ARTICLE

Multivariate analysis of PIXE+XRF and PIXE spectral images

Iva Božičević Mihalić,*[a] Stjepko Fazinić,*[a] , Marko Barac,[a] Andreas Germanos Karydas,[b] Alessandro Migliori,[c] Damir Dorašić,[d] Vladan Desnica,[e] Domagoj Mudronja[f] and Dragica Krstić[g]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

In this work we demonstrate the usefulness of multivariate techniques for analysis of two-dimensional (2D) spectral images obtained by proton beam ionization (PIXE mode) and combined proton beam and photoionization with X-ray tube (PIXE+XRF mode). Two different multivariate analysis approaches were used: (i) Principal Component Analysis (PCA) for dimensionality reduction followed by k-means clustering for the identification of different sample regions, (ii) t-Distributed Stochastic Neighbour Embedding technique (t-SNE). In PCA+k-means clustering similar pixels were grouped in different clusters where a direct connection between individual clusters and elements identified from cluster spectra resulted in the fast image segmentation and identification of different sample regions. t-SNE was used for dimensionality reduction and simple 2D visualization of high dimensional data. Three different cases were investigated: (i) qualitative analysis of 2D spectral maps having pixel spectra with high number of counts per pixel in the full range of measured X-ray energies collected in PIXE+XRF mode excitation, (ii) qualitative and semi-quantitative analysis of 2D spectral maps having pixel spectra with medium to low counts per pixel collected in PIXE mode, (iii) qualitative and quantitative analysis of 2D spectral maps having medium to low statistics per pixel obtained in PIXE mode. In the actual case studies, we identified all the pigments in artificial and real sample that was illumination from a historical book, and quantitatively characterized the identified gold layer and Niello decoration on the archaeological plate of Roman origin. In the last example, we were able to identify the sample regions with similar layer thicknesses, and obtain the layer thickness and elemental concentrations. We demonstrated that high statistics spectra that would contain enough information for qualitative and/or quantitative analysis of major, minor and even trace elements can be deduced using multivariate analysis methods even from low-statistics individual pixel spectra collected during 2D scanning of objects under investigation. This could be of particular importance for sensitive samples that could be damaged during long irradiation.

1 Introduction

Energy dispersive X-ray Emission Spectrometry (ED-XES) techniques offer high analytical potential for non-destructive and non-invasive material characterization and imaging.[1] They can be used for simultaneous detection of almost all the elements in the periodic table. In X-ray Fluorescence (XRF), photoionization is employed as the excitation mechanism with the use of radioisotopes, X-ray tubes or synchrotrons as sources of primary radiation. Electron (as in EPMA – Electron Probe Micro Analysis) or proton beams (PIXE – Particle Induced X-ray Emission) are also routinely used to ionize inner shell electrons and induce X-ray radiation from the sample. These methods, combined with the beam scanning optics or the use of translation stages for sample positioning during data acquisition, enable the collection of two-dimensional (2D) micro or macro elemental image maps based on the obtained X-ray spectra.

Generally PIXE exhibits higher sensitivity for lighter elements, whereas the efficient excitation of elements heavier than iron requires more energetic beams not so often available. Therefore, in the case of typical scanning PIXE analysis with single detection systems, the set-up throughput is expected to be rather low for heavier elements. There are experimental setups were large throughput is obtained by increasing detector solid angles, either using simultaneously several X-ray detectors[2,3] or using large area pixelated detectors.[4] Still, the majority of PIXE mapping systems are based on the use of single standard detectors (such as SDDs or Si(Li) detectors). Sensitivity for heavier elements can alternatively be increased by adding additional X-ray source to the setup. With the availability of miniature, low power and lightweight X-ray tubes it is easy to incorporate X-ray source within the Ion Beam Analysis (IBA) setup. This would help since X-ray fluorescence is more sensitive to the elements whose characteristic radiation is closer to the energy of the primary X-ray beam. Both the IBA spectroscopy methods and XRF can use the same data

[a] Laboratory for Ion Beam Interactions, Division of Experimental Physics, Rudjer Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia. E-mail: ibozicev@irb.hr; stjepko.fazinic@irb.hr
[b] Institute of Nuclear and Particle Physics, Patr. Gregoriou E & 27 Neapoleos Str, 15341 Agia Paraskevi, Greece.
[c] Nuclear Science and Instrumentation Laboratory, International Atomic Energy Agency (IAEA), 2444 Seibersdorf, Austria.
[d] Archaeological Museum, 10000 Zagreb, Croatia
[e] Laboratory for Science and Technology in Art, Academy of Fine Arts in Zagreb, Ilica 85, 10000 Zagreb, Croatia.
[f] Natural Science Laboratory, Croatian Conservation Institute, Grskoviceva 23, 10000 Zagreb, Croatia.
[g] Preservation and Conservation Department, National and University Library in Zagreb, Hrvatske bratske zajednice 4, 10000 Zagreb, Croatia

acquisition modules, and therefore can be easily incorporated in the same analytical setup even for simultaneous use. In particular, for low energy accelerators (2 MeV maximum proton energy), this dynamic integration of the XRF set-up with the external in-air IBA end-station expands significantly the X-ray analytical range and sensitivity in the detection of elements heavier than iron, as their excitation by PIXE would be hampered by the much lower than 2MeV in-air extracted proton beam. The combined PIXE+XRF setup can be configured to obtain similar sensitivity in the full dynamic range of measured X-ray energies (≈ 1-20 keV), that suits well for the measurement of 2D macro elemental image maps based on the simultaneously measured X-ray spectra even with the use of simple X-ray detectors.

Simultaneous PIXE+XRF excitation can be very useful for many practical applications where qualitative screening in the sense of identification of elements' distributions along investigated areas is all what is needed. However, such simultaneous excitation is not suitable for quantitative analysis due to quite different PIXE and XRF excitation mechanisms. When quantitative analysis is of importance, PIXE+XRF excitation can be used for initial screening of heterogeneuos objects and then single ion beam excitation can be peformed on selected areas or points for quantification using dedicated software like GUPIXWIN[5] or GeoPIXE.[6] Such analysis would in principle require longer measurements in order to acquire the spectra with high peak areas to reduce uncertainities in quantitiative analysis. GeoPIXE is a widely used program for obtaining quantitative elemental maps based on the dynamic analysis approach. Pichon et al. developed the mapping system employing multi-detector images in which each pixel of the image is quantified using GUPIXWIN, therefore providing elemental concentration maps.[7,8] Recently GUMAP, a GUPIXWIN based code for extracting regional spectra from list mode files was developed, however, the ROI's selection to obtain summed spectra for quantification purposes still relies on user's input.[9]

Alternativelly, higher statistics spectra that would contain enough information for qualitative and/or quantitative analysis of major, minor and trace elements can be deduced using multivariate analysis methods directly from (low-statistics) individual pixel spectra collected during 2D scanning of the object under investigation. This could be of particular importance for sensitive samples to reduce possible damage. In general, multivariate analysis methods can be very useful as alternative option to traditional analysis of X-ray spectra from two-dimensional arrays, that assumes a comparison of 2D elemental maps based on the characteristic elemental X-ray lines.

Although multivariate data analysis approach is already an established approach in different fields, it is not a common practice in XRF and is especially rarely reported in PIXE 2D analysis. In the case of μXRF imaging of geological and archaeological samples, Vekemans et al. has demonstrated the practical use of Principal Component Analysis (PCA)[10–12] and k-means clustering[13] of XRF elemental images for automated segmentation.[14] Vogt reported the use of PCA for improvement of μXRF analysis and demonstrated that fitted spectra from the PCA filtered data enabled extraction of the overlapping peaks with high accuracy.[15] PCA has been used for historical paint layers identification by means of XRF analysis,[16] but also combined with other techniques like Raman spectroscopy.[17–19] PCA has been applied also in the study of geological samples with XRF.[14,20,21] PCA filtering was shown to be effective in obtaining a better quality of low-intensity trace element images.[22] A multivariate curve resolution alternating least squares (MCR-ALS) approach faced the problem of resolving the overlapping emission peaks in wavelength dispersive XRF.[23]

The t-Distributed Stochastic Neighbour Embedding (t-SNE)[24] technique aproved to be a very promising tool in pigment identification by near infra-red imaging spectroscopy.[25] Pouyet et al. compared capabilities of multivariate methods like PCA, minimum noise function (MNF) and t-SNE in separating different pigments and their mixtures from the visible hyperspectral imaging (HIS) and XRF data.[26] They showed that t-SNE is quite efficient in distinguishing both pure pigments and pigment mixtures even in cases when three different pigments were mixed. PCA and t-SNE proved to be very useful in identification of outlier compositions and for finding patterns in XRF data sets obtained in time-constrained experiments on robotic planetary surveys.[27]

In the case of PIXE, only few studies on the use of multivariate analysis have been found. Swietlicki used PCA and partial least square regression (PLS) on the μPIXE elemental maps obtained together with the ion beam induced luminescence (IBIL) data.[28] Doyle et al. used multivariate curve resolution to reduce the spectral image data sets into a small number of interpretable components that describe the original system.[29] The focus of their study was on low statistics imaging data from which integrated spectra are obtained that can be further quantified using routine analysis programs. This approach was found very useful in the characterization of particulate samples.[30] Recently, 2D PIXE spectral data obtained with a broad beam in-air irradiation of a decorative tile from a historical building were processed by means of non-negative matrix factorization (NMF) and k-means clustering to separate groups of similar pixel data to different clusters.[31]

In this work, we highlight the advantages of using multivariate methods for processing 2D PIXE and combined PIXE + XRF spectral images. The goal of the study is to demonstrate the usefulness of the applied multivariate methods in comparison to traditional analysis of 2D elemental maps. For this purpose, we have acquired 2D elemental images of selected objects at the external ion beam setup of the Rudjer Boskovic Institute Tandem Accelerator Facility (RBI-AF) using a proton beam and miniature X-ray tube as excitation sources. We present three different examples: (i) one where simultaneous PIXE+XRF spectra were collected for qualitative 2D analysis; (ii) the other where PIXE spectra were collected for qualitative and semiquantitative 2D analysis; and (iii) one where PIXE espectra were collected for qualitative and quantitative 2D analysis.

2 Samples, measurements and data analysis

Table 1 Pigments used in the preparation of the sample shown in Fig. 1 (a) (the first column contains marks for identification of the painted lines).

MARK	PIGMENT	FORMULA
B-J	Ultramarine	$3Na_2O \cdot 3Al_2O_3 \cdot 6SiO_2 \cdot Na_2S$
CI	Cinnabar	HgS
DH	Red ochre	$Fe_2O_3 \cdot nH_2O$
L-Centre	Auripigment	As_2S_3
Centre-F	Lithopone	$BaSO_4 + ZnS$
A-Centre	Lead white	$2PbCO_3 \cdot Pb(OH)_2$
Centre-G	Ultramarine + Malachite	
E-Centre	Titanium white	TiO_2
Centre-K	Malachite	$CuCO_3 \cdot Cu(OH)_2$

2.1 Samples

The above mentioned experiments were performed on three different samples. The first experiment was performed on sample 1 that was specifically created as a painted structure on a thick marble block, shown in Fig. 1 (a) and (b). The sample consists of a series of painted lines on a marble surface. The following pigments were used to draw the lines marked in the Fig. 1 and listed in Table 1: ultramarine ($3Na_2O \cdot 3Al_2O_3 \cdot 6SiO_2 \cdot Na_2S$), cinnabar (HgS), red ochre ($Fe_2O_3 \cdot nH_2O$), auripigment (As_2S_3), lithopone ($BaSO_4 + ZnS$), lead white ($2PbCO_3 \cdot Pb(OH)_2$), a mixture of ultramarine and malachite ($CuCO_3 \cdot Cu(OH)_2$), titanium white (TiO_2) and malachite. In-air PIXE+XRF irradiation was performed to acquire related X-ray spectra for qualitative 2D analysis in order to assess the capabilities of multivariate analysis for pigment identification in comparison with the traditional approach of using 2D elemental maps. The sample was created by one of the authors and only after irradiation and analysis of the object, the results were compared with the actual structure that was unknown to analysts during the analysis.

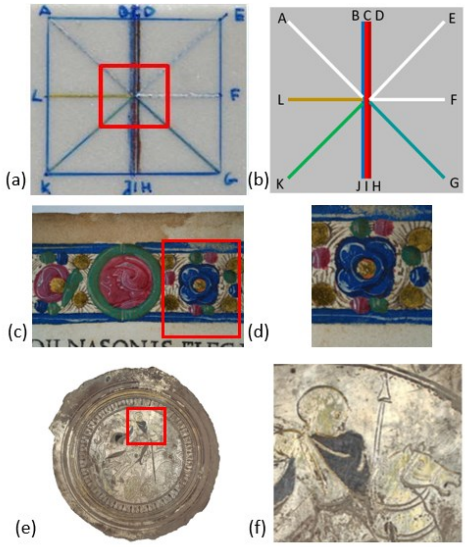

Fig. 1 Sample 1: (a) series of painted lines on marble surface and (b) schematic drawing of the sample. Sample 2: (c) part of the page from the book under analysis showing illumination on top and (d) part of the illumination with the analysed region. Sample 3: (e) decorative medallion of the silver tray (No. 5) from the late Roman period and (f) analysed region of the medallion.

The second experiment was performed on sample 2, that was a selected area of illuminations from the 15th century book of poems by the Roman poet Publius Ovidius Naso published in the printing press of Jacobus Rubeus in Venice in 1474. The book is from the Manuscripts and Old Books Collection of the National and University Library in Zagreb. The book was subject of analysis by a number of non-destructive and non-invasive methods. It is modestly decorated with illuminations. In-air PIXE analysis was performed on some illuminations with the final goal to identify pigments for better understanding of materials and decorating techniques for conservation purposes, but also for a better insight into the development of book painting. For the present analysis, 2D PIXE map has been taken on the selected area (Fig. 1 (c-d)). In addition, X-ray spectra were taken from selected points (individual pixels) representing different pigments.

The third experiment was performed using sample 3, an archaeological find, a silver tray (No. 5) from the late Roman period (Fig. 1 (e) and (f)) that belongs to the collection of luxury silver dishes found in the eastern Croatian town Vinkovci. In today's area of Vinkovci the town Colonia Aurelia Cibalae was located during the Roman period. That was the third-largest town in the province Pannonia Secunda and the birthplace of two Roman emperors. In 2012 altogether 50 silver objects were excavated, of which 21 complete, including trays, plates, bowls, jugs, cups, spoons, etc. The items are significantly fragmented and brittle. The original surface of most objects has been partly preserved but is covered with a thin or thick layer of corrosion products. The analysis of the central area of the decorative medallion has been a part of a series of preliminary investigations carried out before the conservation and restoration treatment on a limited number of findings. The goal is to collect basic data on the elemental composition of the material and corrosion products developed during the centuries to help determine further treatment procedures and also to gain insight into the original production and decoration techniques.

2.2 Measurements

The measurements were performed at the external ion beam end-station, one of the beamlines connected to the 1 MV Tandetron accelerator which can deliver protons with the energy of up to 2 MeV. The proton beam is extracted in air through 8 μm thin Al foil and directed to a sample positioned at the distance of about 8.5 mm from the exit foil. The resulting proton beam energy on the target is about 1580 keV with spot size of 1mm.

Portable lightweight (500 g) Moxtek Magnum transmission anode X-ray tube has been incorporated at the external beam end-station. The tube is equipped with a grounded Rh anode and 0.25 mm thick Be exit window. High voltage between -10 to -50 kV with beam current between 0 to 200 μA can be applied. The tube has been positioned to irradiate the same spot on the sample as the ion beam used for PIXE measurements. When targets are properly positioned, the ion/X-ray beam directions to target normal are 16° and 50°, respectively, with the X-ray detector axis to form 55° with the target normal. A circular

collimator of 0.8 mm in diameter is placed in between the tube anode and sample, at 18 mm and 15 mm, respectively.

Two X-ray detectors have been tested and used: (i) Si(Li) with 3 mm thick crystal and 25 μm Be window; (ii) SDD with 0.45 mm thick crystal and very thin polymer window with thin Al/Mylar absorber to prevent light entering into the detector. The Si(Li) detector was positioned at 24 mm from the target, while SDD at 18 mm from the target.

Our homemade data acquisition system SPECTOR has been used to acquire the X-ray spectra. In the case of the present application, the software has been adapted to collect measured spectra while moving a target to create two-dimensional (2D) maps over the sample surface. The sample holder can accept relatively large objects and lateral scans of up to about 40 cm by 40 cm are possible. The sample can also be moved along the ion beam axis for proper sample alignment. In creating 2D elemental images, SPECTOR uses transient spectra collected at each pixel and through ROI's definition, elemental images are displayed during measurement. All the data can be also collected in list mode and saved to files for later off-line preprocessing, including multivariate analysis.

The XRF excitation channel has been configured to take the advantage of the fact that PIXE exhibits higher sensitivity for lighter elements. Thus, the optimization of the XRF set-up aimed to excite and analyse efficiently trace elements heavier than iron (Z=26) with same spatial resolution to PIXE mode of analysis. For this purpose, the operation of the X-ray tube was tested at a relative high voltage value (35 - 40kV), whereas proper diaphragms were inserted within a specifically designed holder to match the size of the two exciting beams. Further on, the primary X-ray beam from the tube Rh anode was filtered out to improve peak to background ratio for the analysis of those trace elements having characteristic X-ray energies more than about 7 keV. The selection of filters was optimized by means of PyMca calculations[32,33] and through several measurements of scattered radiation from a thick Teflon target. Finally, and in accordance with previous experience in pXRF and handheld XRF analysis[34] a combined filter composed by 50 μm Cu foil and 50 μm Ti foil was inserted in the excitation path. In this way, the combined excitation with protons and X-rays allowed us to obtain optimized and almost equivalent elemental sensitivities for all characteristic X-ray energies up to 15 keV. However, when performing combined excitation, one needs to have in mind the different ionization mechanisms and depth sensitivities, since 1.58 MeV proton beam in the case of a painted structure on marble can penetrate up to a depth of about 25 μm, while the primary X-ray beam can penetrate generally much deeper (for example 850 μm $CaCO_3$ absorbs ≈50% the Rh-Kα X-ray line).

The first experiment was performed using the painted structure on a thick marble block (Fig. 1 (a), (b)). X-ray spectra were obtained from an area of 25x25 mm^2 with the combined proton beam and X-ray tube excitation and using the SD detector. Irradiation time was 1 s/pixel with 25 x 25 pixels. In this way, for each pixel, spectra with good statistics were collected for the selected range of X-ray energies up to 15 keV and above. Although Ca K X-rays dominate in all the spectra (proton beam penetrates through thin pigment layers), X-rays of minor elements are also present with high intensities (Fig. 4).

The second experiment was performed on the illumination from the medieval book (Fig. 1 (c), (d)). An area of about 35x35 mm^2 was selected that contain all the visible pigments. Proton beam was scanned over the sampled area with a scan resolution of 40x40 pixels and with irradiation time of 5 s/pixel to obtain good statistics even in higher energy channels. In all the pixels X-ray spectra contain some contribution from the paper substrate (dominating Ca peak and lower peaks of Si, S, Cl, K, Fe and Cu), as a result of the proton beam penetrating through relatively thin pigment layers. Although absolute quantitative analysis is not possible, relative elemental concentrations can be estimated for mixed painting layers helping to identify the combination of pigments used.

The third experiment was performed on a decorative medallion of the silver tray (Fig. 1 (e), (f)). A selected area of about 30x30 mm^2 was scanned by the low intensity proton beam with a scan resolution of 64x64 pixels, and with the irradiation time of 3 s/pixel. In this case silver L lines dominate in the single pixel spectra while the other channels have low statistics (Fig. 14).

2.3. Data analysis

As a first step, we created 2D elemental maps directly from the measured spectra, which is a common approach usually performed at most laboratories. In case of high statistics single pixel spectra, 2D elemental maps were ploted using net peak intensities, obtained by fitting the measured pixel spectra with the hypermet function, whereas SNIP algorithm was applied for a non-analytical background estimation.[35] In case of low statistics single pixel spectra, 2D elemental maps were created from predefined regions of interests (ROIs) around characteristic X-ray peaks in measured spectra. ROIs were defined using cumulative measured spectra.

Multivariate data analysis was performed by two independent ways: (i) using combined principal component analysis (PCA) for dimensionality reduction and k-means clustering in order to group the data in multidimensional space, (ii) applying independently the t-Distributed Stochastic Neighbour Embedding (t-SNE), as a low dimensional mapping technique. Before multivariate data analysis preprocessing of experimental data was performed. For semiquantitative and quantitative analysis we used raw, unprocessed spectra.

2.3.1. Preprocessing before multivariate analysis. Preprocessing of data includes all the operations on the raw data before multivariate analysis. In this work, raw data are measured X-ray spectra from each pixel belonging to the scanned area of a sample. Since preprocessing can influence the analysis outcome, it should be done with care and taking into account the type of multivariate method, the goal of the analysis and data itself.

In this work, preprocessing includes dead pixel removal, normalization, mean centring and scaling. Dead pixel removal will be addressed in Section 3.1. In our case normalization of the pixel spectra was necessary to account for possible variation of ion doses collected between different pixels. We performed

simple normalization by dividing channel intensities with total X-ray counts for each pixel spectra.

Scaling is important to avoid the loss of useful information contained in the spectral channels with low intensities belonging to the presence of minor and trace elements. Centring is performed to take into account the data offset. We have used the mean centring method: from each channel in a pixel spectrum we subtracted the mean intensity of that channel over all pixels. In this way, the variations of the channel intensities are observed relative to their mean value.

In the analysis we have used three different scaling/centring procedures: (i) auto-scale (Z-normalization or unit variance scaling) of the data, where each channel intensity is first mean centred and then divided (scaled) by standard deviation of the channel; (ii) mean centring without scaling, (iii) square root transform followed by mean centring.[36–38] The last procedure has been used for the spectra with low channel intensities.

2.3.2 Multivariate analysis. Multivariate analysis was performed on the preprocessed pixel spectra by three techniques: Principal Component Analysis (PCA), k-means clustering and t-SNE. PCA is the most widely used multivariate analysis technique for data dimensionality reduction and finding linear relationships in data.[10–12] PCA enables better visualization of the correlation between pixel data and identifies the variables (in this case channel intensities) that describe the grouped pixel data.

K-means clustering is one of the most common and simple algorithms that groups data (individual pixel spectra) in high dimensional space based on their similarity.[13] In this work we used silhouette coefficient as an indicator of the optimal number of clusters. It is calculated with the formula:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (1)$$

where b_i is the mean intra-cluster distance while a_i is the mean distance to the instances of the next closest cluster. The optimal number of clusters is selected for the maximal averaged silhouette score.[39]

t-SNE is a t-Distributed Stochastic Neighbour Embedding technique for dimensionality reduction.[24] Local relationships between points in high dimensions are used to create a low-dimensional mapping represented by t-SNE x and t-SNE y coordinates. Distances between data points (spectra) in high dimensions are used to calculate conditional probabilities with Gaussian distribution representing the similarity between data points. A similar approach is used to construct the probability distribution in low dimensional space using student t-distribution reflecting the similarity in the low dimensional space. The result is that the objects that are not similar in high-dimensional space will be far apart in the low-dimension, while similar objects will be very close. Perplexity is a t-SNE parameter related to the number of effective nearest neighbors. Typical values are in the range between 5 and 50 but for denser data, higher perplexity values are recommended.

3 Results and discussion

3.1 Simultaneous PIXE and XRF excitation: pigments identification from high intensity pixel spectra

Simultaneous excitation by protons and X-rays from the tube was used to collect characteristic X-ray spectra from sample 1. High intensity spectra per pixel were collected in the full range of measured X-ray energies.

As a first step, traditional analysis has been done by creating 2D elemental maps by fitting measured normalized pixel spectra using PyMca software.[32] The result is presented at Fig. 2. It shows 2D spatial distribution of Kα and Kβ X-ray peak areas of Mg, Al and Si, Kα X-rays of S, K, Ca, Ti, Mn, Fe, Cu, Zn and As, as well as L3 X-ray peak areas of Ba, Hg and Pb. Fig. 2 could be used to identify pigments with the aid of one of the available pigment databases (in this case Table 1 can be used). For example, the position of the white painted line (E-Centre) corresponds to the 2D map of Ti, identifying titanium white. Red ochre (D-H) could be associated with the Fe 2D map. Lead white (A-Centre) is associated to Pb 2D map. Cinnabar (C-I) is associated with S and Hg 2D maps. Hg 2D map confirms that the line C-I is painted using cinnabar. Auripigment can be associated to S and As 2D maps. As 2D map confirms its correspondence with the L-centre line. Lithopone pigment (Centre-F) is connected with 2D maps of Ba, S and Zn and can be easily identified from Ba and Zn 2D maps. Ultramarine (B-J) is associated with Al and Si 2D maps (Na K X-ray peaks are not visible in the spectra) and malachite (Centre-K) with Cu 2D map. The line Centre-G is visible in 2D maps of Al, Si and Cu that correlates with the mixture of ultramarine and malachite. Therefore, in principle all the pigments have been identified. However, for unknown samples, this way of pigment identification can be much more complicated and time consuming. Identification may fail in difficult situations, for example in situations when pigments overlap, when 2D elemental maps contain contributions of

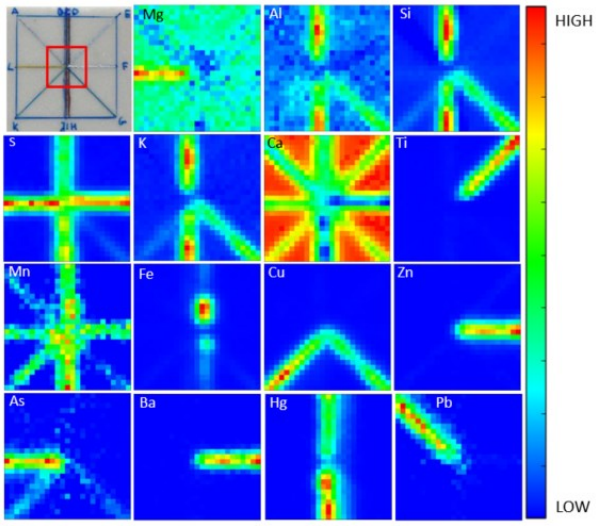

Fig. 2 Sample with the marked exposed area (red square) and preprocessed 2D elemental maps.

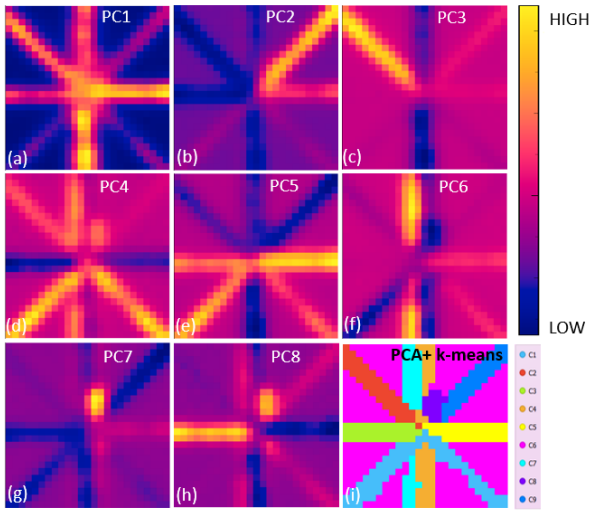

Fig. 3 (a-h) Score images of first 8 principal components obtained by the PCA method. (i) The result of the combined PCA + k-means clustering analysis.

different elements, when related spectra contain overlapping peaks, etc. Also, in this kind of analysis the user input and influence could significantly affect the results since improper selection of characteristic X-rays in the spectra for transfer to 2D elemental maps may lead to incomplete or wrong pigment identification.

As a second step we performed off-line multivariate analysis on the measured individual pixel spectra. In this case the goal was to identify 2D patterns or regions with similar spectra instead of the regions with the same elements. Before multivariate analysis, we performed preprocessing of data. During this phase, one single dead pixel was found, and its value was manually corrected. Correction was done by interpolation of spectra from eight neighbouring pixels. Median value of these eight spectra was calculated and transformed into the dead pixel spectrum. In the case of multiple dead pixels in an image, certain algorithms for locating dead pixels can be applied[40] and then corrected in a similar way. In general, variation of Ca intensities in the spectra was within 5%. However, in several individual pixels very high Ca values were recorded, implying the need for data normalization before multivariate analysis. Normalization to the total X-ray counts was performed. Dead pixel correction and normalization to the total X-ray intensity and data transformation to the input matrix for multivariate analysis was done with in-house Matlab routines (The Mathworks Inc., Natick, USA). Preprocesed data were saved in a hierarchical data format (HDF). PCA combined with k-means clustering and t-SNE analysis was performed on the whole set of X-ray spectra with Orange.[41–43]

Multivariate analysis was performed on the full spectrum data to minimize user input and influence. Such approach enables extraction of additional features from the images that could otherwise be overlooked by a selection of only specific characteristic X-rays in the spectra. The best PCA model was obtained with mean centring of pixel spectra and without scaling the variables. Fig. 3 (a-h) shows the principal component score images for the first 8 components, corresponding to the regions of different pigments on top of the marble plate. These 8 components together explain 99.8 % of the total variance of the sample.

k-means clustering with random initialization and 300 iterations was performed on auto-scaled PC score values. The optimal cluster number of 9 was obtained by the analysis of silhouette coefficient eqn (1). The result of the clustering is shown in Fig. 3 (i). The figure shows the clusters correlated with the actual position of the pigments on the sample. Cluster identification (association with pigments) can be done by observing the result of clustering (Fig. 3 (i)) and average cluster spectra, plotted in

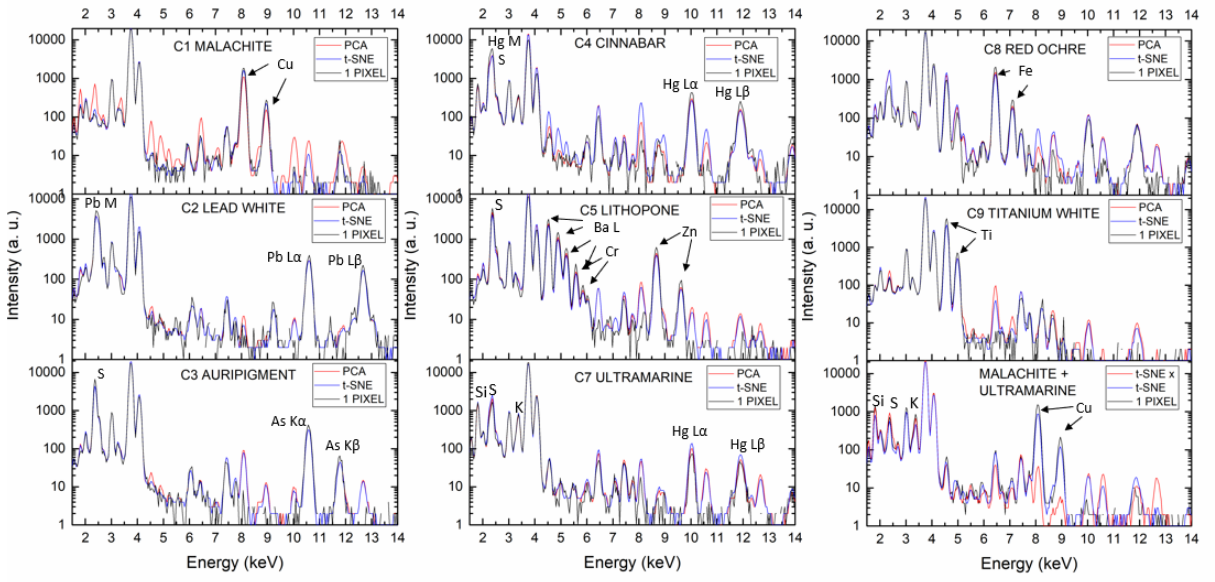

Fig. 4 Averaged spectra obtained by combined PCA and k-means clustering pigments compared with the related spectra obtained by the t-SNE analysis and with the single pixel spectra measured at specific sample positions.

Fig. 4. In such a way, direct association between the cluster and the pigment can be obtained: C1 as malachite due to a presence of Cu, C2 as lead white, C3 with As and S marking the auripigment, C4 with Hg and S as cinnabar, C5 as lithopone containing Ba, S and Zn, C7 as ultramarine, C8 with Fe as red ochre and C9 as titanium white. Cluster 6 belongs to marble support. The only shortcoming observed is that both pure malachite and the mixture of malachite and ultramarine were clustered to a single cluster. In addition to combined PCA and k-means clustering, we analysed the preprocessed spectral data with t-SNE setting the perplexity factor to 26 (Fig. 5). PCA is usually used in combination with clustering (like k-means) to group similar data after dimensionality reduction. Although the grouping of similar data points can be seen in the t-SNE plot and some form of clustering like DBSAN[44] can be performed on the t-SNE output,[25] it should be done with special care. Distances and density in high dimensions are not preserved in low dimensional t-SNE representation so they cannot be used as a measure of similarity in the low dimensional graph.[26] We used t-SNE for simple visualization of similar pixel spectral data. Data segmentation can be performed by manual selection of data points with the simultaneous observation of the average raw spectra and the position of selected pixels in the image. For comparison, the colours in the t-SNE graph (Fig. 5) correspond to the colours obtained with PCA and k-means clustering (Fig. 3 (g)).

The central part of the plot in Fig. 5 corresponds to the marble (purple circle) and the branches spreading from the centre to the individual pigments describe boundary regions between the marble support and specific pigment. Encircled areas at the end of branches are used for identification of pure pigments. The majority of pigments (lead white, auripigment, titanium white and lithopone) are very well separated from each other, matching the results of the PCA and k-means grouping. In some cases, differences are in few pixels, usually from the regions where two pigments overlap, like 2 ultramarine pixels

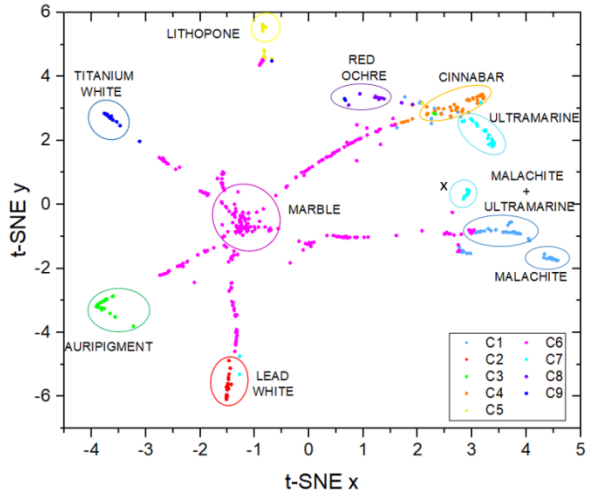

Fig. 5 t-SNE result on the pigment's preprocessed spectra, encircled areas are manually selected points for obtaining t-SNE pigment's spectra.

associated with the lead white group. Exceptions are red ochre, cinnabar and ultramarine which are partly overlapped on the sample. This is reflected in the t-SNE space. In the region of the C1 cluster, two branches can be observed: lower one belonging to the malachite and upper one to the mixture of malachite and ultramarine. Related average spectra obtained from encircled regions in the t-SNE plot are compared with the PCA + k-means spectra and single pixel pigment spectra at Fig. 4. Single pixel spectra are measured spectra from specific sample positions located at the edges of the measured regions in order to avoid interference with other pigments (these data points can be easily identified in the t-SNE plot as the points in the outside regions of the pigment groups – i.e., at the ends of respective branches).

In pure pigments, the average spectra obtained by t-SNE and by combined PCA + k-means approach are essentially the same. Small differences are seen in small peak intensities related to the pixels where two or more pigments intercept. The only discrepancy is seen in case of PCA + k-means C1 spectra and the corresponding t-SNE malachite region spectra. PCA + k-means grouped both malachite and its mixture with ultramarine to the same cluster.

In some cases, the average cluster spectra obtained by combined PCA + k-means method have additional peaks not present in the single pixel spectra (representing specific pigments). The single pixel spectrum of auripigment has a large As peak. PCA + k-means average cluster spectrum shows additionally Cu and Pb peaks. The origin of Cu and Pb contributions is from the regions where auripigment overlaps with malachite and lead white. In a similar way, PCA + k-means average cluster spectrum of lithopone shows some Fe, Cu, Hg and Pb peaks, while the corresponding titanium white shows Fe and Hg. These are indications of the regions where lithopone and titanium white pigments are mixed with the other pigments.

One group of data in the t-SNE map is situated in the region between ultramarine and mixture of malachite and ultramarine (marked with *x* sign in Fig. 5). In PCA + k-means analysis, that is a part of the ultramarine cluster. The *x* data points belong to the edge region between ultramarine and marble support with lower Hg intensities. In the region of lithopone, two separated groups can be seen in t-SNE plot: the upper group represents the central part of the pigment with higher levels of Ba and S while the lower group is from the boundary region of lithopone and marble with lower intensities of Ba and S. Although red ochre was painted as a continuous vertical line, it was localized in the centre of the image where other pigments intercept or even overlap. As a consequence, the elements from neighbouring pigments (Hg, Pb, S, Cu, Ti or Ba, As or Pb) are also present in related spectra.

We have demonstrated that multivariate spectral analysis methods like combined PCA + k-means clustering or t-SNE can be useful for qualitative analysis of two-dimensional arrays of X-ray spectra obtained by simultaneous PIXE – XRF setup. The obtained cluster spectra have higher statistics and enable detection of minor and trace elements without the need for longer measurements. Unfortunately, quantitative analysis in

terms of elemental concentration determination is not possible. However, there is a hope to overcome this problem in future, as it has been reported for the APXS portable PIXE-XRF spectrometer attached to the Mars exploration Rovers for in-situ mineral analyses on the planet Mars.[45,46]

3.2 PIXE excitation: pigments identification from medium to low statistics single pixel spectra

In this example we demonstrate the use of multivariate spectral analysis for pigment identification on a real (unknown) sample. Also, in this case we have used only PIXE excitation, resulting in the individual pixel spectra that in general have medium (low Z elements) to lower-statistics (elements heavier than Ti) peak intensities. Selected area from sample 2 (Fig. 1 (d)) was scanned by the low current proton beam (to avoid paper damage) with the scan resolution of 40x40 pixels and irradiation time of 5 s per pixel. Before multivariate analysis, recorded pixel spectra were normalized to the total spectrum counts using in-house Matlab routine and saved as HDF file.

As in the previous example, 2D elemental maps have been created (Fig. 6) by fitting the normalized pixel spectra with the hypermet function and SNIP non-analytical background estimation using PyMca. Ca, Fe, Cu were obtained using both Kα and Kβ lines, Zn from Kα lines while Ba, Au, Hg and Pb maps were obtained by summation of L3, L2 and L1 X-ray lines. The last two images of Fig. 6 are RGB images with spatial distributions of Pb (red), Zn (green), Fe (blue) and Hg (red), Au (green) and Cu (blue) respectively. In the classical approach, 2D spatial distribution of elements is used for pigment identifications. For example, Au distribution clearly confirms the gold leaves presence on the illumination. Cu is related to blue and green pigments indicating their origin as copper based pigments like azurite or malachite. Red pigment is associated with Pb elemental map where two possibilities appear for the pigment origin: (i) minium or (ii) some organic red pigment mixed with lead white.

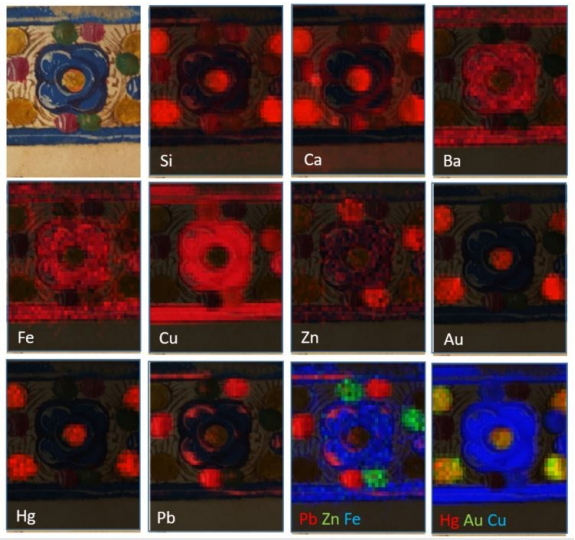

Fig. 6 Scanned selected area of sample 2; 2D elemental maps for 9 elements; and RGB composite maps for 6 elements.

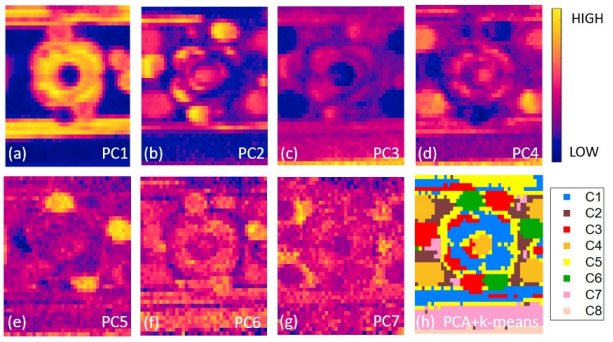

Fig. 7 (a-g) Score images of first 7 principal components obtained by the PCA method. (h) The result of the combined PCA + k-means clustering analysis.

The measured pixel spectra were then analysed by combined PCA and k-means clustering in Orange. The square root transform of the normalized data, followed by mean centring, was performed to account for the measurement uncertainties of the low statistics components in the pixel spectra. From each channel intensity divided by its square root, the channel mean value was subtracted. The resulting preprocessed pixel spectra were then used as an input for the PCA, followed by k-means clustering on the auto-scaled PC scores. The first seven principal components and the resulting image of the k-means clustering are shown in Fig. 7.

Easy pigment identification can be performed by direct comparison of the cluster spectra with the k-means clustering image (Fig. 7 (h)). With each cluster C1-C8, related cluster spectrum is associated (black spectra at Fig. 9), obtained by the summation of the raw pixel spectra belonging to the same cluster. Then C1 is associated with the blue pigment. Presence of Cu in C1 spectrum indicates that the blue pigment basis is most probably natural azurite. This is also supported with the observed minor contribution of Ba and Fe in the spectra, that seem to be usually present in natural azurite pigments.[47] Cluster C2 relates to the regions of the ink drawings on the paper with typical presence of Fe, Cu and Zn related to the iron gall ink.[48–50] Cluster C3 belongs to the red pigment which shows Cu and Pb in the related spectrum. The presence of Pb could indicate that red pigment is minium, but we cannot exclude the possibility that the pigment itself is of the organic origin, not detectable

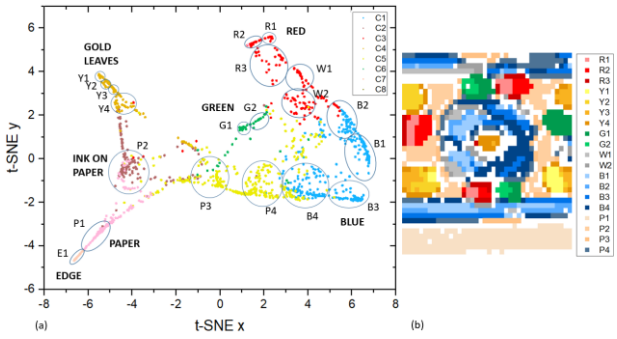

Fig. 8 (a) t-SNE plot with perplexity set to 27, (b) areas on the sample corresponding to the encircled regions at the t-SNE plot.

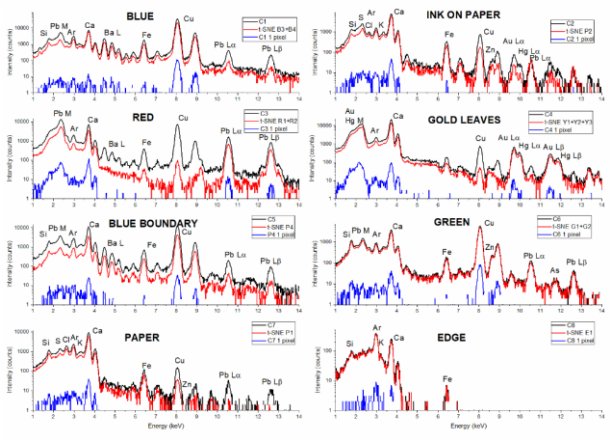

Fig. 9 PIXE spectra obtained after the summation on the PCA+k-means cluster spectra (black) and compared to summed spectra from selected regions of the t-SNE map (red) and single pixel spectra (blue).

with the X-ray spectroscopy techniques such as PIXE, and mixed with lead white. Also, higher Cu intensity is observed in the cluster spectrum, which will be explained latter. Cluster C4 is linked to the gold leaves with Au in the spectrum. Hg is also detected in coincidence with Au. It could be that the vermillion is painted below the gold leaves or Hg is a part of the adhesive ground under the leaves.[49] Cluster C5 is associated with the boundary region between the blue pigment and the paper with the same elements seen as in the blue pigment but with lower X-ray intensities. Cluster C6 relates to the green pigment areas. Cu as a major element is detected, most probably originating from verdigris as commonly used green pigment in illuminated manuscripts, but additional analysis should be performed for final separation among different copper based green pigments. Fe, Zn and As are present as trace elements. Finally, clusters C7 and C8 are associated with the paper regions, with C8 corresponding to the edge of the scan. In the spectra related to clear paper, beside Ca as the major component, Si, S, Cl, K, Fe, Cu and Zn are present.

Independently, t-SNE analysis was performed using the Orange code. The corresponding results with perplexity set to 27 are shown in Fig. 8 (a). Data points in the t-SNE plot (Fig. 8 (a)) are set to the same colour as clusters in Fig. 7 (h). Areas on the sample corresponding to the encircled regions at the t-SNE plot are shown in Fig. 8 (b). Fig. 9 shows summed raw spectra of the t-SNE encircled regions (red spectra), compared with the PCA+k-means cluster spectra (black spectra) and single pixel spectra (blue spectra). Substantial difference between t-SNE encircled regions and PCA+k-means spectra is observed only between the C3 and the corresponding t-SNE R1+R2 spectra. The C3 spectrum has visible Ba X-ray lines and higher Cu/Pb intensity ratio, compared to t-SNE R1+R2 spectrum. From the comparisons of the sample image (Fig. 1 (d)), the image of the k-means clustering (Fig. 7 (h)), and the t-SNE result shown at the Fig. 8 (b), one can conclude that C3 cluster spectrum relates to the red pigment (t-SNE R1 and R2 regions) but also includes contribution from the white areas on the blue flower (t-SNE W1 and W2 regions). This is clearly seen from the comparison of average pixel spectra obtained from the individual t-SNE selected regions in Fig. 10 (b) and (e): R1, R2 spectra related to the red pigment without Ba vs W1 and W2 spectra from white pigment on the flower with Ba and higher Cu contribution. Also Ca intensities in the average pixel spectra from the R1 and R2 regions are higher compared to the values found in the paper P1 average spectra (Fig. 10 (b)), probably due to the mixture of chalk ($CaCO_3$) with red pigment.

Y1, Y2 and Y3 t-SNE regions situated within the PCA+k-means C4 (gold leaves) have similar spectra (Fig. 10 (c)) with small variation in Au/Hg intensity ratio, while Y4 is identified as the edge region of the gold leaves. Hg originates from the layer below the gold leaves either as a part of the vermilion pigment or as a component of the gold ground preparation layer. It is observed that the intensity ratio Hg-L/Au-L increases from its minimum value in Y1 to its maximum value in Y3 region, probably due to a smaller thickness of gold leaves, which is reflected in lower absorption of Hg lines and higher Hg/Au intensity ratio. Also, it is observed that Ca intensities in the average pixel spectra (Fig 10 (c)) are higher compared to the Ca intensity in t-SNE P1 average paper spectrum due to its presence in the gold ground preparation layer.[49] Since Ca originates in the layer below the gold leaves, Ca Kα and Kβ X-ray lines are differently absorbed for different thicknesses of the gold leaves. Ca Kα/Ca Kβ intensity ratio increases 14 % from Y1 to Y3 region, supporting the fact of thinner gold layer in Y3 region, as it is also seen from Hg L/Au L intensity ratios.

For the green pigment, two regions on the t-SNE plot were selected (Fig. 8 (a)): G1 related to the darker green areas on the sample and G2 associated with the regions of white lines on the green pigment. Higher Pb/Cu intensity ratio in G2 regions is a consequence of lead white lines painted on top of the green pigment. Beside Cu and Pb, additionally Si, K, Fe, Zn and As are seen in the spectrum. Trace element concentration ratios were roughly estimated using thick target approximation in GUPIXWIN. The following estimates for the trace element concentrations relative to copper were obtained: Fe to Cu

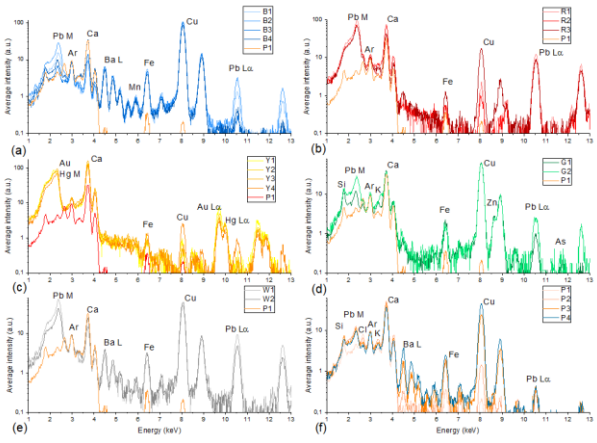

Fig. 10 Average spectra obtained from the selected (encircled) t-SNE regions (see Fig. 8 (a)). Pigment spectra are compared with the paper spectrum (t-SNE P1 region at Fig. 8 (a)).

concentrations at the level of 1 %, Zn to Cu at the level of 5 % and As at the level of 4 % of Cu concentrations.

Four different regions within C1 cluster were selected on the t-SNE plot (Fig 8 (a) and (b)) corresponding to the four shades of the blue pigment: B1 and B2 regions related to the left part of the flower and the blue stripe below the flower (see Fig. 1). B3 and B4 regions are from the right part of the flower and blue stripe above. It is clearly seen that the darker shades contain less Pb as an indication that shades of blue are the result of mixing with the lead white. This conclusion is in the line with the average spectra with lower Pb/Cu intensity ratio for B3 and B4 regions (Fig. 10 (a)). As in the case of green pigment, thick target approximation in GUPIXWIN was used for rough quantitative analysis to address the trace elements concentrations relative to Cu. Ba is at the level of 6 % of Cu concentration, which is about one order magnitude higher than the values found in selected natural azurite pigments in illuminated manuscripts of Spanish, Italian, French and Dutch origin dated from 13th to 16th century.[47] Fe concentrations are 2 % of Cu concentrations, values similar to concentrations found in the above mentioned natural azurite pigments.[47] Mn concentrations found are at the level of 3 % of Cu concentrations.

Illuminators usually drew their preliminary designs with ink or charcoal. In this case sketches were done in ink. These sketches were often reworked in ink that had been diluted so that it could be more easily concealed beneath the pigments. The observed presence of Fe and some other elements (Ca, K, Cu, Zn, ...) almost everywhere suggests that iron gall ink was used to draw the sketches.

Compared to the classical approach in the pigment identification, multivariate analysis enables fast separation of different pigment areas on the sample. In parallel, total or average pigment spectra can be used for fast identification of all major and trace elements present in the pigment. Application of t-SNE analysis on the PIXE spectral images additional information can be gathered, otherwise impossible to obtain by pure comparison of elemental maps. In the above case, t-SNE enabled localization of the regions with different thicknesses of gold leaves and separating different shades of blue pigments.

3.3 PIXE excitation: analysis of metallic object, low statistics single pixel spectra

Fig. 11 (a) shows the area of about 3x3 cm that was scanned by low intensity proton beam with a scan resolution of 64x64 pixels, and with irradiation time of 3 seconds per pixel. Excitation by protons was useful to limit the spectral response as much as possible only to the surface layer.

As a first step, the pixel spectra were saved in HDF file and imported into PyMca to obtain 2D elemental maps from related X-ray peak areas. The pixel spectra were in general of low intensity, especially for higher energy X-rays. Therefore 2D elemental maps are based on the measured intensities (i.e. the spectra were not fitted). Fig. 11 (b-k) shows the obtained 2D elemental maps of the elements identified in the cumulative (total) spectrum- (Si, S, Cl, Ca, Fe, Cu, Ag, Au, Hg) on top of the decorative medallion. Ca, Fe and Cu maps have a contribution from both Kα and Kβ lines. For Au L and Hg L maps summation

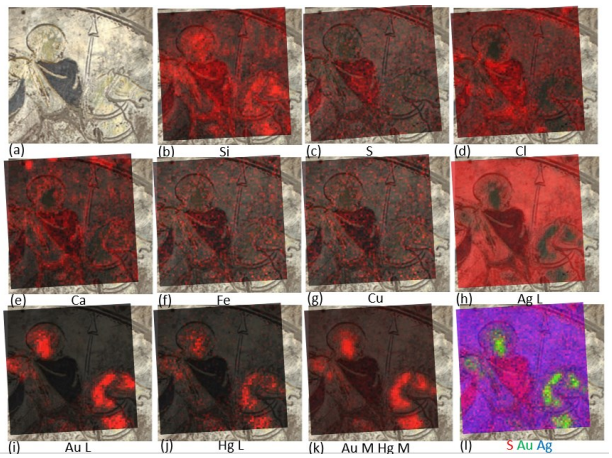

Fig. 11 (a) Image of the analyzed surface area; (b-k) areal 2D elemental maps; and (l) RGB composite map marks for 3 main elements.

of Lα, Lβ and Lγ lines was performed, while the Au M-Hg M map contains overlapping signals from Au M and Hg M X-rays. The Fig. 11 (l) shows an RGB image of S (red), Au (green) and Ag (blue) spatial distribution.

For quantitative analysis, longer irradiation on selected pixels can be performed to obtain higher statistics in related spectra used for analysis. Alternative approach is to apply some kind of pattern recognition algorithm (like multivariate analysis) to the already measured low statistics pixel spectra and sum the spectra from similar areas to obtain good statistics spectra. Advantages of this approach are: avoiding the need for long measurements and minimizing user influence on the analysis procedure.

Before the multivariate analysis, pixel spectra were normalized to the total spectrum counts to compensate for possible beam fluctuation during the measurements. Spectra were normalized with the in-house Matlab routine. PCA + k-means clustering and independently t-SNE on spectral data was performed in Orange. A combination of PCA and k-means clustering was applied in the analysis. To account for the measurement uncertainties of the low statistics components in the pixel spectra, we applied the combination of square root transform and mean centring. The resulting preprocessed (normalized, scaled and mean centred)

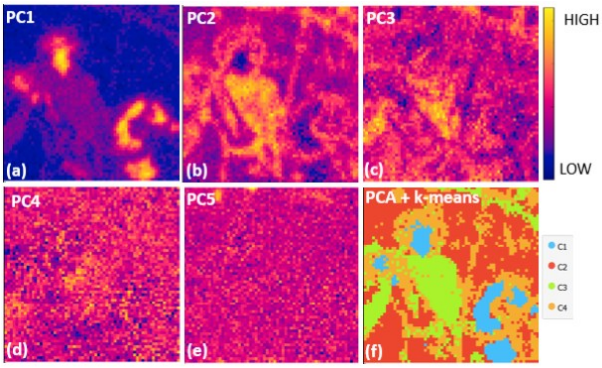

Fig. 12 (a-e) Score images of first 5 principal components obtained by the PCA method. (f) The result of the combined PCA + k-means clustering analysis.

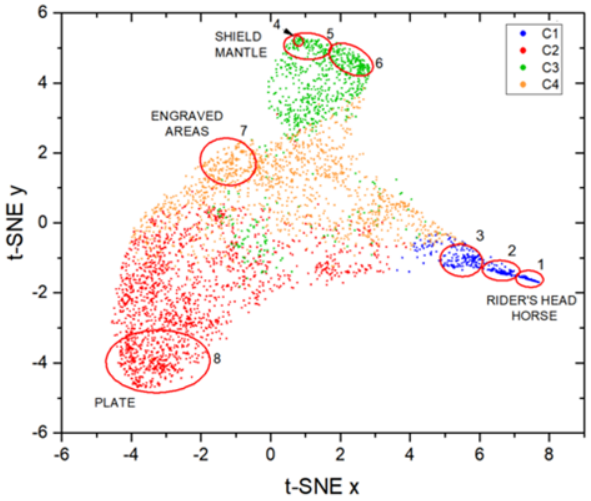

Fig. 13 t-SNE result with perplexity set to 22.

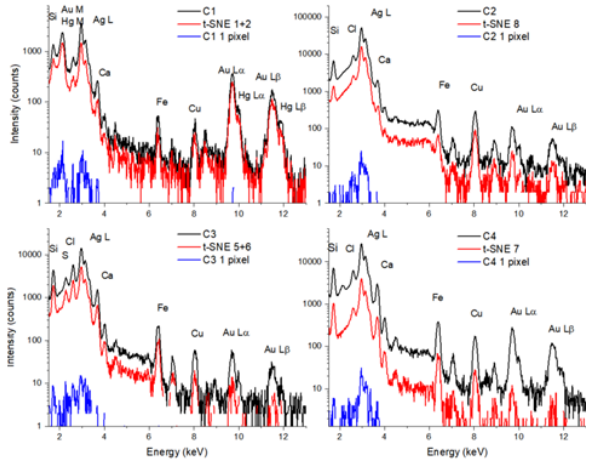

Fig. 14 PIXE spectra obtained after the summation on the PCA+k-means cluster spectra (black) and compared to summed spectra from selected regions of the t-SNE map (red) and single pixel spectra (blue)

pixel spectra were then used as input for the PCA, k-means clustering and t-SNE analysis.

Fig. 12 (a-e) shows the first 5 principal components obtained by the PCA method. The result of the combined PCA + k-means clustering analysis is shown at Fig. 12 (f). With each cluster C1-C4, related cluster spectrum is associated (black spectra at Fig. 14), obtained by the summation of the raw pixel spectra belonging to the same cluster. From the summed spectra, clusters can be easily identified: C1 relates to the gilding on the rider's head and horse with Au and Hg; C2 relates to the silver plate; C3 corresponds to the compound of Ag and S while C4 is associated with engraved areas having Ca, Cl and Si as impurities (originating from the ground where the plate was found).

t-SNE analysis was performed with perplexity parameter 22. The results are shown at Fig. 13. For better comparison with the result of the combined PCA and k-means clustering, data points are in the same colours as the clusters from Fig. 12 (f). Several (encircled) regions in the t-SNE plot were selected and summed spectra from these regions were compared with the PCA+k-means cluster spectra and with the single pixel spectra.

Although low statistics one-pixel spectra present an important drawback for this experiment, PCA+k-means or t-SNE analysis overcome this limitation. The resulting red spectra shown at Fig. 14 show that t-SNE regions 1 and 2 relate with the C1 cluster (these are very similar spectra); region 8 with C2; regions 5 and 6 with C3; and region 7 with C4.

t-SNE can be used for fast and easy selection of different sample regions with similar spectra. To demonstrate this, the most characteristic parts of the C1 and C3 clusters are divided into three regions each. Selection of the region 1 in Fig. 13 as a part of the C1 cluster shows that it relates to the central part of the horse and rider's head (Fig. 15 (a)) having the highest contribution of the Au and Hg (Fig. 15 (b)), while regions 2 and 3 correspond to boundaries between the gilding and engraved areas, having smaller Au and Hg contributions. Similar approach is used in investigating the area of rider's shield, with average spectra obtained from three regions within C3 with different S contributions (Fig. 15 (c)).

We used the summed spectra related to three encircled C1 components to quantitatively determine elemental composition. The related spectra (Fig. 15 (b)) show large

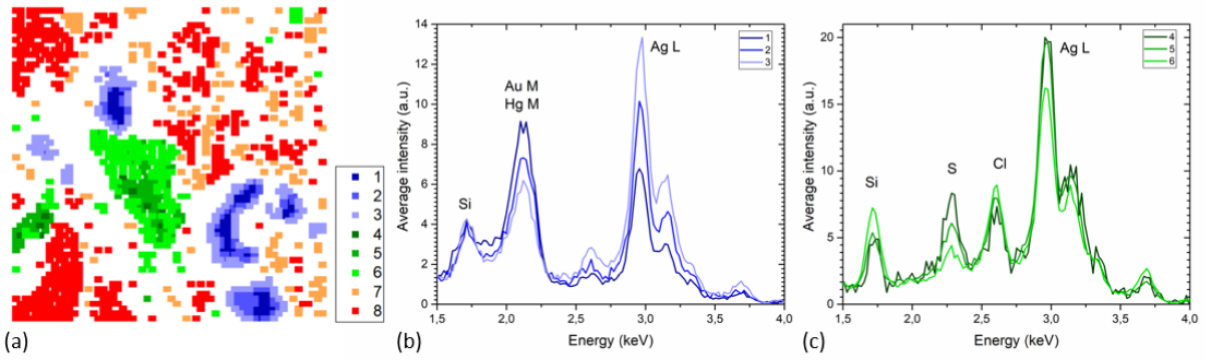

Fig. 15 (a) corresponding positions on the sample from the encircled regions of the t-SNE plot, average spectra obtained from the t-SNE (encircled) regions within (b) C1 and (c) C3 clusters.

Table 2 Quantitative analysis results of selected regions 1-6 from t-SNE plot.

	Concentrations (weight %) and relative uncertainties			Layer thickness (μg/cm²)
Region	Au	Hg		
1	86.9 (6 %)	13.1 (15 %)		670
2	82.8 (6 %)	17.2 (15 %)		570
3	84 (6 %)	16 (15 %)		450
	Ag	S	Cu	
4	84.7 (6 %)	14.8 (14 %)	-	≥ 3.9 μm
5	89 (6 %)	10.7 (14 %)	0.3 (40 %)	< 3.9 μm
6	91.5 (6 %)	8 (15 %)	0.5 (40 %)	< 3.9 μm

contributions of Ag, Au and Hg L X-rays. Some contributions of Si, Cl and Ca are visible. We assume that Ag is from the plate body (bulk), Au and Hg are present as thin surface layer and that Si, Cl and Ca are parasitic elements. GUPIXWIN was used for analysis using layered sample (one thin layer + bulk) option with iterative procedure. The results of the analysis are given in Table 2. They show that the Au/Hg layer (spectrum related to area 1) has the maximal thickness of about 670 μg/cm² with Au concentration of about 87 w% and Hg concentration of about 13 w%. Fig. 15 clearly shows the relationship between the three spectra and actual regions on the 2D image map. The analysis clearly demonstrates that the fire-gilding process was applied to create this decoration. Similar analysis was performed on three encircled C3 components with higher S contributions. Related spectra are shown in Fig. 15 (c). We can see from Fig. 15 that these spectra relate to the shield and mantle areas of the 2D image map. Direct comparison with 2D elemental maps (Fig. 11 (b-k)) shows that the shield and mantle areas are dark in colour and are supposed to be made by the Niello technique. It has been reported in the literature that Niello was used to decorate metalwork throughout the Roman empire using silver sulphide (Ag_2S).[51] In this case, we performed quantitative analysis using GUPIXWIN, assuming thick target. The reason is the presence of Ag in the bulk and in the supposed surface layer of Ag_2S. The results are also included in Table 2 and they show different concentrations of S and Ag. The highest concentration of S, related to encircled area 4, corresponds to the concentration of S in Ag_2S. In the case of the other two spectra, S concentration is lower. We therefore assume that the spectrum from the area 4 is from the Niello layer, and that it should be quite thick for sulphur K X-rays to yield the expected concentration according to the layer stoichiometry. We have further calculated that 99% of the S yield from the Ag_2S layer originates from the top 3.9 μm and therefore this should be the minimum thickness of the corresponding layer. The spectra from the other two encircled areas would correspond to thinner Niello layers as the Ag contributions in the related spectra increased.

4. Conclusion

Generally, PIXE exhibits higher sensitivity for lighter elements, whereas the efficient excitation of elements heavier than iron requires more energetic beams (more than 3 MeV) not so often available. Therefore, in the case of typical scanning PIXE analysis with single detection systems, the setup throughput is expected to be rather low for heavier elements. One way to increase sensitivity is the use of large area detectors (i.e. by the increase of the detector solid angle). Sensitivity for heavier elements imaging can alternatively be increased by adding low power and lightweight X-ray tube to Ion Beam Analysis setup. Simultaneous PIXE+XRF excitation mode can be very useful for many practical applications where elemental areal distributions along investigated areas are all what is needed. However, that excitation mode is not suitable for quantitative analysis due to quite different PIXE and XRF excitation mechanisms. In both PIXE+XRF or PIXE excitation modes, higher statistics spectra that would contain enough information for qualitative and/or quantitative analysis of major, minor and even trace elements can be deduced using multivariate analysis methods from individual pixel spectra collected during 2D scanning of objects under investigation. This kind of analysis can be performed on low statistics individual pixel spectra. This could be of particular importance for sensitive samples that could be damaged during long or intense ion beam irradiation.

In this work we demonstrate the usefulness of multivariate spectral analysis to PIXE+XRF and PIXE 2D spectral images. Three cases were investigated: (i) qualitative analysis of spectra with high number of counts per pixel in the full range of measured X-ray energies collected in PIXE+XRF mode; (ii) qualitative and semi-quantitative analysis of spectra with medium to low counts per pixel collected in PIXE mode (i.e. on the spectra having high sensitivity in the low energy X-ray region and low count rates at the high energy part of the spectra), and (iii) qualitative and quantitative analysis of spectra with low counts per pixel measured in PIXE mode.

In the first case, we analysed measured 2D spectral maps with high statistics individual pixel spectra of various pigments obtained by simulteneous PIXE + XRF mode. Multivariate analysis clearly separated characteristic spectra of different pigments. High statistics average spectra (Fig. 4) were obtained for each pigment, showing clearly major, minor and trace elements corresponding to particular pigments. t-SNE analysis clearly separated pigment mixtures from pure pigments.

In the second case, we performed 2D analysis of illuminations from the 15th century book using ion beam excitation (PIXE mode). The goal was to identifiy the pigments used by the author and to demonstrate the power of multivariate analysis compared to traditional analysis. Multivariate analysis was applied to the measured 2D spectral maps with typical individual pixel PIXE spectra that show high sensitivity at the low energy part of the spectra and modest to low sensitivity for the corresponding high energy part. t-SNE analysis clearly separated the spectra (Fig. 9, 10) from different sample regions (Fig. 8) and helped to identify the pigments used by the author.

In the third case, where we analysed 2D spectral maps with low statistics pixel spectra (obtained in the PIXE mode) of a decorated archaeological silver plate, multivariate analysis enabled to identify pixels with similar spectra that can be summed and used for quantitative analysis. Using this approach, we were able to identify the sample regions with similar layer thicknesses, and obtain the layer thickness and

elemental concentrations. In the actual case study, we characterized the identified gold layer and Niello decoration on the archaeological plate of Roman origin.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

IBM, SF, VD and DM acknowledge the support from the Croatian Science Foundation project Hi-REXS [Grant Number 9429].

References

1 A. Adriaens, *Spectrochim. Acta, Part B*, 2005, **60**, 1503–1516.

2 T. Calligaro, J. C. Dran, J. Salomon and P. Walter, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2004, **226**, 29–37.

3 J. Salomon, J.-C. Dran, T. Guillou, B. Moignard, L. Pichon, P. Walter and F. Mathis, *Appl. Phys. A: Mater. Sci. Process.*, 2008, **92**, 43–50.

4 J. S. Laird, C. G. Ryan, R. Kirkham, T. Satoh and A. Pages, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2017, **404**, 15–20.

5 J. L. Campbell, N. I. Boyd, N. Grassi, P. Bonnick and J. A. Maxwell, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2010, **268**, 3356–3363.

6 C. G. Ryan, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1995, **104**, 377–394.

7 L. Pichon, L. Beck, P. Walter, B. Moignard and T. Guillou, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2010, **268**, 2028–2033.

8 L. Pichon, T. Calligaro, Q. Lemasson, B. Moignard and C. Pacheco, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2015, **363**, 48–54.

9 J. L. Russell, J. L. Campbell, N. I. Boyd and J. F. Dias, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2018, **417**, 46–50.

10 K. Pearson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, **2**, 559–572.

11 H. Hotelling, *Journal of Educational Psychology*, 1933, **24**, 417–441.

12 I. T. Jolliffe and J. Cadima, *Philos. Trans. R. Soc., A*, 2016, **374**, 1–16.

13 J. MacQueen, *Proc. Fifth Berkeley Symp. Math. Stat. Probab.*, 1967, **1**, 281–297.

14 B. Vekemans, K. Janssens, L. Vincze, A. Aerts, F. Adams and J. Hertogen, *X-Ray Spectrom.*, 1997, **26**, 333–346.

15 S. Vogt, J. Maser and C. Jacobsen, *J. Phys. IV*, 2003, **104**, 617–622.

16 G. Sciutto, P. Oliveri, S. Prati, M. Quaranta, S. Bersani and R. Mazzeo, *Anal. Chim. Acta*, 2012, **752**, 30–38.

17 M. Sawczak, A. Kamińska, G. Rabczuk, M. Ferretti, R. Jendrzejewski and G. Śliwiński, *Appl. Surf. Sci.*, 2009, **255**, 5542–5545.

18 A. Deneckere, L. de Vries, B. Vekemans, L. Van de Voorde, F. Ariese, L. Vincze, L. Moens and P. Vandenabeele, *Appl. Spectrosc.*, 2011, **65**, 1281–1290.

19 A. Deneckere, B. Vekemans, L. Voorde, P. De Paepe, L. Vincze, L. Moens and P. Vandenabeele, *Appl. Phys. A: Mater. Sci. Process.*, 2012, **106**, 363–376.

20 T. M. Shanahan, J. T. Overpeck, J. B. Hubeny, J. King, F. S. Hu, K. Hughen, G. Miller and J. Black, *Geochem., Geophys. Geosyst.*, 2008, **9**, 1–14.

21 M. A. Rodriguez, P. G. Kotula, J. J. M. Griego, J. E. Heath, S. J. Bauer and D. E. Wesolowski, *Powder Diffr.*, 2012, **27**, 108–113.

22 S. Aida, T. Matsuno, T. Hasegawa and K. Tsuji, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2017, **402**, 267–273.

23 J. B. Ghasemi, M. K. Rofouei and N. Amiri, *X-Ray Spectrom.*, 2015, **44**, 75–80.

24 L. van der Maaten and G. Hinton, *Journal of Machine Learning Research*, 2008, **9**, 2579–2605.

25 B. Grabowski, W. Masarczyk, P. Głomb and A. Mendys, *Journal of Cultural Heritage.*, 2018, **31**, 1–12.

26 E. Pouyet, N. Rohani, A. K. Katsaggelos, O. Cossairt and M. Walton, *Pure Appl. Chem.*, 2018, **90**, 493–506.

27 D. R. Thompson, D. T. Flannery, R. Lanka, A. C. Allwood, B. D. Bue, B. C. Clark, W. T. Elam, T. A. Estlin, R. P. Hodyss, J. A. Hurowitz, Y. Liu and L. A. Wade, *Astrobiology*, 2015, **15**, 961–976.

28 Erik Swietlicki, N. P.-O. Larsson and Changyi Yang, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1993, **77**, 195–202.

29 B. L. Doyle, P. P. Provencio, P. G. Kotula, A. J. Antolak, C. G. Ryan, J. L. Campbell and K. Barrett, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2006, **249**, 828–832.

30 A. J. Antolak, D. H. Morse, P. G. Grant, P. G. Kotula, B. L. Doyle and C. B. Richardson, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2007, **261**, 470–474.

31 T. F. Silva, G. F. Trindade and M. A. Rizzutto, *X-Ray Spectrom.*, 2018, **47**, 372–381.

32 V. A. Solé, E. Papillon, M. Cotte, P. Walter and J. Susini, *Spectrochim. Acta, Part B*, 2007, **62**, 63–68.

33 PyMca, http://pymca.sourceforge.net/, (accessed December 2020).

34 A. G. Karydas, *Ann. Chim.*, 2007, **97**, 419–432.

35 C. G. Ryan, E. Clayton, W. L. Griffin, S. H. Sie and D. R. Cousens, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 1988, **34**, 396–402.

36 M. J. Crawley, *Statistical Computing: An Introduction to Data Analysis using S-Plus*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2002.

37 J. Maindonald and W. J. Braun, *Data Analysis and Graphics Using R – an Example-Based Approach*, Cambridge University Press, Cambridge, UK, 2010.

38 R. R. Sokal and F. J. Rohlf, *Biometry: The Principles and Practices of Statistics in Biological Research*, W. H. Freeman, New York, USA, 1994.

39 L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, Inc., Hoboken, USA, 1990.

40 M. Vidal and J. M. Amigo, *Chemom. Intell. Lab. Syst.*, 2012, **117**, 138–148.

41 J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik and B. Zupan, *Journal of Machine Learning Research.*, 2013, **14**, 2349−2353.

42 M. Toplak, G. Birarda, S. Read, C. Sandt, S. M. Rosendahl, L. Vaccari, J. Demšar and F. Borondics, *Synchrotron Radiat. News*, 2017, **30**, 40–45.

43 Orange Data Mining, https://orange.biolab.si/ (accessed December 2020).

44 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, *Proc. Second Int.*

Conf. Knowl. Discov. Data Min., 1996, **2**, 226–231.
45 J. L. Campbell, A. M. McDonald, G. M. Perrett and S. M. Taylor, *Nucl. Instrum. Methods Phys. Res., Sect. B*, 2011, **269**, 69–81.
46 G. M. Perrett, J. L. Campbell, S. Glasauer and R. Pardo, *X-Ray Spectrom.*, 2014, **43**, 359–366.
47 L. M. Smieska, R. Mullett, L. Ferri and A. R. Woll, *Appl. Phys. A: Mater. Sci. Process.*, 2017, **123**.
48 W. Malzer, O. Hahn and B. Kanngiesser, *X-Ray Spectrom.*, 2004, **33**, 229–233.
49 K. Trentelman, C. S. Patterson and N. Turner, in *Handheld XRF for Art and Archaeology*, eds. A. N. Shugar and J. L. Mass, Leuven University Press, Leuven, Belgium, 1st edn., 2012, pp. 159–190.
50 R. Viegas, N. Franco, L. C. Alves, M. T. Peña, E. Alves and V. Corregidor, *Microsc. Microanal.*, 2015, **21**, 156–157.
51 S. La Niece, *Antiq. J.*, 1983, **63**, 279–297.

