradation of plastics used in modern and contemporary art, building a robust degradation model of each analysed material that can be used to predict and classify the degradation state of artworks and to check priorities of intervention in the museum collections [127]. A portable NIR spectrometer was successfully used to quantitatively determine with a PLS model the proportion of oil to lacquer in ancient Chinese lacquer objects (a proportion that is crucial to discern the quality of the lacquer objects) [128]. Different portable XRF spectrometers together with PLS modelling were used to predict with moderate accuracy different elements on a Viking site in Denmark (suggesting the use of more cost-efficient XRF data than ICP-MS data) [129] and to analyse the surface layer of different artworks belonging to a museum collection in Caserta (Italy) to study different gilding techniques, finding interesting results such as the composition of the surfaces and the presence of porporina due to bad restorations [95].

3.4. Classification and discrimination

Classification techniques [130] are employed to categorize objects or samples into one or more predefined classes or categories based on a set of inherent characteristics (chemical measurements in the framework of this paper, which, as we pointed out in section 2.1, are mainly of multivariate nature). In the context of cultural heritage, this may involve for instance categorizing artifacts by historical periods, identifying artistic styles, or grouping materials used in archaeological objects. Some examples of classification in the field of cultural heritage with the help of chemical data is the classification of Hungarian medieval silver coins according to historical periods using elemental analysis from XRF data [131] or the classification of Neolithic potteries of 6th millennium

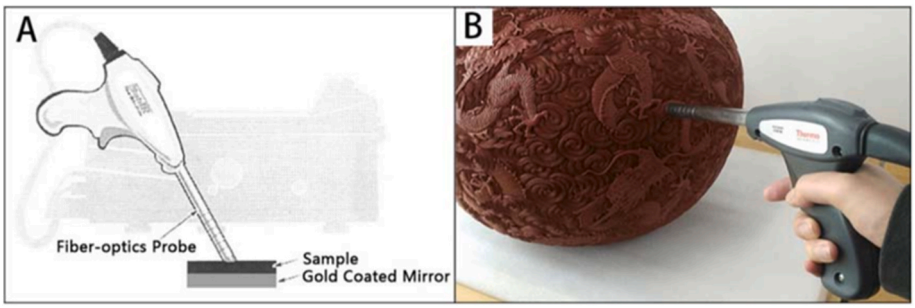


Fig. 8. FTNIR instrument used in the collection of spectra of ancient Chinese artwork. Collecting the spectra from a standard sample (A) and an historic object (B). Reproduced with permission from Ref. [123].

BC according to the excavation levels using FTIR and Raman data [132].

Multivariate data from chemical measurements are typically organized in the **X** matrix and the **Y** matrix contains the assigned classes (the information of the classes may be expressed as integer numbers or as text, depending on the specific classification method and algorithm used).

Classification techniques are supervised [133], meaning that they require prior knowledge of classes or categories in which the samples are divided. The new unknown samples can be classified into the predefined classes or categories. This contrasts with techniques like PCA, which is an unsupervised technique (there is no prior knowledge of classes or categories of the samples) and is primarily used for exploring data patterns and reducing dimensionality without relying on predefined categories.

Classification techniques are mainly divided in two groups of techniques: modelling (or class-modelling) and discriminant techniques. On one hand, modelling techniques focus on calculating individual models for each predefined class, capturing the specific characteristics and variations associated with each class. In these techniques, each class is modelled individually and independently from the other classes. On the other hand, discriminant techniques aim to find the most effective combination of variables that maximally separates the predefined classes. Discriminant techniques rely on the differences between samples from different classes and result in hypersurfaces (multidimensional surfaces), dividing the space variable in as many regions as the number of classes. An important consequence of this distinction among the two groups of techniques is that when a discriminant technique is used, one sample is always classified into one (and only one) of the available classes, while using a modelling technique, one sample may be classified into none, one, or more than one of the available classes. These differences can be seen in Fig. 9 that illustrates a simple example with the different approaches to the classification of three classes in a two-dimensional space (two variables).

Among the modelling techniques [135], the most used is soft independent modelling of class analogies (SIMCA) [136]. SIMCA finds a PCA model for each pre-defined category or class and two basic statistics (score and orthogonal distance of the sample to each PCA model) are used to decide if a specific sample belongs to a class. A distance (d) from the kth sample to the Ath PCA model can be calculated as:

$$d_{k,A} = \sqrt{OD_{k,A}^2 + SD_{k,A}^2} \quad \text{Eq. 7}$$

where $OD_{k,A}$ is the orthogonal distance from the kth sample to the Ath PCA model and $SD_{k,A}$ is the score distance of the kth sample in the Ath PCA model. The $d_{k,A}$ value in Eq. (7) is then compared with a threshold distance to the model of class A. The decision rule for classifying a sample into a PCA model varies among the different versions of SIMCA implementations [137]. Other modelling techniques, but much less used than SIMCA, are for instance unequal-class modelling (UNEQ) [138] or potential function techniques [134].

Among the discriminant techniques [134], one of the most used ones is partial least squares – discriminant analysis (PLS-DA) [139]. In PLS-DA, a PLS regression model between the independent variables in the **X** matrix and a vector **y** containing the assigned classes as integer numbers is calculated. Typically, 0 and 1 are used as integer numbers for the two assigned classes. In case of having more than two classes, the PLS2 version of PLS, able to cope with several dependent variables, may be used, and the technique is therefore called PLS2-DA [140]. In this case, there are as many columns in **Y** as you have defined classes. Each column is then a dummy vector of 0s and 1s, with 1 indicating the membership of that specific class, and 0 that it is not a member of that specific class. A threshold or discriminant line, between 0 and 1, is established to separate two classes, and a sample is classified using the projected value of the PLS model, which is a real number rather than an integer. The sample is classified to class 1 if the prediction is larger than the threshold or assigned to class 0 if the prediction is lower than the threshold.

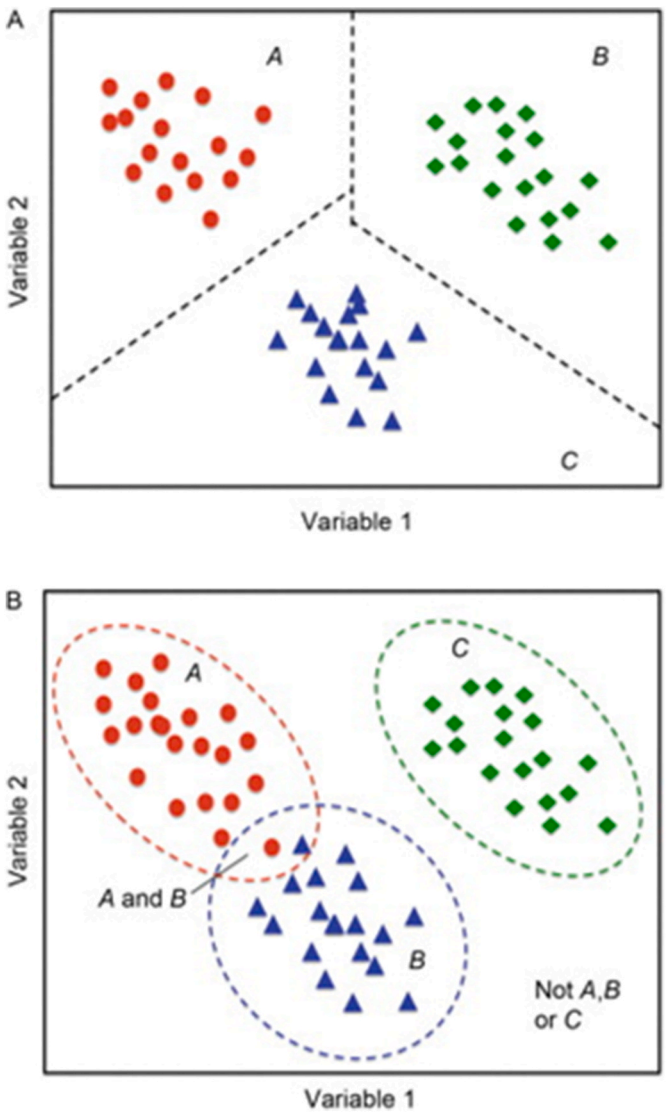


Fig. 9. Classification using (A) discriminant and (B) modelling techniques for a problem involving three classes in a two-dimensional space. (A) Discriminant approach divides the space into three non-overlapping regions corresponding to the different categories so that a sample is univocally assigned. (B) Class-modelling defines an individual model for the different categories, so that regions where a sample is classified into more than one category or into none can occur. Reproduced with permission from Ref. [134].

Another used discriminant technique is linear discriminant analysis (LDA) [134]. LDA aims to identify the linear combinations of variables (linear discriminant functions) that best discriminate between predefined classes in a dataset. The primary goal of LDA is to maximize the separation between classes while minimizing the variation within each class. A sophistication of LDA is quadratic discriminant analysis (QDA), in which quadratic decision boundaries are modelled, making it suitable for datasets where within-class variation varies across classes.

k nearest neighbours (kNN) is the simplest discriminant technique, but despite of its simplicity, is one of the least used. In this technique, the class is determined by the majority class among its k nearest neighbours. To calculate the distance between samples, typically the Euclidean distance is employed. A crucial parameter is the selection of k, the number of nearest neighbours, and usually small values of k (3 or 5) produce better classification results [134]. Despite the mathematical simplicity of the technique, it has shown to work well in many real cases, except in

those cases where there is a clear difference in the number of objects in the different classes or categories.

Finally, other discriminant classification techniques, although much less used in cultural heritage, are for instance support vector machine (SVM)-based methods for classification [141,142] or artificial neural networks (ANN) [143].

Validation is a crucial step to ensure that the model generalizes well to unseen data and performs reliably in real-world scenarios. Cross-validation or external test set can be used for this purpose as in regression models. It is worthwhile to note that when dealing with imbalanced datasets, it's essential to ensure that each class is represented proportionally in both the training and validation sets. The topic of validation has been widely covered in the literature and for instance the reader can look at reference [111]. Regardless of the classification technique used, several parameters are used to assess the classification performance. The most widely used are sensitivity (the ability to correctly identify samples belonging to a specific class), specificity (the ability to correctly identify samples that do not belong to a specific class) and precision (the ability to avoid wrong predictions in one class) [144], but other parameters such as accuracy, negative predictive value (NPV), false positive rate (FPR) and false negative rate (FNR) may also be used [145], although less frequently. As with regression modelling, classification is typically used in combination with exploratory techniques, especially PCA. Classification techniques applied to cultural heritage using multivariate chemical data have been more prevalent in studies compared to regression modelling. This is likely because classification represents one of the fundamental chemometric challenges within the realm of cultural heritage analysis. Among these techniques, SIMCA and PLS-DA are the most utilized, although others have also been employed. Most of the reviewed applications using data from analytical techniques with autocorrelated variables (i.e., spectra), are from studies of the 21st century, while the reviewed applications using data with non-correlated variables extend further back in time to the late 1970s.

Beginning with data derived from analytical techniques utilizing independent variables, chemometric methods are frequently employed in the investigation of Roman archaeological remains. For instance, in the analysis of different Roman glasses discovered in Norway, AAS and XRF were employed, and SIMCA was utilized to categorize the samples based on different decoration classes. [146], Seven major and minor elements were analysed in Roman pottery using ICP-AES and AAS, and SIMCA was used to discriminate objects produced in different geographical areas [147], 86 Roman amphorae sherds from a shipwreck were analysed using NAA in order to classify the provenance of the amphora [148] and 160 amphorae dating to the 5th century AD were analysed by XRF and SIMCA and artificial neural networks were employed to classify their geographical provenance with good classification abilities [149]. Pottery and ceramics are other frequently used materials in which chemometrics is applied for classification purposes. In this way, ancient Mesopotamian ceramics and clay were analysed using INAA and SIMCA was used to identify pottery workshops in the Sumerian society [150]. Archaeological pottery from Cyprus was also analysed with INAA and LDA was used to identify clay types used and not used in the manufacturing processes [151]. Italian archaeological pottery ranging in date from the mid-7th century BC to the beginning of the 3rd century BC was analysed by AAS and SIMCA was applied to find that most of the sherds came from local kilns, with a probable Ionian origin for some other sherds and some other ones from unknown origin [152].As the last example, ceramic samples from the Banda traditional area (west-central Ghana) were analysed with NAA and SIMCA in order to determine with good classification performances the geographical origin of the samples [153].

Portable instrumentation in analytical techniques that use independent variables has also been applied in the recent years in combination with classification. To name some applications, a portable XRF spectrometer (using only the information about the composition of the elements, not the whole spectra) was used to analyse pottery samples from an archaeological site in Mexico and PLS-DA was used to find differences among the detected classes, which can be interpreted as different manufacturing processes [94] (Fig. 10 shows the results of the PLS-DA model assigning pottery samples to a pre-defined class). Another portable XRF sensor (again using only the information about the composition of the elements) was used to analyse 42 elements in granites in heritage buildings. SIMCA and Naïve Bayes classification (a set of supervised learning algorithms based on applying the theorem of Bayes) were used to find the most similar quarry rocks for appropriate restoration of historical buildings [154].

Focussing on applications that use classification in data from analytical techniques using autocorrelated variables, IR spectroscopies are again the most used instrumental techniques. In this way, FTIR was used in a variety of works such as the analysis of pre-Roman ceramics from 75 shards excavated in Puglia (Italy) where SIMCA was used to classify the samples in different classes according to the mineralogical composition [155]. It was also used in the analysis of polymeric resins where the use of LDA allowed the identification of the main components in commercial varnishes employed for art purposes [156]. IR spectroscopy was also employed in the analysis of ancient Persian wall paintings where SIMCA was used to classify the binding media of unknown samples [157], in the analysis of paper relic where different classification techniques (PLS-DA, SVM, LDA and SIMCA) successfully classified the different types of paper [158], and in the analysis of archaeological amber where LDA was able to predict the provenance of the amber from a historical tomb in China [159]. Raman spectroscopy has also been used in several works. For instance, pigments used in the field of cultural heritage were analysed with an instrument prototype combining micro-Raman and XRF and PLS-DA was successfully applied in the classification of pigments [160,161], lipidic paint binders and PLS-DA yielded satisfactory results of the different types of binders studied [162], and data from commercial artist paints were analysed using LDA to identify the crystalline structure of an important pigment in 20th century artworks, what may retrieve information on the production process of the pigment at the moment [163]. Other techniques using correlated variables have also been used in cultural heritage samples. For instance, GC and MS were used in the analysis of the organic patinas of old buildings in towns and villages in the Apennines (Italy) and PLS-DA was used to find a relationship between the age, history, environment and anthropic location of the monuments [164]. MALDI-TOF mass spectrometry (matrix-assisted laser desorption/ionization-time of flight - mass spectrometry) was used in the analysis of protein binders in

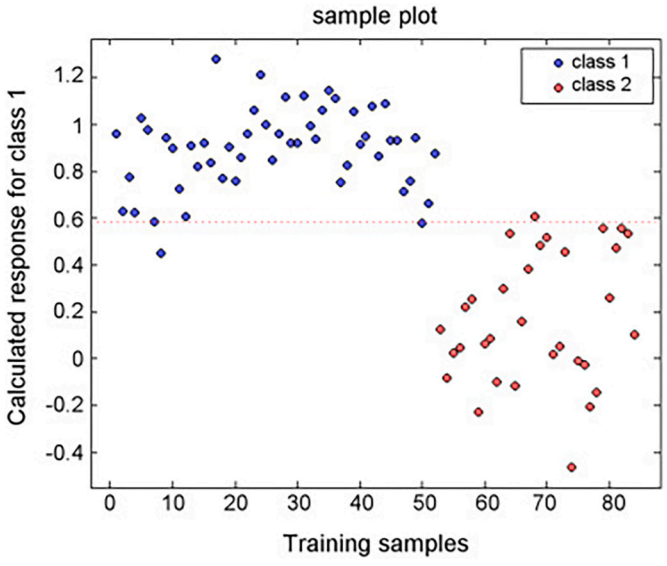


Fig. 10. Results of the PLS-DA assigning pottery samples to a pre-defined class. Reproduced with permission from Ref. [94].

historical paints and SIMCA helped in the classification of the binding medium of the painted paper background from the altarpiece of a 16th century church in Slovakia [165]. Botanical source of starch in ancient manuscripts was analysed by micellar electrokinetic capillary chromatography method with UV diode-array detection (MEKC–DAD) and LDA was used to determine the botanical source of the starch [166]. Direct analysis in real time mass spectrometry (DART-MS) was used for the non-invasive analysis of the Dead Sea scrolls and LDA was employed to investigate two different parchment treatments [167]. DART-MS was also used to analyse archaeological wood and PLS-DA was used to distinguish between three classes of wood (severely decayed archaeological wood, moderately decayed archaeological wood and recent wood) [96]. Finally, a LIBS prototype was used to analyse medieval ceramics (Islamic and Byzantine from different mounds) and PLS-DA was used to identify different sample classes [168].

Portable instrumentation in analytical techniques that use dependent variables has also been applied in the recent years in combination with classification. For instance, a portable LIBS instrumentation was used in the analysis of wall paintings with a reference database of commercial pigments traditionally used in murals and oil techniques following ancient recipes, and SIMCA and PLS-DA were used for the classification of pigments [169]. A portable NIR spectrometer was used to analyse Scandinavian Stone Age rock paintings and SIMCA was used to find that some painted elks could be separated from each other while others were similar (helping to answer questions about origin, age and weathering) and PLS-DA helped in discriminating the background from the paints [170]. Wall paintings from a Roman basilica were analysed using a variety of techniques (Raman, XRF, time gated laser induced fluorescence spectroscopy -TG-LIF, UV–Vis–NIR multispectral imaging, FTIR, VIS/NIR/SWIR) and PLS-DA was used to identify, among others, the original pigments palette used by the artist [171].

3.5. Data pre-processing

Using autocorrelated and non-correlated variables also has a significant impact in the data pre-processing before the chemometric analysis. In non-correlated variables, data pre-processing is typically applied to the columns, to the variables of the data matrix, to normalize the raw analytical data to avoid wrong conclusions due to the different order of magnitude and range of variation of the data. This is the case for instance when dealing with analysis from major and minor components or when different methods of analysis have been used and one is dealing with analysis with different units (e.g. elements in concentration units and elements in percentage values). Without normalization, techniques that use the reduction of dimensionality would take the directions of the original variables having the highest values, which do not necessarily represent the best data variability. Normalization ensures that all variables carry equal weight in the analysis, preventing the different techniques from being biased towards variables with larger numerical values. The most widely used pre-processing for column normalization is autoscaling, which scales the concentration of each variable to zero mean value and unit variance, although other techniques such as Pareto scaling, vast scaling or range scaling are also used [172].

When using data with autocorrelated variables, data pre-processing is typically applied to the rows, to the samples of the data matrix. Spectroscopic (including UV, NIR, MIR, FTIR or Raman) data are typical examples, but other examples are chromatographic or fluorescence (mainly X-ray fluorescence) data. Different sources of unwanted variation may affect in a different way to the samples, so applying the data pre-processing to the rows one tries to reduce these unwanted effects highlighting only the relevant variation among the samples. The data can be significantly influenced by a variety of processes, depending on the specific part of the electromagnetic spectrum used [172] and also on the specific measurement mode [173]. Some sources of variation may arise from inherent limitations of the instrument (e.g. instrument drifts or noise in virtually all the reviewed instrumental techniques), from variations among samples (e.g. particle size or homogeneity level affecting light scattering in NIR data), from features intrinsic to the measurement mode (e.g. the presence of slope in ATR-FTIR spectra since the reflectance is getting lower as the wavenumber decreases), or from the need to correct the peak alignment to reduce the shift effect of peaks in chromatographic or XRF data. Experimental conditions can generate variations in raw data, hence the importance of studying all sources of variability associated with the analytical procedure itself (see as example [174,175]). Different pre-processing methods are used, among others, to remove or decrease the instrumental noise (smoothing techniques) [172,176,177], to remove the offset baseline or the slope (detrending, first and second order derivatives) [67,172,176], to resolve overlapped peaks (second derivative) [178], to remove light scattering effects (standard normal variate, SNV, or multiplicative scatter correction, MSC) [67,172] or specific algorithms to correctly align peaks [179].

4. Practical considerations

Many researchers ponder the following question: how can one construct a proficient model?

However, the initial inquiry that should come to mind when considering the application of chemometrics for data examination is whether the samples being assessed accurately represent the subject of study. Regarding cultural heritage, the sample size is frequently constrained, as there are instances where it is not feasible to obtain as many samples as desired.

We need to consider that models, whether they are qualitative or quantitative, aim to elucidate the data provided to them for explanation. When the samples collected and examined accurately represent the context under investigation, multivariate analysis allows for the drawing of conclusions that can be extended to the entire context. Conversely, when the samples fail to genuinely depict the context, the results obtained from their study through multivariate analysis remain restricted to that specific subset of samples. This holds true in most experiments, and it is particularly relevant to the multivariate analysis techniques examined in this article. In fact, these methods aim to capture overarching patterns within the analysed samples in relation to the selected study variables.

Frequently, it is claimed that multivariate analysis requires a substantial sample size. Consequently, researchers sometimes resort to creating duplicates (replicates) of their samples in an effort to augment the sample size. Undoubtedly, an increased sample size enables the model to better capture trends, groups, clusters, or predict relevant properties within the data or find patterns for explorative purposes. Actually, a larger sample size introduces more variability into the model. However, it is important to note that replicate samples do not contribute to this type of information. Replicates, i.e. replicate analyses on the same sample, hold significant importance in capturing the inherent variability within each sample analysed with a certain type of measurement. Their significance becomes more pronounced when dealing with highly inhomogeneous samples. In such cases, using the mean value in the model could be advisable, as it enables each sample to be represented by a single row in the data matrix, providing a more comprehensive description of individual samples.

In regression and classification models, the sample size plays a significant role as it determines the feasibility of validation methods that can be applied to the model. These models should be constructed using one set of samples (calibration set) and subsequently validated using a separate set of samples that is independent from the initial set (test set). In cases where the number of samples is extremely limited, this approach is impractical, and as a result, cross-validation becomes necessary. Cross-validation serves as a reliable estimate of how a model operates but it should not be regarded as a conclusive validation method (for a more in-depth exploration of the subject of model validation, we recommend reading [111,180]). It is important to keep in mind that

when implementing variable selection systems in PLS models, the optimization process should be carried out exclusively on the calibration set. Subsequently, the selected variables should be tested on the validation sets. Once more, choosing variables from a limited number of samples may result in the selection of variables that are overly influenced by those specific samples, making them potentially unsuitable for future samples (to delve deeper into this topic, it is recommended to read [181–185]).

When discussing data visualization models, particularly PCA, or regression models like PLS, it is crucial to bear in mind that their operation relies on discerning differences among samples. Exceptionally distinct and unique samples have the potential to significantly distort the model, even altering the percentage of the explained variance in models with and without these unique samples. So, it is essential to exercise caution to ensure that this does not undermine the quest for the overarching trends within the samples being studied. In such a scenario, it might be advantageous to consider, following the initial model, the removal of the atypical sample or samples [186]. Subsequently, one can observe whether the PCA model is better able to account for other sources of variability within the dataset, or if the PLS is more efficient in predicting the interested property.

5. Conclusions

Analytical chemistry and chemometric data analysis significantly bolsters the study and conservation of cultural heritage. Both fields are almost inseparable today, implying that one significantly benefits from the other. This review does not aim to encompass all the literature on the use of chemometric techniques in cultural heritage analysis. Instead, it serves as motivation for those yet to employ these techniques in their data processing endeavours. The goal is to offer the necessary groundwork for cultural heritage scientists to embark on utilizing these techniques in their own studies, which often entail multivariate data. This involves grasping the fundamentals of the most utilized methods and drawing inspiration from existing applications documented in the literature. Apart from the cited applications, numerous methodological articles have been referenced in this review. It is important to note that while chemometrics utilizes specialized software, it is a misconception to assume that the software autonomously handles everything or makes their own decisions. The true expertise lies with the scientist in the field, who must possess a solid understanding of the fundamentals to effectively utilize the software, which primarily aids in the analytical process. The study of cultural heritage involves a significant interdisciplinary approach. While the cultural heritage scientist may rely on the expertise of a chemometrician for data processing, having a foundational understanding of this discipline proves beneficial for all involved. It enables better sampling and experiment design and a comprehensive comprehension of the strengths and limitations inherent in this approach, also employing a shared language that enhances the effectiveness and simplifies the sharing of research findings.

CRediT authorship contribution statement

Jordi Riu: Writing – original draft, Resources. **Barbara Giussani:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Data availability

No data was used for the research described in the article.

Acknowledgements

JR would like to acknowledge the financial support from MCIN/AEI/10.13039/501100011033/and from FEDER 'Una manera de hacer Europa' (project PID2022-136649OB-I00).

References

[1] L. Eriksson, J. Gottfries, T. Lundstedt, J. Trygg, Editorial, J. Chemom. 21 (2007) 397.
[2] S. Wold, Chemometrics; what do we mean with it, and what do we want from it? Chemometr. Intell. Lab. Syst. 30 (1995) 109–115, https://doi.org/10.1016/0169-7439(95)00042-9.
[3] P. Geladi, K. Esbensen, The start and early history of chemometrics: selected interviews. Part 1, J. Chemom. 4 (1990) 337–354.
[4] K. Esbensen, P. Geladi, The start and early history of chemometrics: selected interviews. Part 2, J. Chemom. 4 (1990) 389–412, https://doi.org/10.1002/cem.1180040604.
[5] A. Borman, New directions in analytical chemistry, Anal. Chem. 53 (1981).
[6] S. Wold, M. Sjöström, Chemometrics, present and future success, Chemometr. Intell. Lab. Syst. 44 (1998) 3–14, https://doi.org/10.1016/S0169-7439(98)00075-6.
[7] B.R. Kowalski, Chemometrics, Anal. Lett. 11 (1978) xi–xiii, https://doi.org/10.1080/00032717808059728.
[8] B.R. Kowalski, Chemometrics, Anal. Chem. 52 (1980).
[9] R.M. Belchamber, D. Betteridge, Y.T. Chow, T.J. Sly, A.P. Wade, The application of computers in chemometrics and analytical chemistry, Anal. Chim. Acta 150 (1983) 115–128, https://doi.org/10.1016/S0003-2670(00)85464-1.
[10] G. Kateman, Evolutions in chemometrics, Analyst 115 (1990) 487–493, https://doi.org/10.1039/AN9901500487.
[11] D. Brodnjak Voncina, Chemometrics in bioanalytical chemistry, Nov. Biotechnol. 9 (2009) 211, https://doi.org/10.1007/978-3-030-82381-8_26.
[12] N. Kumar, A. Bansal, G.S. Sarma, R.K. Rawal, Chemometrics tools used in analytical chemistry: an overview, Talanta 123 (2014) 186–199, https://doi.org/10.1016/j.talanta.2014.02.003.
[13] E. Metcalfe, S.J. Haswell, Journal of what ? SCIENTIFIC tools of the trade, Nature 335 (1988) 463.
[14] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J. M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools, Anal. Bioanal. Chem. 409 (2017) 5891–5899, https://doi.org/10.1007/s00216-017-0517-1.
[15] R.G. Brereton, J. Jansen, J. Lopes, F. Marini, A. Pomerantsev, O. Rodionova, J. M. Roger, B. Walczak, R. Tauler, Chemometrics in analytical chemistry—part II: modeling, validation, and applications, Anal. Bioanal. Chem. 410 (2018) 6691–6704, https://doi.org/10.1007/s00216-018-1283-4.
[16] S. Wold, Chemometrics, why, what and where to next? J. Pharm. Biomed. Anal. 9 (1991) 589–596, https://doi.org/10.1016/0731-7085(91)80183-A.
[17] G. Musumarra, M. Fichera, Chemometrics and cultural heritage, Chemometr. Intell. Lab. Syst. 44 (1998) 363–372, https://doi.org/10.1016/S0169-7439(98)00069-0.
[18] G. Visco, P. Avino, Employ of multivariate analysis and chemometrics in cultural heritage and environment fields, Environ. Sci. Pollut. Res. 24 (2017) 13863–13865, https://doi.org/10.1007/s11356-017-9205-0.
[19] J.M. Madariaga, Analytical chemistry in the field of cultural heritage, Anal. Methods 7 (2015) 4848–4876, https://doi.org/10.1039/c5ay00072f.
[20] M. Magdy, Analytical techniques for the preservation of cultural heritage: frontiers in knowledge and application, Crit. Rev. Anal. Chem. 52 (2022) 1171–1196, https://doi.org/10.1080/10408347.2020.1864717.
[21] M.D.L. de Castro, A. Jurado-López, The role of analytical chemists in the research on the cultural heritage, Talanta 205 (2019) 120106, https://doi.org/10.1016/j.talanta.2019.07.001.
[22] B. Borg, M. Dunn, A. Ang, C. Villis, The application of state-of-the-art technologies to support artwork conservation: literature review, J. Cult. Herit. 44 (2020) 239–259, https://doi.org/10.1016/j.culher.2020.02.010.
[23] C. Jones, C. Duffy, A. Gibson, M. Terras, Understanding multispectral imaging of cultural heritage: determining best practice in MSI analysis of historical artefacts, J. Cult. Herit. 45 (2020) 339–350, https://doi.org/10.1016/j.culher.2020.03.004.
[24] A. Filopoulou, S. Vlachou, S.C. Boyatzis, Fatty acids and their metal salts: a review of their infrared spectra in light of their presence in cultural heritage, Molecules 26 (2021), https://doi.org/10.3390/molecules26196005.
[25] D. Thickett, B. Pretzel, FTIR surface analysis for conservation, Herit. Sci. 8 (2020) 1–10, https://doi.org/10.1186/s40494-020-0349-8.
[26] G. Bitossi, R. Giorgi, M. Mauro, B. Salvadori, L. Dei, Spectroscopic techniques in cultural heritage conservation: a survey, Appl. Spectrosc. Rev. 40 (2005) 187–228, https://doi.org/10.1081/ASR-200054370.
[27] F. Rosi, L. Cartechini, D. Sali, C. Miliani, Recent trends in the application of fourier transform infrared (FT-IR) spectroscopy in Heritage Science: from non-

invasive FT-IR, in: L. Sabattini, I. Dorothé van der Werf (Eds.), Chem. Anal. Cult. Herit., De Gruyter, Berlin (Germany), 2020, pp. 121–150, https://doi.org/10.1515/psr-2018-0006.
[28] R.A. Crocombe, Portable spectroscopy, Appl. Spectrosc. 72 (2018) 1701–1751, https://doi.org/10.1177/0003702818809719.
[29] G.L. Liu, S.G. Kazarian, Recent advances and applications to cultural heritage using ATR-FTIR spectroscopy and ATR-FTIR spectroscopic imaging, Analyst 147 (2022) 1777–1797, https://doi.org/10.1039/d2an00005a.
[30] J. Striova, A.D. Fovo, R. Fontana, Reflectance imaging spectroscopy in heritage science, Riv. Del Nuovo Cim. 43 (2020) 515–566, https://doi.org/10.1007/s40766-020-00011-6.
[31] E. Catelli, G. Sciutto, S. Prati, M.V. Chavez Lozano, L. Gatti, F. Lugli, S. Silvestrini, S. Benazzi, E. Genorini, R. Mazzeo, A new miniaturised short-wave infrared (SWIR) spectrometer for on-site cultural heritage investigations, Talanta 218 (2020) 121112, https://doi.org/10.1016/j.talanta.2020.121112.
[32] T. Raicu, F. Zollo, L. Falchi, E. Barisoni, M. Piccolo, F.C. Izzo, Preliminary identification of mixtures of pigments using the paletteR package in R—the case of six paintings by andreina rosa (1924–2019) from the international gallery of modern art Ca' pesaro, venice, Heritage 6 (2023) 524–547, https://doi.org/10.3390/heritage6010028.
[33] B. Giussani, G. Gorla, J. Riu, Analytical chemistry strategies in the use of miniaturised NIR instruments: an overview, Crit. Rev. Anal. Chem. (2022), https://doi.org/10.1080/10408347.2022.2047607.
[34] A. Gibson, Medical imaging applied to heritage, Br. J. Radiol. 96 (2023).
[35] L. Rampazzi, V. Brunello, F.P. Campione, C. Corti, L. Geminiani, S. Recchia, M. Luraschi, Non-invasive identification of pigments in Japanese coloured photographs, Microchem. J. 157 (2020) 105017, https://doi.org/10.1016/j.microc.2020.105017.
[36] D. Chiriu, F.A. Pisu, P.C. Ricci, C.M. Carbonaro, Application of Raman spectroscopy to ancient materials: models and results from archaeometric analyses, Materials 13 (2020), https://doi.org/10.3390/MA13112456.
[37] M.C. Caggiani, P. Colomban, Raman microspectroscopy for Cultural Heritage studies, Phys. Sci. Rev. 3 (2019) 1–18, https://doi.org/10.1515/psr-2018-0007.
[38] A. Rousaki, P. Vandenabeele, In situ Raman spectroscopy for cultural heritage studies, J. Raman Spectrosc. 52 (2021) 2178–2189, https://doi.org/10.1002/jrs.6166.
[39] J. Jehlička, A. Culka, Critical evaluation of portable Raman spectrometers: from rock outcrops and planetary analogs to cultural heritage – a review, Anal. Chim. Acta 1209 (2022), https://doi.org/10.1016/j.aca.2021.339027.
[40] P. Silveira, T. Falcade, Applications of energy dispersive X-ray fluorescence technique in metallic cultural heritage studies, J. Cult. Herit. 57 (2022) 243–255, https://doi.org/10.1016/j.culher.2022.09.008.
[41] A. Botto, B. Campanella, S. Legnaioli, M. Lezzerini, G. Lorenzetti, S. Pagnotta, F. Poggialini, V. Palleschi, Applications of laser-induced breakdown spectroscopy in cultural heritage and archaeology: a critical review, J. Anal. At. Spectrom. 34 (2019) 81–103, https://doi.org/10.1039/c8ja00319j.
[42] V. Detalle, X. Bai, The assets of laser-induced breakdown spectroscopy (LIBS) for the future of heritage science, Spectrochim. Acta Part B At. Spectrosc. 191 (2022) 106407, https://doi.org/10.1016/j.sab.2022.106407.
[43] J.S. Cabral, C.R. Menegatti, G. Nicolodelli, Laser-induced breakdown spectroscopy in cementitious materials: a chronological review of cement and concrete from the last 20 years, TrAC - Trends Anal. Chem. 160 (2023) 116948, https://doi.org/10.1016/j.trac.2023.116948.
[44] B. Giussani, D. Monticelli, L. Rampazzi, Role of laser ablation-inductively coupled plasma-mass spectrometry in cultural heritage research: a review, Anal. Chim. Acta 635 (2009), https://doi.org/10.1016/j.aca.2008.12.040.
[45] S. Carter, R. Clough, A. Fisher, B. Gibson, B. Russell, Atomic spectrometry update: review of advances in the analysis of metals, chemicals and materials, Royal Soc. Chem. (2021), https://doi.org/10.1039/d1ja90049h.
[46] G. Festa, G. Romanelli, R. Senesi, L. Arcidiacono, C. Scatigno, S.F. Parker, M.P. M. Marques, C. Andreani, Neutrons for cultural heritage—techniques, sensors, and detection, Sensors 20 (2020) 502.
[47] C. Scatigno, G. Festa, Neutron imaging and learning algorithms: new perspectives in cultural heritage applications, J. Imag. 8 (2022), https://doi.org/10.3390/jimaging8100284.
[48] M. Shahid, J. Wertz, I. Degano, M. Aceto, M.I. Khan, A. Quye, Analytical methods for determination of anthraquinone dyes in historical textiles: a review, Anal. Chim. Acta 1083 (2019) 58–87, https://doi.org/10.1016/j.aca.2019.07.009.
[49] M.S. Zalaffi, N. Karimian, P. Ugo, Review—electrochemical and SERS sensors for cultural heritage diagnostics and conservation: recent advances and prospects, J. Electrochem. Soc. 167 (2020) 037548, https://doi.org/10.1149/1945-7111/ab67ac.
[50] A. Doménech-Carbó, Electrochemistry in archaeology and art conservation, Isr. J. Chem. 61 (2021) 113–119, https://doi.org/10.1002/ijch.202000056.
[51] M. Tomassetti, F. Marini, L. Campanella, A. Coppa, Study of modern or ancient collagen and human fossil bones from an archaeological site of middle Nile by thermal analysis and chemometrics, Microchem. J. 108 (2013) 7–13, https://doi.org/10.1016/j.microc.2012.11.006.
[52] S. Prati, F. Rosi, G. Sciutto, P. Oliveri, E. Catelli, C. Miliani, R. Mazzeo, Evaluation of the effect of different paint cross section preparation methods on the performances of Fourier transformed infrared microscopy in total reflection mode, Microchem. J. 110 (2013) 314–319, https://doi.org/10.1016/j.microc.2013.04.016.
[53] J.A. Remolá, M.S. Larrechi, F.X. Rius, Chemometric characterization of 5th century A.D. amphora-producing centres in the Mediterranean, Talanta 40 (1993) 1749–1757, https://doi.org/10.1016/0039-9140(93)80093-7.
[54] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37–52, https://doi.org/10.1016/0169-7439(87)80084-9.
[55] R.G. Brereton, Applied Chemometrics for Scientists, John Wiley & Sons Ltd., Chichester, UK, 2007.
[56] R.G. Brereton, Chemometrics. Data Driven Extraction for Science, second ed., John Wiley & Sons Ltd., Chichester, UK, 2018.
[57] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Berveke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 2005.
[58] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (2014) 2812–2831, https://doi.org/10.1039/c3ay41907j.
[59] M. Li Vigni, C. Durante, M. Cocchi, Exploratory data analysis, in: F. Marini (Ed.), Data Handl. Sci. Technol., Amsterdam, 2013, pp. 55–126, https://doi.org/10.1016/B978-0-444-59528-7.00003-X.
[60] I. Kuleff, R. Djingova, G. Djingov, Provenience study of medieval Bulgarian glasses by NAA and cluster analysis, Archaeometry 27 (1985) 185–193, https://doi.org/10.1111/j.1475-4754.1985.tb00361.x.
[61] F.X. Rius, M.S. Larrechi, C. Benet, E. Subias, D.L. Massart, A. Thielemans, The application of multivariate techniques to data from Spanish glass-making objects from the Roman Era, Anal. Chim. Acta 225 (1989) 69–81, https://doi.org/10.1016/S0003-2670(00)84594-8.
[62] P. Mirti, M. Aceto, M.C. Preacco Ancona, Campanian pottery from ancient Bruttium (Southern Italy): scientific analysis of local and imported products, Archaeometry 40 (1998) 311–329, https://doi.org/10.1111/j.1475-4754.1998.tb00840.x.
[63] G. Visco, E. Gregori, M. Tomassetti, L. Campanella, Probably counterfeit in Roman Imperial Age: pattern recognition helps diagnostic performed with inductive coupled plasma spectrometry and thermogravimetry analysis of a torso and a head of Roman Age marble statue, Microchem. J. 88 (2008) 210–217, https://doi.org/10.1016/j.microc.2007.11.013.
[64] C. Corti, L. Rampazzi, C. Ravedoni, B. Giussani, On the use of trace elements in ancient necropolis studies: overview and ICP-MS application to the case study of Valdaro site, Italy, Microchem. J. 110 (2013) 614–623, https://doi.org/10.1016/j.microc.2013.07.001.
[65] V. Lazic, M. Romani, L. Pronti, M. Angelucci, M. Cestelli-Guidi, M. Mangano, R. Fantoni, Identification of materials in oil paintings through studies of correlations and ratios between the element line intensities during LIBS stratigraphy, Spectrochim. Acta Part B At. Spectrosc. 201 (2023) 106601, https://doi.org/10.1016/j.sab.2022.106601.
[66] T. Vettor, V. Sautter, S. Pont, C. Harivel, L. Jolivet, I. Moretti, J.C. Moretti, Delos archaeological marbles: a preliminary geochemistry-based quarry provenance study, Archaeometry 63 (2021) 907–922, https://doi.org/10.1111/arcm.12655.
[67] E. Marengo, M.C. Liparota, E. Robotti, M. Bobba, Monitoring of paintings under exposure to UV light by ATR-FT-IR spectroscopy and multivariate control charts, Vib. Spectrosc. 40 (2006) 225–234, https://doi.org/10.1016/j.vibspec.2005.09.005.
[68] C. Mazzuca, L. Micheli, F. Marini, M. Bevilacqua, G. Bocchinfuso, G. Palleschi, A. Palleschi, Rheoreversible hydrogels in paper restoration processes: a versatile tool, Chem. Cent. J. 8 (2014) 10, https://doi.org/10.1186/1752-153X-8-10.
[69] A. Sarmiento, M. Pérez-Alonso, M. Olivares, K. Castro, I. Martínez-Arkarazo, L. A. Fernández, J.M. Madariaga, Classification and identification of organic binding media in artworks by means of Fourier transform infrared spectroscopy and principal component analysis, Anal. Bioanal. Chem. 399 (2011) 3601–3611, https://doi.org/10.1007/s00216-011-4677-0.
[70] G. Sciutto, P. Oliveri, S. Prati, E. Catelli, I. Bonacini, R. Mazzeo, A multivariate methodological workflow for the analysis of FTIR chemical mapping applied on historic paint stratigraphies, Int. J. Anal. Chem. 2017 (2017), https://doi.org/10.1155/2017/4938145.
[71] G. Sciutto, S. Prati, I. Bonacini, P. Oliveri, R. Mazzeo, FT-NIR microscopy: an advanced spectroscopic approach for the characterisation of paint cross-sections, Microchem. J. 112 (2014) 87–96, https://doi.org/10.1016/j.microc.2013.09.021.
[72] M. Maguregui, U. Knuutinen, K. Castro, J.M. Madariaga, Raman spectroscopy as a tool to diagnose the impact and conservation state of Pompeian second and fourth style wall paintings exposed to diverse environments (House of Marcus Lucretius), J. Raman Spectrosc. 41 (2010) 1400–1409, https://doi.org/10.1002/jrs.2671.
[73] E.M. Rodrigues, E. Hemmer, Trends in hyperspectral imaging: from environmental and health sensing to structure-property and nano-bio interaction studies, Anal. Bioanal. Chem. 414 (2022) 4269–4279, https://doi.org/10.1007/s00216-022-03959-y.
[74] H. Liang, Advances in multispectral and hyperspectral imaging for archaeology and art conservation, Appl. Phys. Mater. Sci. Process 106 (2012) 309–323, https://doi.org/10.1007/s00339-011-6689-1.
[75] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. Essential steps before image analysis, Chemometr. Intell. Lab. Syst. 117 (2012) 138–148, https://doi.org/10.1016/j.chemolab.2012.05.009.
[76] J.M. Amigo, Practical issues of hyperspectral imaging analysis of solid dosage forms, Anal. Bioanal. Chem. 398 (2010) 93–109, https://doi.org/10.1007/s00216-010-3828-z.
[77] G. Sciutto, P. Oliveri, S. Prati, M. Quaranta, S. Bersani, R. Mazzeo, An advanced multivariate approach for processing X-ray fluorescence spectral and hyperspectral data from non-invasive in situ analyses on painted surfaces, Anal. Chim. Acta 752 (2012) 30–38, https://doi.org/10.1016/j.aca.2012.09.035.
[78] G. Capobianco, M.P. Bracciale, D. Sali, F. Sbardella, P. Belloni, G. Bonifazi, S. Serranti, M.L. Santarelli, M. Cestelli Guidi, Chemometrics approach to FT-IR

hyperspectral imaging analysis of degradation products in artwork cross-section, Microchem. J. 132 (2017) 69–76, https://doi.org/10.1016/j.microc.2017.01.007.
[79] G. Sciutto, S. Legrand, E. Catelli, S. Prati, C. Malegori, P. Oliveri, K. Janssens, R. Mazzeo, Macroscopic mid-FTIR mapping and clustering-based automated data-reduction: an advanced diagnostic tool for in situ investigations of artworks, Talanta 209 (2020) 120575, https://doi.org/10.1016/j.talanta.2019.120575.
[80] G. Capobianco, C. Pelosi, G. Agresti, G. Bonifazi, U. Santamaria, S. Serranti, X-ray fluorescence investigation on yellow pigments based on lead, tin and antimony through the comparison between laboratory and portable instruments, J. Cult. Herit. 29 (2018) 19–29, https://doi.org/10.1016/j.culher.2017.09.002.
[81] V. Andrić, M. Gajić-Kvaščev, D.K. Crkvenjakov, M. Marić-Stojanović, S. Gadžurić, Evaluation of pattern recognition techniques for the attribution of cultural heritage objects based on the qualitative XRF data, Microchem. J. 167 (2021) 106267, https://doi.org/10.1016/j.microc.2021.106267.
[82] L. Geminiani, F.P. Campione, C. Canevali, C. Corti, B. Giussani, G. Gorla, M. Luraschi, S. Recchia, L. Rampazzi, Historical silk: a novel method to evaluate degumming with non-invasive infrared spectroscopy and spectral deconvolution, Materials 16 (2023), https://doi.org/10.3390/ma16051819.
[83] F. Modugno, E. Ribechini, M.P. Colombini, Chemical study of triterpenoid resinous materials in archaeological findings by means of direct exposure electron ionisation mass spectrometry and gas chromatography/mass spectrometry, Rapid Commun. Mass Spectrom. 20 (2006) 1787–1800, https://doi.org/10.1002/rcm.2507.
[84] F. Modugno, F. Di Gianvincenzo, I. Degano, I.D. van der Werf, I. Bonaduce, K. J. van den Berg, On the influence of relative humidity on the oxidation and hydrolysis of fresh and aged oil paints, Sci. Rep. 9 (2019) 5533, https://doi.org/10.1038/s41598-019-41893-9.
[85] R.G. Brereton, Chemometrics in analytical chemistry, Analyst 112 (1987) 1635–1657, https://doi.org/10.1039/an9871201635.
[86] R. Todeschini, D. Ballabio, V. Consonni, F. Grisoni, A new concept of higher-order similarity and the role of distance/similarity measures in local classification methods, Chemometr. Intell. Lab. Syst. 157 (2016) 50–57, https://doi.org/10.1016/j.chemolab.2016.06.013.
[87] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recogn. Lett. 31 (2010) 651–666, https://doi.org/10.1016/j.patrec.2009.09.011.
[88] P. Zerzucha, B. Walczak, Concept of (dis)similarity in data analysis, Trends Anal. Chem. 38 (2012) 116–128, https://doi.org/10.1016/j.trac.2012.05.005.
[89] G. Rauret, E. Casassas, F.X. Rius, M. Muñoz, Cluster analysis applied to spectrochemical data of European mediaeval stained glasses, Archaeometry 29 (1987) 240–249, https://doi.org/10.1111/j.1475-4754.1987.tb00417.x.
[90] V. Argyropoulos, A characterization of the compositional variations of roman samian pottery manufactured at the lezoux production centre, Archaeometry 37 (1995) 271–285, https://doi.org/10.1111/j.1475-4754.1995.tb00743.x.
[91] P. Fermo, E. Delnevo, M. Lasagni, S. Polla, M. de Vos, Application of chemical and chemometric analytical techniques to the study of ancient ceramics from Dougga (Tunisia), Microchem. J. 88 (2008) 150–159, https://doi.org/10.1016/j.microc.2007.11.012.
[92] W.M. Badawy, A.Y. Dmitriev, V.Y. Koval, V.S. Smirnova, O.E. Chepurchenko, V. V. Lobachev, M.O. Belova, A.M. Galushko, Formation of reference groups for archaeological pottery using neutron activation and multivariate statistical analyses, Archaeometry 64 (2022) 1377–1393, https://doi.org/10.1111/arcm.12793.
[93] S. Columbu, S. Carboni, S. Pagnotta, M. Lezzerini, S. Raneri, S. Legnaioli, V. Palleschi, A. Usai, Laser-Induced Breakdown Spectroscopy analysis of the limestone Nuragic statues from Mont'e Prama site (Sardinia, Italy), Spectrochim. Acta Part B At. Spectrosc. 149 (2018) 62–70, https://doi.org/10.1016/j.sab.2018.07.011.
[94] P. López-García, D. Argote-Espino, K. Fačevicová, Statistical processing of compositional data. The case of ceramic samples from the archaeological site of Xalasco, Tlaxcala, Mexico, J. Archaeol. Sci. Rep. 19 (2018) 100–114, https://doi.org/10.1016/j.jasrep.2018.02.023.
[95] J. Brocchieri, E. Scialla, A. Manzone, G.O. Graziano, A. D'Onofrio, C. Sabbarese, An analytical characterization of different gilding techniques on artworks from the Royal Palace (Caserta, Italy), J. Cult. Herit. 57 (2022) 213–225, https://doi.org/10.1016/j.culher.2022.08.014.
[96] R. Li, J. Guo, N. Macchioni, B. Pizzo, G. Xi, X. Tian, J. Chen, J. Sun, X. Jiang, J. Cao, Z. Zhang, Y. Yin, Characterisation of waterlogged archaeological wood from Nanhai No. 1 shipwreck by multidisciplinary diagnostic methods, J. Cult. Herit. 56 (2022) 25–35, https://doi.org/10.1016/j.culher.2022.05.004.
[97] Z. Chen, A. Gu, X. Zhang, Z. Zhang, Authentication and inference of seal stamps on Chinese traditional painting by using multivariate classification and near-infrared spectroscopy, Chemometr. Intell. Lab. Syst. 171 (2017) 226–233, https://doi.org/10.1016/j.chemolab.2017.10.017.
[98] E. Scialla, P. Improda, J. Brocchieri, M. Cardinali, A. Cerasuolo, A. Rullo, A. Zezza, C. Sabbarese, Study of 'cona degli ordini' by colantonio with IR and XRF analyses, Heritage 6 (2023) 1785–1803, https://doi.org/10.3390/heritage6020095.
[99] P. Colomban, A. Tournié, On-site Raman identification and dating of ancient/modern stained glasses at the Sainte-Chapelle, Paris, J. Cult. Herit. 8 (2007) 242–256, https://doi.org/10.1016/j.culher.2007.04.002.
[100] J.N. Miller, Basic statistical methods for analytical chemistry. Part 2. calibration and regression methods. A review, Analyst 116 (1991) 3–14, https://doi.org/10.1039/AN9911600003.
[101] R.G. Brereton, Introduction to multivariate calibration in analytical chemistry, Analyst 125 (2000) 2125–2154, https://doi.org/10.1039/b003805i.
[102] R. Bro, Multivariate calibration: what is in chemometrics for the analytical chemist? Anal. Chim. Acta 500 (2003) 185–194, https://doi.org/10.1016/S0003-2670(03)00681-0.
[103] F. Westad, M. Bevilacqua, F. Marini, Regression, in: F. Marini (Ed.), Data Handl. Sci. Technol., Elsevier B.V., Amsterdam, 2013, pp. 127–170, https://doi.org/10.1016/B978-0-444-59528-7.00004-1.
[104] S. Wold, M. Sjostrom, L. Eriksson, S. Sweden, PLS-regression, a basic tool of chemometrics, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130, https://doi.org/10.1016/S0169-7439(01)00155-1.
[105] A. Soleymani, H. Jahangir, M.L. Nehdi, Damage detection and monitoring in heritage masonry structures: systematic review, Construct. Build. Mater. 397 (2023) 132402, https://doi.org/10.1016/j.conbuildmat.2023.132402.
[106] M. Mishra, Machine learning techniques for structural health monitoring of heritage buildings: a state-of-the-art review and case studies, J. Cult. Herit. 47 (2021) 227–245, https://doi.org/10.1016/j.culher.2020.09.005.
[107] L. Liu, T. Miteva, G. Delnevo, S. Mirri, P. Walter, L. de Viguerie, E. Pouyet, Neural networks for hyperspectral imaging of historical paintings: a practical review, Sensors 23 (2023) 2419, https://doi.org/10.3390/s23052419.
[108] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Non-linear regression methods in NIRS quantitative analysis, Talanta 72 (2007) 28–42, https://doi.org/10.1016/j.talanta.2006.10.036.
[109] W. Ni, L. Nørgaard, M. Mørup, Non-linear calibration models for near infrared spectroscopy, Anal. Chim. Acta 813 (2014) 1–14, https://doi.org/10.1016/j.aca.2013.12.002.
[110] T. Fearn, Assessing Calibrations: SEP,RPD,RER,R2, NIR News, vol. 13, 2002, pp. 12–14.
[111] F. Westad, F. Marini, Validation of chemometric models - a tutorial, Anal. Chim. Acta 893 (2015) 14–24, https://doi.org/10.1016/j.aca.2015.06.056.
[112] P. Oliveri, C. Malegori, R. Simonetti, M. Casale, The impact of signal pre-processing on the final interpretation of analytical outcomes – a tutorial, Anal. Chim. Acta 1058 (2019) 9–17, https://doi.org/10.1016/j.aca.2018.10.055.
[113] M. Alewijn, H. van der Voet, S. van Ruth, Validation of multivariate classification methods using analytical fingerprints – concept and case study on organic feed for laying hens, J. Food Compos. Anal. 51 (2016) 15–23, https://doi.org/10.1016/j.jfca.2016.06.003.
[114] D. Pérez-Guaita, J. Kuligowski, B. Lendl, B.R. Wood, G. Quintás, Assessment of discriminant models in infrared imaging using constrained repeated random sampling – cross validation, Anal. Chim. Acta 1033 (2018) 156–164, https://doi.org/10.1016/j.aca.2018.05.019.
[115] K. Kjeldahl, R. Bro, Some common misunderstandings in chemometrics, J. Chemom. 24 (2010) 558–564, https://doi.org/10.1002/cem.1346.
[116] J. Ezenarro, D. Schorn-García, L. Aceña, M. Mestres, O. Busto, R. Boqué, J-Score: a new joint parameter for PLSR model performance evaluation of spectroscopic data, Chemometr. Intell. Lab. Syst. 240 (2023) 104883, https://doi.org/10.1016/j.chemolab.2023.104883.
[117] N.E.G. Lövestam, E. Swietlicki, PIXE analysis and imaging of papyrus documents, Nucl. Instrum. Methods Phys. Res. B. 45 (1990) 307–310, https://doi.org/10.1016/0168-583X(90)90841-H.
[118] L. Rampazzi, A. Pozzi, A. Sansonetti, L. Toniolo, B. Giussani, A chemometric approach to the characterisation of historical mortars, Cement Concr. Res. 36 (2006) 1108–1114, https://doi.org/10.1016/j.cemconres.2006.02.002.
[119] E. Marengo, M.C. Liparota, E. Robotti, M. Bobba, Multivariate calibration applied to the field of cultural heritage: analysis of the pigments on the surface of a painting, Anal. Chim. Acta 553 (2005) 111–122, https://doi.org/10.1016/j.aca.2005.07.061.
[120] T. Trafela, M. Strlič, J. Kolar, D.A. Lichtblau, M. Anders, D.P. Mencigar, B. Pihlar, Nondestructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation, Anal. Chem. 79 (2007) 6319–6323, https://doi.org/10.1021/ac070392t.
[121] D. Lichtblau, M. Strlič, T. Trafela, J. Kolar, M. Anders, Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics - a new instrument, Appl. Phys. Mater. Sci. Process 92 (2008) 191–195, https://doi.org/10.1007/s00339-008-4479-1.
[122] L. Cséfalvayová, M. Pelikan, I. Kralj Cigić, J. Kolar, M. Strli, Use of genetic algorithms with multivariate regression for determination of gelatine in historic papers based on FT-IR and NIR spectral data, Talanta 82 (2010) 1784–1790, https://doi.org/10.1016/j.talanta.2010.07.062.
[123] A. Gu, J.R. Min, N. Wang, Y. Lei, Study of tung oil content in ancient lacquer by noninvasive quantitative methods: near infrared and chemometrics, Stud. Conserv. 67 (2022) 373–380, https://doi.org/10.1080/00393630.2021.1945860.
[124] C. Clementi, W. Nowik, A. Romani, D. Cardon, M. Trojanowicz, A. Davantès, P. Chaminade, Towards a semiquantitative non invasive characterisation of Tyrian purple dye composition: convergence of UV-Visible reflectance spectroscopy and fast-high temperature-high performance liquid chromatography with photodiode array detection, Anal. Chim. Acta 926 (2016) 17–27, https://doi.org/10.1016/j.aca.2016.04.022.
[125] M.O. Bachler, M. Bišćan, Z. Kregar, I. Jelovica Badovinac, J. Dobrinić, S. Milošević, Analysis of antique bronze coins by Laser Induced Breakdown Spectroscopy and multivariate analysis, Spectrochim. Acta Part B At. Spectrosc. 123 (2016) 163–170, https://doi.org/10.1016/j.sab.2016.08.010.
[126] P. Lemberge, I. De Raedt, K.H. Janssens, F. Wei, P.J. Van Espen, Quantitative analysis of 16-17th century archaeological glass vessels using PLS regression of EPXMA and μ-XRF data, J. Chemom. 14 (2000) 751–763, https://doi.org/10.1002/1099-128X(200009/12)14:5/6<751::AID-CEM622>3.0.CO;2-D.
[127] M. Manfredi, E. Barberis, E. Marengo, Prediction and classification of the degradation state of plastic materials used in modern and contemporary art, Appl.

Phys. Mater. Sci. Process 123 (2017) 35, https://doi.org/10.1007/s00339-016-0663-x.
[128] Y. Fu, Q. Xiao, S. Zong, S. Wei, Characterization and quantitation study of ancient lacquer objects by NIR spectroscopy and THM-Py-GC/MS, J. Cult. Herit. 46 (2020) 95–101, https://doi.org/10.1016/j.culher.2020.06.015.
[129] P.L.K. Trant, S.M. Kristiansen, S.M. Sindbæk, Visible near-infrared spectroscopy as an aid for archaeological interpretation, Archaeol. Anthropol. Sci. 12 (2020) 280, https://doi.org/10.1007/s12520-020-01239-3.
[130] F. Marini, Classification methods in chemometrics, Curr. Anal. Chem. 6 (2010) 72–79, https://doi.org/10.2174/157341110790069592.
[131] A. Rácz, K. Héberger, R. Rajkó, J. Elek, Classification of Hungarian medieval silver coins using x-ray fluorescent spectroscopy and multivariate data analysis, Herit. Sci. 1 (2013) 2, https://doi.org/10.1186/2050-7445-1-2.
[132] S. Akyuz, F. Guliyev, S. Celik, A.E. Ozel, V. Alakbarov, Investigations of the Neolithic potteries of 6th millennium BC from Göytepe-Azerbaijan by vibrational spectroscopy and chemometric techniques, Vib. Spectrosc. 105 (2019) 102980, https://doi.org/10.1016/j.vibspec.2019.102980.
[133] R.G. Brereton, Pattern recognition in chemometrics, Chemometr. Intell. Lab. Syst. 149 (2015) 90–96, https://doi.org/10.1016/j.chemolab.2015.06.012.
[134] M. Bevilacqua, R. Bucci, A.D. Magrì, A.L. Magrì, R. Nescatelli, F. Marini, Classification and class-modelling, in: Data Handl. Sci. Technol., Elsevier, 2013, pp. 171–233, https://doi.org/10.1016/B978-0-444-59528-7.00005-3.
[135] M. Forina, P. Oliveri, S. Lanteri, M. Casale, Class-modeling techniques, classic and new, for old and new problems, Chemometr. Intell. Lab. Syst. 93 (2008) 132–148, https://doi.org/10.1016/J.CHEMOLAB.2008.05.003.
[136] S. Wold, Pattern recognition by means of disjoint principal components models, Pattern Recogn. 8 (1976) 127–139, https://doi.org/10.1016/0031-3203(76)90014-5.
[137] A.L. Pomerantsev, O.Y. Rodionova, Popular decision rules in SIMCA: critical review, J. Chemom. 34 (2020) e3250, https://doi.org/10.1002/cem.3250.
[138] M.P. Derde, D.L. Massart, UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution, Anal. Chim. Acta 184 (1986) 33–51, https://doi.org/10.1016/S0003-2670(00)86468-5.
[139] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (2013) 3790–3798, https://doi.org/10.1039/c3ay40582f.
[140] H. Yangming, H. Yue, S. Xiangzhong, G. Jingxian, X. Yanmei, M. Shungeng, Comparison of a novel PLS1-DA, traditional PLS2-DA and assigned PLS1-DA for classification by molecular spectroscopy, Chemometr. Intell. Lab. Syst. 209 (2021) 104225, https://doi.org/10.1016/j.chemolab.2020.104225.
[141] J. Luts, F. Ojeda, R. Van De Plas, B. De Moor, S. Van Huffel, J.A.K. Suykens, A tutorial on support vector machine-based methods for classification problems in chemometrics, Anal. Chim. Acta 665 (2010) 129–145, https://doi.org/10.1016/j.aca.2010.03.030.
[142] Y. Xu, S. Zomer, R.G. Brereton, Support vector machines: a recent method for classification in chemometrics, Crit. Rev. Anal. Chem. 36 (2006) 177–188, https://doi.org/10.1080/10408340600969486.
[143] F. Marini, R. Bucci, A.L. Magrì, A.D. Magrì, Artificial neural networks in chemometrics: history, examples and perspectives, Microchem. J. 88 (2008) 178–185, https://doi.org/10.1016/j.microc.2007.11.008.
[144] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.
[145] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemometr. Intell. Lab. Syst. 174 (2018) 33–44, https://doi.org/10.1016/j.chemolab.2017.12.004.
[146] O.H.J. Christie, J.A. Brenna, E. Straume, Multivariate classificatin of Roman glasses found in Norway, Archaeometry 21 (1979) 233–241, https://doi.org/10.1111/j.1475-4754.1979.tb00257.x.
[147] R. Aruga, P. Mirti, A. Casoli, Application of multivariate chemometric techniques to the study of Roman pottery (terra sigillata), Anal. Chim. Acta 276 (1993) 197–204, https://doi.org/10.1016/0003-2670(93)85056-.
[148] R.J. Taylor, V.J. Robinson, D.J.L. Gibbins, An investigation of the provenance of the Roman amphora cargo from the plemmirio B shipwreck, Archaeometry 39 (1997) 9–21, https://doi.org/10.1111/j.1475-4754.1997.tb00787.x.
[149] J.A. Remolà, J. Lozano, I. Ruisánchez, M.S. Larrechi, F.X. Rius, J. Zupan, New chemometric tools to study the origin of amphorae produced in the Roman Empire, TrAC - Trends Anal. Chem. 15 (1996) 137–151, https://doi.org/10.1016/0165-9936(95)00091-7.
[150] K. Heydorn, I. Thuesen, Classification of ancient mesopotamian ceramics and clay using SIMCA for supervised pattern recognition, Chemometr. Intell. Lab. Syst. 7 (1989) 181–188, https://doi.org/10.1016/0169-7439(89)80122-4.
[151] R.H. King, Provenance of clay material used in the manufacture of archaeological pottery from Cyprus, Appl. Clay Sci. 2 (1987) 199–213, https://doi.org/10.1016/0169-1317(87)90031-7.
[152] T. Rotunno, L. Sabbatini, M. Corrente, A provenance study of pottery from archaeological sites near Canosa, Puglia (Italy), Archaeometry 39 (1997) 343–354, https://doi.org/10.1111/j.1475-4754.1997.tb00811.x.
[153] C. Pizarro, N. Pérez-del-Notario, C. Sáenz-González, S. Rodríguez-Tecedor, J. M. González-Sáiz, Matching past and present ceramic production in the Banda area (Ghana): improving the analytical performance of neutron activation analysis in archaeology using multivariate analysis techniques, Archaeometry 1 (2012) 101–113, https://doi.org/10.1111/j.1475-4754.2011.00601.x.
[154] Z.S. Duma, A. Surakka, P. Härmä, H. Laxström, T. Sihvonen, S.P. Reinikainen, Tool for similarity identification of rapakivi granites in heritage buildings, J. Cult. Herit. 58 (2022) 229–236, https://doi.org/10.1016/j.culher.2022.10.013.
[155] G.E. De Benedetto, B. Fabbri, S. Gualtieri, L. Sabbatini, P.G. Zambonin, FTIR-chemometric tools as aids for data reduction and classification of pre-Roman ceramics, J. Cult. Herit. 6 (2005) 2005–2011, https://doi.org/10.1016/j.culher.2005.06.004.
[156] J. Peris-Vicente, M.J. Lerma-García, E. Simó-Alfonso, J.V. Gimeno-Adelantado, M. T. Doménech-Carbó, Use of linear discriminant analysis applied to vibrational spectroscopy data to characterize commercial varnishes employed for art purposes, Anal. Chim. Acta 589 (2007) 208–215, https://doi.org/10.1016/j.aca.2007.03.001.
[157] Z. Haghighi, A.H. Karimy, F. Karami, A. Bagheri Garmarudi, M. Khanmohammadi, Infrared spectroscopic and chemometric approach for identifying binding medium in Sukias mansion's wall paintings, Nat. Prod. Res. 33 (2019) 1052–1060, https://doi.org/10.1080/14786419.2015.1108974.
[158] J. Xia, J. Zhang, Y. Zhao, Y. Huang, Y. Xiong, S. Min, Fourier transform infrared spectroscopy and chemometrics for the discrimination of paper relic types, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 219 (2019) 8–14, https://doi.org/10.1016/j.saa.2018.09.059.
[159] T. Zhao, M. Peng, M. Yang, R. Lu, Y. Wang, Y. Li, Effects of weathering on FTIR spectra and origin traceability of archaeological amber: the case of the Han Tomb of Haihun Marquis, China, J. Archaeol. Sci. 153 (2023) 105753, https://doi.org/10.1016/j.jas.2023.105753.
[160] P.M. Ramos, I. Ruisánchez, Data fusion and dual-domain classification analysis of pigments studied in works of art, Anal. Chim. Acta 558 (2006) 274–282, https://doi.org/10.1016/j.aca.2005.10.066.
[161] P.M. Ramos, I. Ruisánchez, K.S. Andrikopoulos, Micro-Raman and X-ray fluorescence spectroscopy data fusion for the classification of ochre pigments, Talanta 75 (2008) 926–936, https://doi.org/10.1016/j.talanta.2007.12.030.
[162] E. Manzano, J. García-Atero, A. Dominguez-Vidal, M.J. Ayora-Cañada, L. F. Capitán-Vallvey, N. Navas, Discrimination of aged mixtures of lipidic paint binders by Raman spectroscopy and chemometrics, J. Raman Spectrosc. 43 (2012) 781–786, https://doi.org/10.1002/jrs.3082.
[163] C. Defeyt, J. Van Pevenage, L. Moens, D. Strivay, P. Vandenabeele, Micro-Raman spectroscopy and chemometrical analysis for the distinction of copper phthalocyanine polymorphs in paint layers, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 115 (2013) 636–640, https://doi.org/10.1016/j.saa.2013.04.128.
[164] F. De Angelis, A. Di Tullio, R. Ceci, R. Quaresima, M. Marsili, Application of multivariate analysis for recognition of organic patinas on stone monuments, J. Separ. Sci. 25 (2002) 29–36, https://doi.org/10.1002/1615-9314(20020101)25:1/2<29::AID-JSSC29>3.0.CO;2-1.
[165] W. Fremout, S. Kuckova, M. Crhova, J. Sanyova, S. Saverwyns, R. Hynek, M. Kodicek, P. Vandenabeele, L. Moens, Classification of protein binders in artist's paints by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry: an evaluation of principal component analysis (PCA) and soft independent modelling of class analogy (SIMCA), Rapid Commun. Mass Spectrom. 25 (2011) 1631–1640, https://doi.org/10.1002/rcm.5027.
[166] M.E. Castillo-Valdivia, A. López-Montes, T. Espejo, J.L. Vílchez, R. Blanc, Identification of starch and determination of its botanical source in ancient manuscripts by MEKC-DAD and LDA, Microchem. J. 112 (2014) 75–81, https://doi.org/10.1016/j.microc.2013.09.019.
[167] M. Manfredi, E. Robotti, G. Bearman, F. France, E. Barberis, P. Shor, E. Marengo, Direct analysis in real time mass spectrometry for the nondestructive investigation of conservation treatments of cultural heritage, J. Anal. Methods Chem. 2016 (2016) 6853591, https://doi.org/10.1155/2016/6853591.
[168] B. Genc Oztoprak, M.A. Sinmaz, F. Tülek, Composition analysis of medieval ceramics by laser-induced breakdown spectroscopy (LIBS), Appl. Phys. Mater. Sci. Process 122 (2016) 557, https://doi.org/10.1007/s00339-016-0085-9.
[169] S. Duchene, V. Detalle, R. Bruder, J. Sirven, Chemometrics and laser induced breakdown spectroscopy (LIBS) analyses for identification of wall paintings pigments, Curr. Anal. Chem. 6 (2009) 60–65, https://doi.org/10.2174/157341110790069600.
[170] J. Linderholm, P. Geladi, C. Sciuto, Field-based near infrared spectroscopy for analysis of Scandinavian Stone Age rock paintings, J. Near Infrared Spectrosc. 23 (2015) 227–236, https://doi.org/10.1255/jnirs.1172.
[171] M. Romani, G. Capobianco, L. Pronti, F. Colao, C. Seccaroni, A. Puiu, A.C. Felici, G. Verona-Rinati, M. Cestelli-Guidi, A. Tognacci, M. Vendittelli, M. Mangano, A. Acconci, G. Bonifazi, S. Serranti, M. Marinelli, R. Fantoni, Analytical chemistry approach in cultural heritage: the case of Vincenzo Pasqualoni's wall paintings in S. Nicola in Carcere (Rome), Microchem. J. 156 (2020) 104920, https://doi.org/10.1016/j.microc.2020.104920.
[172] J. Engel, J. Gerretzen, E. Szymańska, J.J. Jansen, G. Downey, L. Blanchet, L.M. C. Buydens, Breaking with trends in pre-processing? TrAC - Trends Anal. Chem. 50 (2013) 96–106, https://doi.org/10.1016/j.trac.2013.04.015.
[173] Y.Y. Pu, C. O'Donnell, J.T. Tobin, N. O'Shea, Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing, Int. Dairy J. 103 (2020) 104623, https://doi.org/10.1016/j.idairyj.2019.104623.
[174] G. Gorla, P. Taborelli, H.J. Ahmed, C. Alamprese, S. Grassi, R. Boqué, J. Riu, B. Giussani, Miniaturized NIR spectrometers in a nutshell: shining light over sources of variance, Chemosensors 11 (2023), https://doi.org/10.3390/chemosensors11030182.
[175] G. Gorla, P. Taborelli, C. Alamprese, S. Grassi, B. Giussani, On the importance of investigating data structure in miniaturized NIR spectroscopy measurements of food: the case study of sugar, Foods 12 (2023), https://doi.org/10.3390/foods12030493.

[176] R. Gautam, S. Vanga, F. Ariese, S. Umapathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy, EPJ Tech. Instrum. 2 (2015) 8, https://doi.org/10.1140/epjti/s40485-015-0018-6.
[177] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC - Trends Anal. Chem. 28 (2009) 1201–1222, https://doi.org/10.1016/j.trac.2009.07.007.
[178] L.C. Lee, C.Y. Liong, A.A. Jemain, A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum, Chemometr. Intell. Lab. Syst. 163 (2017) 64–75, https://doi.org/10.1016/j.chemolab.2017.02.008.
[179] C.H. Hsiao, Y.H. Lai, S.Y. Kuo, Y.H. Cai, C.H. Lin, Y.S. Wang, A dynamic data correction method for enhancing resolving power of integrated spectra in spectroscopic analysis, Anal. Chem. 92 (2020) 12763–12768, https://doi.org/10.1021/acs.analchem.0c00737.
[180] E. Lopez, J. Etxebarria-Elezgarai, J.M. Amigo, A. Seifert, The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples, Anal. Chim. Acta 1275 (2023) 341532, https://doi.org/10.1016/j.aca.2023.341532.
[181] K. Baumann, M. Von Korff, H. Albert, A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part II. Practical applications, J. Chemom. 16 (2002) 351–360, https://doi.org/10.1002/cem.729.
[182] K. Baumann, H. Albert, M. Von Korff, A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part I. Search algorithm, theory and simulations, J. Chemom. 16 (2002) 339–350, https://doi.org/10.1002/cem.730.
[183] C.M. Andersen, R. Bro, Variable selection in regression-a tutorial, J. Chemom. 24 (2010) 728–737, https://doi.org/10.1002/cem.1360.
[184] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, Chemometr. Intell. Lab. Syst. 118 (2012) 62–69, https://doi.org/10.1016/j.chemolab.2012.07.010.
[185] Y.H. Yun, H.D. Li, B.C. Deng, D.S. Cao, An overview of variable selection methods in multivariate analysis of near-infrared spectra, TrAC - Trends Anal. Chem. 113 (2019) 102–115, https://doi.org/10.1016/j.trac.2019.01.018.
[186] A. Smiti, A critical overview of outlier detection methods, Comput. Sci. Rev. 38 (2020) 100306, https://doi.org/10.1016/j.cosrev.2020.100306.