Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2026, **151**, 2563

Micro-SORS and machine learning for the non-invasive reference-free study of subsurface pigment degradation

A. Lux, *[a,b,c] J. Schlanz, [d] A. Botteon,[a] L. Monico, [e] P. Matousek, [c] R. I. Dima,[d] C. Conti [a] and P. Strobbia *[d]

In this study, we present the development of micro-spatially offset Raman spectroscopy (micro-SORS) methods and data analysis routines for the study of pigment degradation processes in the cultural heritage field, exploiting micro-SORS ability to non-invasively investigate the inner portions of turbid materials. The purpose of the study is to demonstrate an automated reference-free method to visualize through micro-SORS mapping the distribution of degradation both on and below the surface. The need arises from the handling of large datasets provided by micro-SORS mapping, which are often troublesome to analyse manually and usually require prior knowledge of the sample composition. Unaged and artificially aged painted mock-up samples were analysed with micro-SORS mapping, and conventional map reconstruction was compared with both supervised and unsupervised learning methods. Representative features in the micro-SORS spectra, able to distinguish unaltered pigments and degradation products, were automatically selected through machine learning techniques, revealing hidden patterns and correlations. Through the important spectral features (wavenumbers) and clustering analysis, quantitative micro-SORS degradation maps were created to identify degradation patterns also below the sample surface. Unlike previous studies that only use supervised or unsupervised learning, both are combined in this study to ensure the relevance of the selected spectral features and discover correlations among spectra through clustering techniques. This approach can be valid also for other scientific fields, such as forensic or biomedical, where data visualization and pattern identification are essential.

Received 7th November 2025,
Accepted 4th March 2026

DOI: 10.1039/d5an01176k

rsc.li/analyst

Introduction

Non-invasive analyses have become a necessity in many fields, as the possibility of avoiding damage or alteration to the analysed sample is particularly relevant for biomedical, forensics or heritage science applications. Spatially Offset Raman Spectroscopy (SORS),[1–3] among other non-invasive techniques[4,5] based on vibrational spectroscopy, is a powerful method to obtain non-invasive information of the subsurface molecular composition of optically turbid samples. It is able to retrieve deep Raman photons from the inner part of materials thanks to a physical separation between the collection and excitation spots.[6] Since photons generated from deep scattering events are more likely to travel further sideways than surface photons, they can be collected at a different portion of the sample surface away from the illumination area.[6] This can also be achieved in a defocusing setting, which means collecting spectra while gradually moving away from the sample surface. The overall intensity collected by the detection system decreases as the offset (or defocusing step) increases: however, the top layer signal decays more rapidly compared to the bottom layer, thus allowing the evaluation of the composition of the inner portions.

Micro-SORS[7–9] is a variant that couples SORS with microscopy, allowing a higher spatial resolution at the micrometre scale. Initially developed for cultural heritage applications,[10–12] it has rapidly expanded to other fields.[13–15] This technique is most suitable in heritage science, as samples are often precious artwork that require non-invasive measurements and possess complex structures, often comprising micrometre layers. Degradation processes in polychrome cultural heritage objects usually consist of environmental weath-

[a]*Institute of Heritage Science, National Research Council (CNR-ISPC), Via Cozzi 53, 20125 Milan, Italy. E-mail: alberto.lux@mail.polimi.it*
[b]*Sapienza University of Rome, Faculty of Literature, Department of Classics, Piazzale Aldo Moro 5, 00185 Rome, Italy*
[c]*Central Laser Facility, Research Complex at Harwell, STFC Rutherford Appleton Laboratory, UKRI, Harwell Oxford, OX11 0QX, UK*
[d]*Department of Chemistry, University of Cincinnati, 312 College Dr., Cincinnati, USA. E-mail: stroppbo@ucmail.uc.edu*
[e]*Institute of Chemical Sciences and Technologies "Giulio Natta", National Research Council, (CNR-SCITEC), Via Elce di Sotto 8, 06123 Perugia, Italy*

ering, resulting in diffusion of by-products within the original materials. They are typically challenging not only to prevent but also to study and mitigate, as they often involve heterogeneity of various types (*i.e.*, morphological or compositional), which render local point measurements unreliable to evaluate the overall extent of the degradation. In this scenario, imaging and mapping techniques permit the investigation to relatively large areas, providing a visualization tool to estimate the distribution of a certain compound. Micro-SORS coupled with mapping techniques extends the non-invasive visualization of degradation to the inner parts of the sample, providing Raman maps at different depths.

Moreover, multiple unknown degradation substances can be found during the investigation, which could create complex composite signals that greatly differ from literature references. Identifying and labelling correctly all characteristic bands can be even more demanding with large datasets, as we obtain in case of micro-SORS maps. In this case, machine learning can be very helpful, as it is able to identify hidden patterns and relationships among spectral features (in our case, we mostly consider wavenumbers) that a visual examination would likely miss. Machine learning techniques aid in pinpointing complex patterns that indicate subtle changes in material composition, which could be hidden by conventional data analysis methods. These methods are also able to effectively identify and classify correct categories of previously established labels, paving the way for the automatic definition of the analysed sample.

Machine learning has already been applied to cultural heritage for many purposes,[16,17] including automated annotation for metadata completion,[18] data structuring in geomatics,[19] automated trend detection for archaeological site discovery and preservation,[20] digital restoration of paintings,[21] original colour palette reconstruction[22] and assessment and 3D mapping of decay evidence in archaeological structures.[23,24] In this study, we present for the first time the coupling between micro-SORS mapping and machine learning techniques to create an automated and reference-free method able to distinguish the extent and stratigraphy of degraded and intact pigment. Unlike previous studies,[23] we employed a combination of unsupervised and supervised learning to validate the results.

Workflow

To demonstrate the proposed reference-free method to create the Raman maps representing the extent of degradation, we studied a pair of aged and unaged samples for two pigments, with unaged being the intact material and the aged being subjected to an artificial aging protocol. The samples have been analysed with *three methods*, as a step-by-step validation for

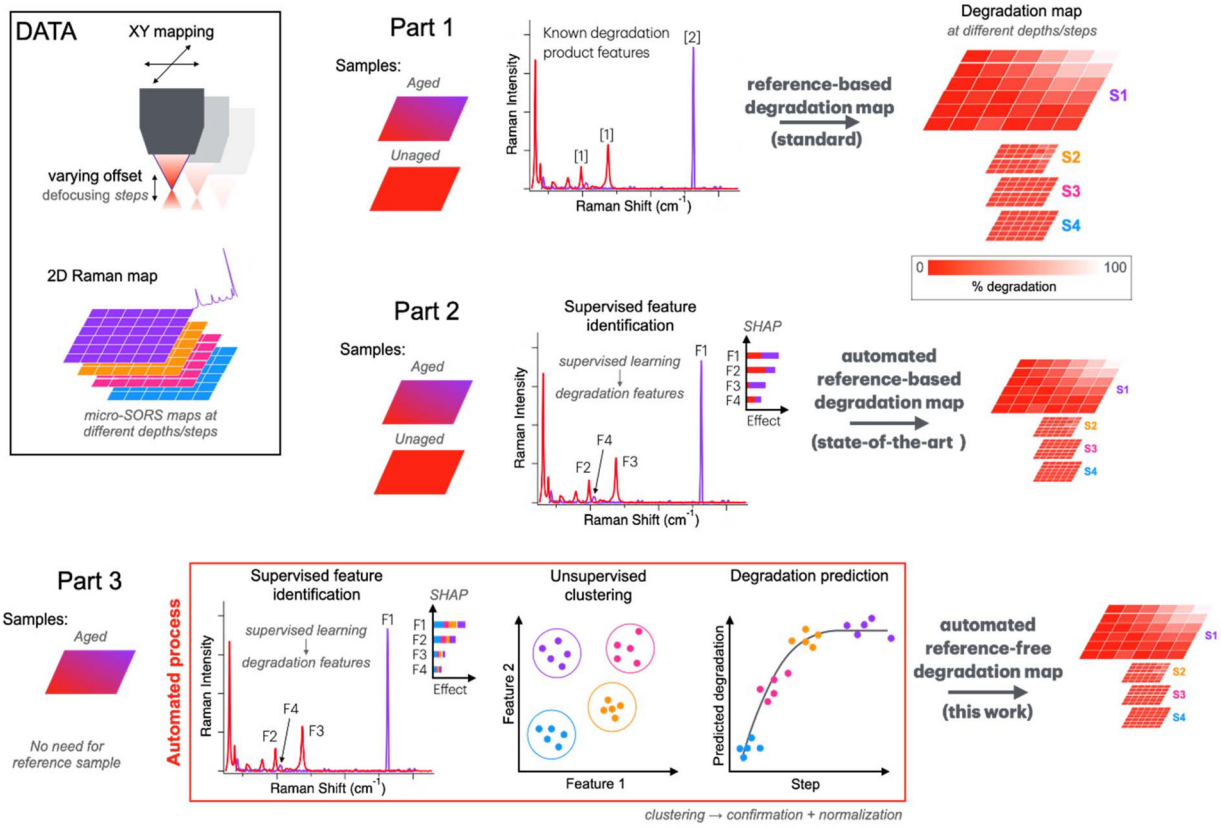

Fig. 1 Overview of the proposed methodology.

the use of machine learning to achieve reference-free degradation maps. In brief, the aged/unaged samples are first evaluated using maps that are based on known degradation-specific bands, retrieved from literature. Then, these characteristic bands are identified by a supervised machine learning analysis of differences between aged and unaged: therefore, the prior knowledge is given by the unaged sample itself. Finally, we focus on the aged sample only (no reference) and used the micro-SORS data to obtain the important features/bands. This process was first performed *via* supervised machine learning to validate and tune the model to separate pigments by degradation, followed by unsupervised learning to quantify degradation. Through a sufficient amount of algorithm tuning, the third part is fully capable of generating reference-free degradation map from micro-SORS data. Importantly, the combination of supervised and unsupervised learning is needed due to the absence of a clear reference. This specific workflow is shown below in Fig. 1 and discuss in further detail below.

The first part (Part 1) consisted in the collection of the micro-SORS maps of both aged and unaged samples, followed by the experimental micro-SORS map reconstruction *via* an in-house developed script. The maps were reconstructed by plotting the ratio of the intensities of the Raman bands representative of degradation (numerator) and of the intact material (denominator). Although this is an easily implemented tool, it requires prior knowledge of the sample to select the correct Raman bands. The second part (Part 2) exploits machine learning *via* supervised analysis of the surface of the unaged and aged samples. It compares the differences of their surface spectra, automatically providing the representative spectral features (as wavenumbers of key Raman bands) able to identify degradation and removing the need of retrieving them from literature. Using the supervised analysis on well-distinguished and labelled unaged and aged spectra allows the verification that the method can correctly identify differences between Raman spectra. However, in a real case scenario it is very unlikely to have access to the unaged counterpart of the sample to compare with the altered one. To address this issue, in the third part (Part 3) we applied the same supervised machine learning method (automatic identification of the relevant spectral features) to the subsurface micro-SORS datasets (defocusing steps), considering that the innermost portions of the sample are statistically more similar to the intact material than the surface ones. Therefore, the representative spectral features are selected through a comparison between the surface and the deepest micro-SORS map. However, because sub-surface degradation is not linear and not easy to parse, the supervised-feature model struggled to distinguish between different defocusing steps. Unsupervised learning was used to identify outliers and cluster together similar spectra in order to confirm the features associated with degradation and to build a model to output degradation level. The model was used to quantitatively map relative degradation. To demonstrate the accuracy of this process (Part 3), we compared the obtained features and the resulting maps with the ones obtained from Part 1. The specific workflow is shown in Fig. 1.

Materials and methods

Samples

Both sets of samples have been created in pairs on 25 × 25 mm squared PTFE supports, and then one element of the pair was subjected to an aging process (UVA visible light – $\lambda \geq 320$ nm, irradiance ~10^5 μW cm^{-2}, RH ≥ 95%, T = 40–45 °C, 23–26 days). The first pair of samples contained red lead (Pb_3O_4), a bright red-orange pigment, mixed with linseed oil, using a 4 : 1 pigment-to-binder weight ratio (Fig. 2a). To reflect the historical use of orpiment across various binding media from Antiquity through the 17th century,[25–28] the second set of samples was made up of orpiment (α-As_2S_3) mixed with Arabic gum in a 2 : 1 weight ratio of pigment to binder (Fig. 2b). More specifically, while arsenic sulphide was traditionally mixed with gum Arabic in Ancient Egypt,[25] it appears in combination with linseed oil in various 17th century paintings.[27,28] The sample comprising orpiment and linseed oil was prepared as well, but it revealed to be exceedingly fluorescent to obtain satisfactory results, and therefore we only kept the orpiment mixed with Arabic gum. Although this approach introduces additional variability among the samples (different binders for the two pigments), the use of diverse mock-ups enabled us to test the methodology on historically different systems, allowing the evaluation of its effectiveness across multiple scenarios.

We chose these pigments because they are frequently used in polychrome cultural heritage objects since antiquity, and are known to undergo chemical transformations when exposed to different environmental conditions of light and humidity, a phenomenon well-documented in literature.[26,28–32] Moreover, the aged samples presented different intricacies: the signal coming from the first one is strongly shielded by fluorescence, whereas spectra of the second sample present signals of multiple degradation products. Moreover, visual examination of the optical images (Fig. S1 and S2 in the SI) indicates that the two pigments have different degradation processes, which results in different spatial features: red lead shows heterogeneous degradation, with particularly affected

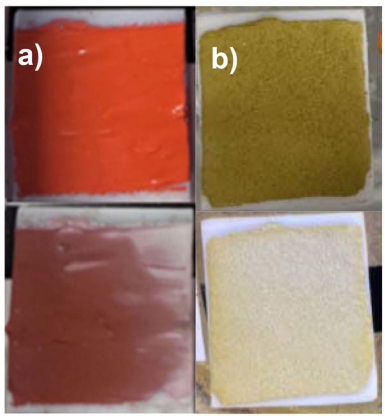

Fig. 2 Pictures of (a) red lead and (b) orpiment paint mock-ups before (top) and after artificial aging (bottom).

regions, whereas orpiment shows degradation spread over the whole surface.

Although two pairs of samples were analysed to have a solid basis to validate our method, for the sake of clarity and readability we chose to keep the orpiment sample in the main text, given its intrinsic complexity in terms of multiple degradation products, and the red lead sample in the SI. In any case, the analytical procedure is the same for both samples.

Micro-SORS protocol

The micro-SORS instrument is a Renishaw InVia Qontor prototype specifically adapted for micro-SORS measurements.[9] The instrument features a Renishaw CTX-1024A256-FDU Peltier cooled (−70 °C) NIR enhanced CCD camera, with a pixel dimension of 26 × 26 μm^2, and provides a spectral resolution of 1–2 cm^{-1}. A 785 nm excitation wavelength and ×20 microscope objective were used. Micro-SORS maps were acquired using the defocusing micro-SORS variant which consists in subsequent measurements at gradually incremented distances between the sample surface and the microscope objective:[2] first, an optical image of the samples was acquired (Fig. S1 and S2), over which we collected 507 spectra for each map, arrayed in a 39 × 13 grid over a 4256 × 1320 μm^2 optical image. Therefore, the analyses began by collecting an array of spectra on both the aged and unaged samples to generate Raman maps of the surface and subsurface levels. Four different defocusing depths were selected starting from the surface (0 μm defocusing distance, or 'imaged'). The defocusing steps for red lead samples were 75, 150, 225 and 300 μm; those of orpiment samples were 100, 200, 300 and 400 μm. The different choice in the defocusing steps was dictated by the different behaviour of the two samples: red lead signal drastically reduced after few defocusing steps, most likely due to stronger absorption, thus suggesting slightly shallower analyses; orpiment instead shows very efficient scattering even at high defocusing depth. The amount of defocusing steps was chosen balancing two factors: on one side, the number of probed depths has to be sufficiently high to be representative; on the other side it is important to select the correct spacing among defocusing steps to observe appreciable spectral differences. Since too many defocusing steps also create extremely large datasets that complicate the analyses, the number of defocusing steps was limited to five. In the case of red lead, we employed 0.5 mW laser power, with an integration time set to 20 s (0.5 s acquisition for 40 accumulations). For orpiment, 0.05 mW laser power was used, along with an integration time of 30 s (3 s acquisition for 10 accumulations). The spectral range of acquisition was set to 100–1350 cm^{-1}, since all their characteristic Raman bands fall within this range. All spectra were collected through WiRe software. In case of the orpiment sample and of the unaged red lead sample, the baseline correction was performed directly through the WiRe software (Intelligent Fitting, a polynomial-based algorithm); the aged red lead sample was extremely fluorescent and required a more precise approach, for which we used the SNIP algorithm[33] *via* an in-house developed script. Given the extremely high scattering efficiency of the orpiment pigment, the background contribution in the raw spectra was already minimal, and baseline correction had little effect on the final spectra.

The reconstruction of the Raman maps in Part 1 has been achieved *via* an in-house developed Python script, which was preferred from the automatic reconstruction provided by the WiRe software in order to have greater versatility and manipulation over the outcomes. The representative bands of the compounds of interest have been selected from literature, and an intensity ratio between these has been calculated for each spectrum in the map. The intensity ratios were then normalised to the highest one among all maps, creating an array of comparable values. These values were then plotted, creating an experimental heatmap that provides a visualization of the distribution and extent of decay product. It is important to notice that the spatial distribution of various pigments in the aged and unaged samples cannot be compared, since these were created separately, as mentioned before. Thus, they should be considered as individual samples, and comparison should be made only on the intensity of the signals of interest at a certain defocusing step.

Machine learning protocol

Micro-SORS spectra were characterized using supervised machine learning with the Pycaret library.[34] The purpose of the model in Part 2 was to distinguish between the surface (0 μm defocusing step) of the unaged and aged samples using their Raman spectral differences. The datasets were pre-processed using baseline correction and normalization made *via* an in-house developed script or through the WiRe instrument software. Each spectrum in the red lead sample dataset presented 807 wavenumber values (features) ranging from 119–1099 cm^{-1}, whereas orpiment sample comprised 232 features ranging from 150–450 cm^{-1}. In case of orpiment, shifts below 150 cm^{-1} were dropped to remove instrumental noise. Ten spectra chosen randomly from the unaged and aged datasets were kept aside as a holdout dataset to be used for the validation of the model. In order to prevent overfitting, the number of features was reduced through the feature selection methods described in the SI, only completed on the training set. The 3 remaining features for red lead and the 5 remaining features for orpiment were used to create the binary classification models to correctly classify aged and unaged spectra. The tuned and original Extra Trees Classifiers were chosen as described in the SI for red lead and orpiment, respectively. SHapley Additive exPlanations (SHAP) plots were created to analyse the model outputs.[35] The model also predicted the labels (Unaged or Aged) of the holdout dataset. A confusion matrix was calculated to test the validity of the final model.

Unsupervised learning using clustering was completed on all the data (including the holdout set) with 3 features for red lead and 5 features for orpiment to separate the unaged and aged samples based on degradation. Different methods, metrics, and cluster numbers were compared, but Euclidean and ward were selected with 11 and 9 clusters, respectively, relying on the Calinski-Harabasz, Silhouette, and Davies

Bouldin scores (Fig. S3). Features present in the peak used to normalize the data were not analysed further. Based on the intensities of the features (3 and 5) for each sample and the minimum and maximum average intensities across the clusters for each feature, the percentages of relative degradation were calculated and discussed in the SI.

Using similar methods and parameters as above, a multi-classification machine learning model was created for Part 3 to characterize the aged spectra from the 5 defocusing steps for each sample and to estimate the amount of degradation in the red lead and orpiment samples. The spectra were labelled with classes ranging from 0–4. The surface was labelled as class 0 and the largest defocusing step was labelled as class 4. After keeping aside 10 spectra from each class as a holdout set, the data consisted of 497 spectra from each class with 807 features ranging from 118–1100 cm^{-1} for red lead and 497 spectra from each class with 232 features ranging from 150–450 cm^{-1} for orpiment. Feature selection was completed using the training set as described previously. Tuned Linear Regression was chosen for red lead and tuned Light Gradient Boosting Machine classifier was chosen for orpiment as discussed in the SI. To analyse the outputs of the final model and determine the most important features, SHAP plots were created. The accuracy values were low due to the heterogeneity of each class, so the classes were further investigated using unsupervised learning.

Unsupervised learning using similar methods as above was completed on the defocusing step data with the 3 selected features for red lead and the 21 selected features for orpiment to separate the classes based on intensity and features that correspond to intact and degraded pigments. The percent degradation was also calculated as discussed in the SI. In the following sections, the surface and defocusing steps are referred to as classes, whereas the Raman bands associated with the degradation products will be named degradation peaks or degradation bands.

Results and discussion

Orpiment

Orpiment (α-As_2S_3) has a strong Raman scattering cross section intensity, with the main characteristic bands at 354, 311 and 294 cm^{-1} (Fig. 3). These Raman bands correspond to the different vibrational transitions: the ones localized around 300–400 cm^{-1} are related to the As–S bond stretching, whereas those that fall around 200 cm^{-1} are assigned to S–As–S bending modes. It is very common that the pigment powders can include other very similar compounds, such as other arsenic sulphides like realgar (As_4S_4), which has been found in the samples. Realgar can be clearly distinguished from orpiment through the bands at 220, 192 and 182 cm^{-1} (corresponding to As–As stretching vibrations) while the main peaks are strongly overlapped (355 and 354 cm^{-1}, respectively).[36–43] Interestingly, other compounds have been found in the intact sample: arsenolite (As_2O_3) with bands at 370 and 269 cm^{-1},

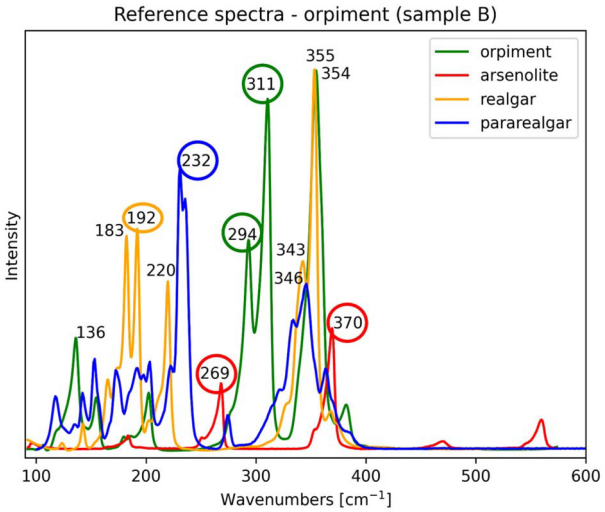

Fig. 3 Reference spectra of orpiment.

and pararealgar (As_4S_4) which can be distinguished through its intense band at 232 cm^{-1}; the first one forms from the oxidation of both realgar and pararealgar, whereas the latter is due to the light-induced transformation of realgar.[27,37,44] This confirms the high reactivity possessed by these pigments. In Fig. 4 the complete set of surface spectra of aged and unaged orpiment is shown (the rest are shown in Fig. S31). Realgar bands appear only in the unaged datasets, possibly indicating that after the aging process any trace of this pigment was transformed into pararealgar and then arsenolite (since pararealgar signal in the aged dataset is almost absent). In fact, arsenolite, clearly visible at 269 cm^{-1}, was the primary degradation product for the aged sample. It is worth noting that the orpiment signal at 354 cm^{-1} is on average less intense in the aged sample, since it was in part transformed into arsenolite.

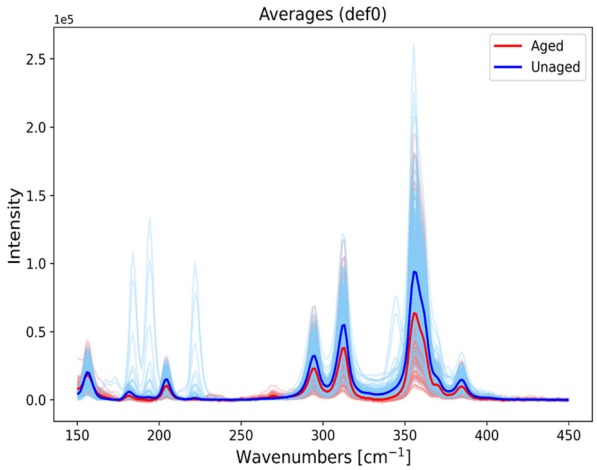

Fig. 4 Full set of surface spectra and their averages of both the aged and unaged orpiment.

Orpiment – Part 1. Micro-SORS measurements and experimental maps

In order to reconstruct the micro-SORS maps, we employed the secondary characteristic band of orpiment (311 cm^{-1} as the primary; 354 cm^{-1} is too close to the primary realgar band) as the denominator of the intensity ratio, since it identifies the intact pigment. As numerator, to account for the multiple degradation products in the sample, we employed different Raman bands for each map. More specifically, in case of arsenolite, we chose the band at 269 cm^{-1} since the primary one (370 cm^{-1}) is partially overlapped to the main band of orpiment, rendering its analysis troublesome. The arsenolite maps are shown below in Fig. 5, whereas the others (mostly discussing pararealgar and realgar presence) are shown in Fig. S32 and S33.

Starting from the right panels (Fig. 5b), the presence of arsenolite in the aged sample is straightforward on the surface map, and the ratio decreases as the defocusing increases. It appears that from the second defocusing step the amount of degradation reaches a plateau, as the intensity ratios vary less than before. This could possibly indicate that the penetration depth of the transformation of orpiment into arsenolite is rather shallow, and most is concentrated on the surface. In the case of the unaged sample, the left panels (Fig. 5a) also provide an expected result, since overall the presence of arsenolite is lower than in the aged sample. However, a very bright pixel is found at the second defocusing step, possibly indicating the presence of a cluster of arsenolite beneath the surface. It is quite particular to find a cluster of degraded substance within the material without having any hint of it on the surface: it is possible that during the defocusing measurement, the enlarging laser spot illuminated a cluster of degradation products on the surface, and it incorrectly assigned it to the 200 defocusing step. A more thorough investigation on this matter is presented in the SI.

Orpiment – Part 2. Unaged *vs.* aged orpiment (surface)

Supervised machine learning. The five features selected to compare the unaged and aged spectra – 368, 330, 292, 269,

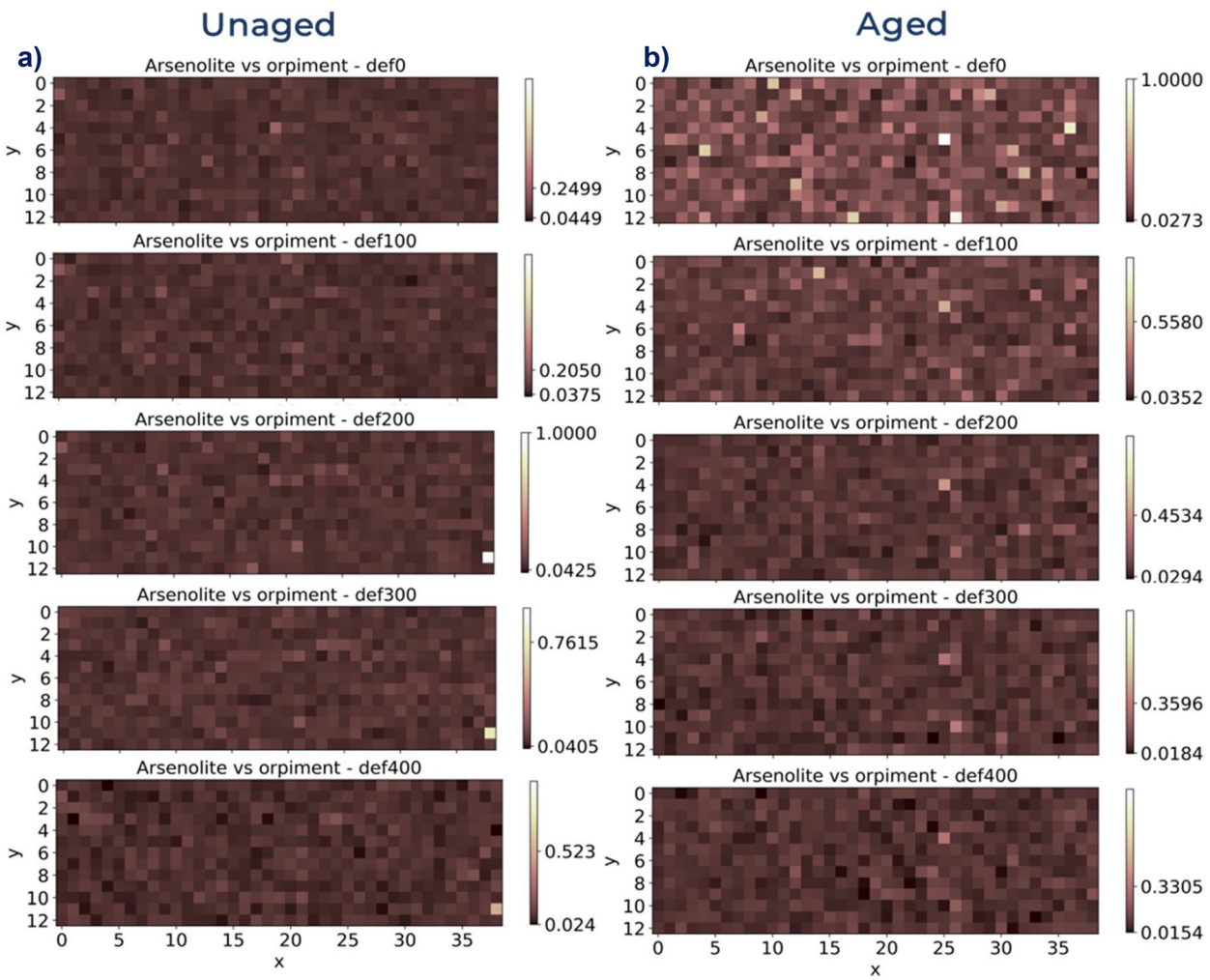

Fig. 5 Micro-SORS maps of the (a) unaged (left panels) and (b) aged (right panels) orpiment, plotting arsenolite *vs.* orpiment. The values of the colorbars indicate the maximum and minimum of each individual level with respect to the general scale.

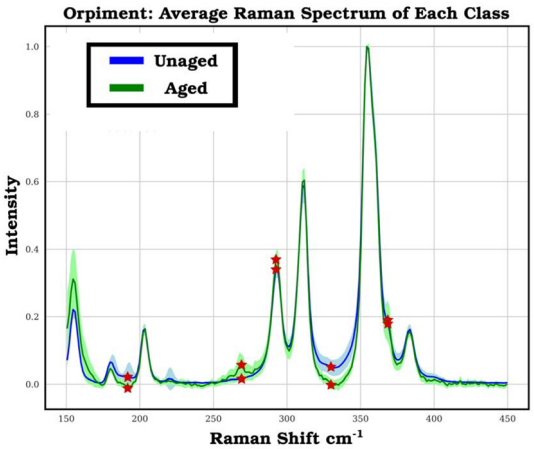

Fig. 6 The average unaged and aged spectra with the five selected features denoted as red stars.

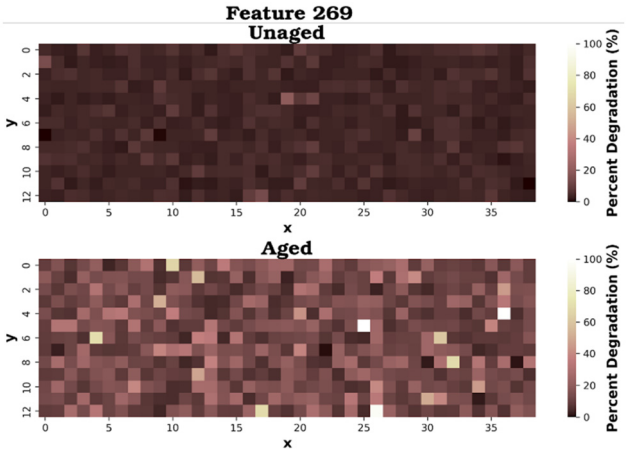

Fig. 7 The percent degradation of the unaged and aged orpiment samples using feature 269. A more degraded area of the sample is represented by white, and a more intact area of the sample is represented by dark red.

and 192 – are displayed in Fig. 6. Feature 292 corresponds to orpiment, and features 368 and 269 represent the degradation product, arsenolite. Feature 192 captures peaks corresponding to the intact realgar pigment. Using feature selection, features 368 and 269 corresponding to the degraded product arsenolite are automatically selected even when multiple intact and degradation peaks are present. The SHAP plot in Fig. S35a shows that feature 330 has the greatest impact on the model output. Interestingly, feature 330 does not correspond to an actual Raman band, but to a difference in the background signal between peaks 311 and 354. The difference is most likely stemming from the aging process, since the applied baseline removal was the same for both datasets. Although it is not clear as to how the aging process created this variation, this is particularly relevant, because only machine learning could pick up such a spectral difference and employ it to distinguish between the aged and unaged spectra, whereas a manual inspection would have probably ignored it. Including feature 330 along with the four other selected features also allows the Extra Trees model to correctly classify the 10 unaged and 10 aged spectra from the holdout dataset, as shown in the confusion matrix in Fig. S35b. Therefore, feature selection with machine learning can be used to identify important features and their corresponding compounds among the unaged and aged orpiment spectra.

Unsupervised learning. Unsupervised learning is used to further investigate the degradation across the surface of each orpiment sample. The average spectra for the nine clusters are shown in Fig. S36a, and it shows the presence of orpiment, arsenolite, and realgar. Since so few spectra contain pararealgar and realgar, they can easily become buried and averaged out by the signal of other pigments. Therefore, the pararealgar and realgar spectra are hard to identify manually within a large number of spectra. Clustering separates the spectra based on their intact and degradation products (Fig. S36b), which can make identifying minority pigments such as pararealgar and realgar easier. Detailed discussion over the clusters is provided in the SI (Fig. S37–S40).

Calculating degradation. Since pararealgar and realgar are not present in many spectra, a feature that represents them is not automatically selected. However, feature 269 is automatically selected and represents the degradation product, arsenolite. The percent degradation heatmap calculated using feature 269 is shown in Fig. 7. The degradation heatmaps calculated using the other four features are shown in Fig. S41. The 368 percent degradation heatmap is not used as a main reference since it overlaps with the large orpiment peak, and the other features do not correspond to a signal of a degradation product. However, as discussed previously, these features still allow the model to accurately predict whether the spectra are unaged or aged. Additionally, not all degradation products may be represented by feature selection, especially if there are a very low amount of degradation products present in the spectra. Three dark red (intact) cells (0 × 5, 7 × 22, and 11 × 34) and three white/yellow (degraded) cells (5 × 25, 4 × 36, and 6 × 31) from the aged heatmap were compared as discussed in the SI (Fig. S42). The experimental heatmaps (Fig. 5) and the heatmaps shown in Fig. 7 are very similar, as also shown by the comparison map in Fig. S43. The results show that unsupervised learning combined with feature selection can be used to separate spectra based on their degradation and allow for easier identification of minority intact and degraded pigments, especially with a complex sample like orpiment.

Orpiment – Part 3. Aged orpiment subsurface study

Machine learning. Since machine learning with feature selection was able to distinguish between the unaged and aged orpiment spectra using important intact and degraded features, we investigated whether it could distinguish between five different defocusing steps which vary in the degree of degradation. The 21 features used to make the final Light

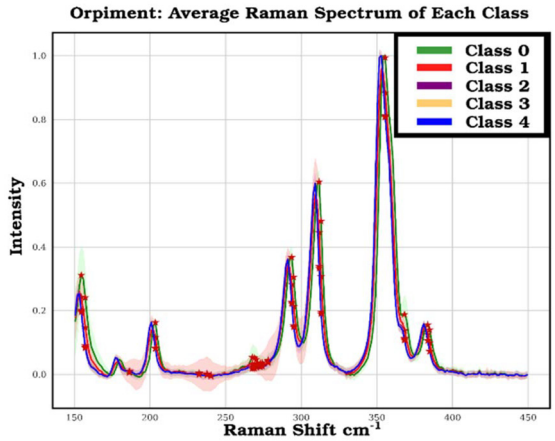

Fig. 8 The average aged defocusing steps for the five defocusing steps (classes 0–4) with the 21 selected features denoted as red stars. Class 0 corresponds to the surface and class 4 corresponds to the deepest defocusing step.

Gradient Boosting Machine classification model with the five defocusing steps are displayed in Fig. 8. Many features were needed for the orpiment defocusing steps since the spectra are very complex in terms of the amount of signals belonging to different compounds. More features were also needed for orpiment Part 3 compared to Part 2 since the spectra were obtained from the same sample, which caused them to be more similar and require more features for the model to distinguish them.

The top twelve features from the SHAP plot in Fig. 9 correspond to the intact and degraded pigments–orpiment, realgar, arsenolite, and pararealgar. Features 155, 157, and 295 represent orpiment. Feature 204 corresponds to both orpiment and pararealgar, and feature 356 corresponds to orpiment and realgar but is not used for further analysis since it is present in the peak used to normalize the data. The top feature in the SHAP plot corresponds to the orpiment feature 157, but it contributes little to class 0. However, the second most important feature (385, belonging to orpiment) contributes the most to class 0 and represents all the classes. Feature 384 most likely corresponds to the same peak, with a slight difference in wavenumber value due to statistical variation. Features around 235 are probably linked to the presence of traces of pararealgar, since it is in its characteristic spectral region. The SHAP plot also shows that feature 268 contributes about equally to all five classes, which is not observed for any other feature. This is coherent considering that features 268 and 368 represent the major degradation product, arsenolite. Features 238 and 186 are in a pararealgar and realgar peak, respectively, which shows that the important features are able to highlight peaks that do not appear important when analysing the average spectra. For example, features 238 and 186 do not appear to represent peaks in the average spectra since they correspond to the minority pigments that are not present in most spectra. Since these features are important using feature selection, it

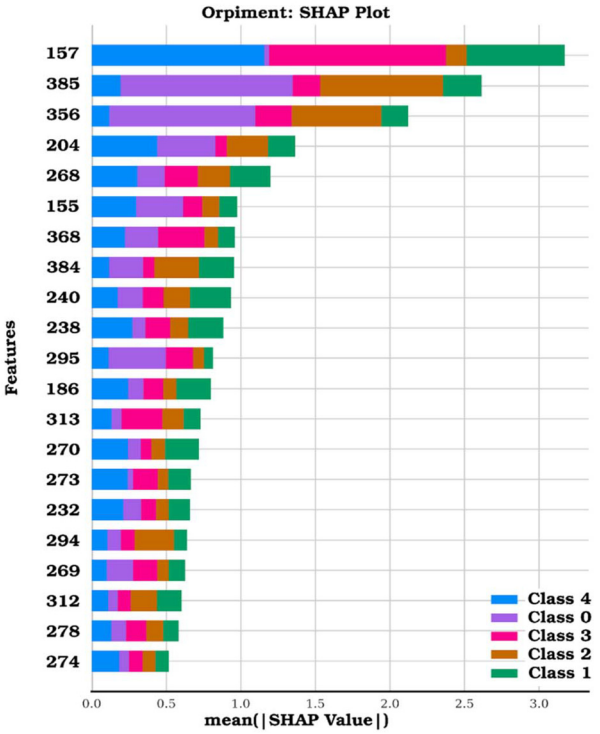

Fig. 9 SHAP summary plot of the orpiment defocusing steps (classes) using the Light Gradient Boosting Machine classification model.

hints that minority pigments may still have a relevant influence over the final outcome.

After the model with 21 features was created, it predicted the labels of the 10 spectra from each class in the holdout dataset. As shown in Fig. 10, class 0 is the only class that correctly predicted each spectrum. Most of the spectra are not being labelled the same as their true labels, especially those from classes 1–3. More than half of the spectra from class 1 are labelled as class 2, and the mislabelled spectra in class 4

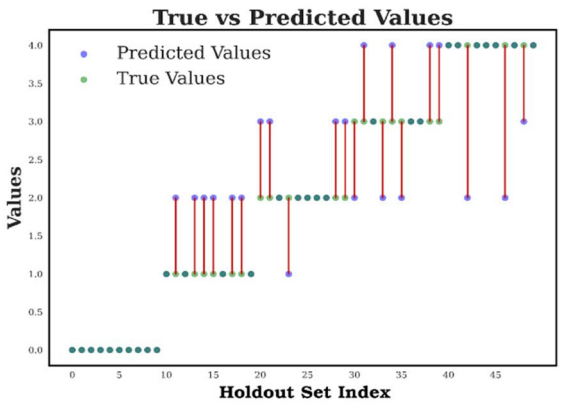

Fig. 10 The orpiment offset Light Gradient Boosting Machine classification model outputs using the ten spectra from each class of the holdout dataset. Each red line represents the difference between the true and predicted values.

are labelled as class 2 or 3. Among the mislabelled data in classes 2 and 3, almost all are labelled as class 2, 3, or 4. Therefore, at least the last three defocusing steps appear to be very similar in composition, which reveals that the degradation may begin to decrease and plateau after the first two defocusing steps, which is coherent with the preliminary observations made on the micro-SORS maps in Fig. 5 and in the SI. Additionally, the experimental micro-SORS map in the second defocusing step shows unexpected signals and a lot of variation in the intensity of some Raman bands, which might shed light on the large variation in the model's predicted labels for class 1. The presence of the realgar impurity, which can cause pararealgar to form, also contributes to the variation in spectral characteristics among the defocusing steps.

Unsupervised learning. Due to the heterogeneous natures of the five orpiment defocusing steps, especially the second, unsupervised learning exploiting hierarchical clustering was used to cluster the spectra. The clustering analysis is discussed in the SI in Fig. S44–S45. The distribution of the classes into clusters confirms the similarities in the last three classes, the large variability in the second class, and the heterogeneity of each defocusing step as discussed in the SI. Since the classes are heterogeneous with few areas containing realgar and pararealgar, these spectra become buried by the orpiment and arsenolite spectra, as discussed previously and in the SI. Therefore, unsupervised learning can be used after feature selection to visually analyse the spectra in greater detail. Using this method, similar spectra are grouped together, which can highlight outliers that contain unexpected signals, coming from compounds such as realgar and pararealgar, that may be missed upon manual inspection.

Calculating degradation. The percent degradation was calculated using feature 268 (Fig. 11) since it represents the main degradation product, arsenolite. The heatmaps created using the other 19 features and the overall percent degradation are shown in Fig. S46–S55. The degradation using feature 268 captures the heterogeneity in each defocusing step, with the surface being the most degraded. The areas of high degradation in class 0 (white cells at 5 × 25, 4 × 36, and 6 × 31) correspond with the same areas of high degradation found in the aged sample of Part 2. As the defocusing steps increase, the degradation should decrease. Indeed, class 0 appears to be highly degraded, and class 1 is less degraded than class 0. As expected, class 4 appears comparable to class 3, possibly indicating high similarity in composition, as suggested before. To investigate this degradation pattern, two very degraded (white) cells and two intact (reddish) cells were selected for further analysis as discussed in the SI (Fig. S56–58).

Another area of interest on the heatmaps in Fig. 11 and S48a is cell 6 × 36, which appears very degraded in the heatmaps in class 1 for features 268 and 238. Similar as discussed in the SI of Part 1, the presence of pararealgar and the noise amount appear to strongly influence the final visualization. Even though noise may affect the degradation calculation, the heatmaps are still able to highlight areas of high and low degradation. Unusual patterns in the heatmaps can also identify minority pigments, such as pararealgar. Therefore, it is important to use a combination of heatmaps that represent different features (and peaks) to gain a better understanding of how the pigment degrades as the defocusing steps increase. The comparison between the experimental heatmaps (Fig. 5) and the machine learning heatmaps (Fig. 11) is shown in Fig. S60.

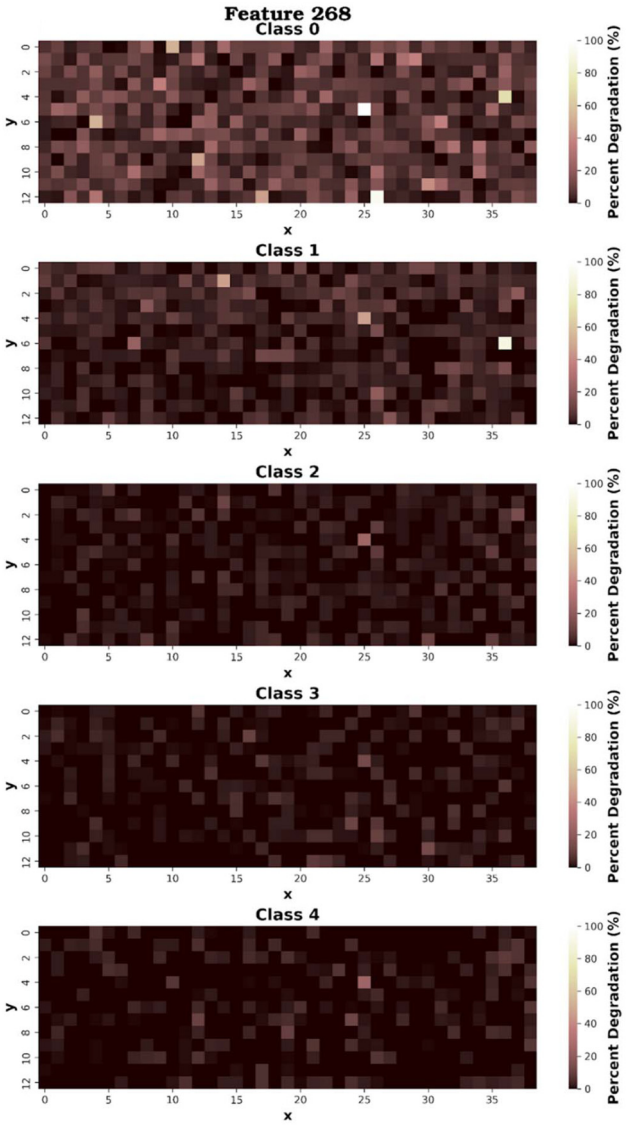

Fig. 11 The percent degradation of the orpiment defocusing steps using features 268. A more degraded area of the sample is represented by white, and a more intact area is represented by dark red.

Conclusions

We have employed micro-SORS mapping for the subsurface investigation of degradation products of pigments in mock-up samples, in combination with both supervised and unsupervised machine learning, to identify hidden patterns and corre-

lations. We were able to demonstrate that machine learning is capable of recognizing degradation patterns, invisible to the human eye, when comparing the spectral differences among surface and subsurface spectra, aiding the speed and reliability of the outcomes interpretation. The automatic reference-free analysis allows an accurate evaluation of the presence and distribution of degradation products, without the requirement of having prior knowledge over the analysed substance. This result was demonstrated by comparing the important features from Part 2–3 with the reference-selected features in Part 1. The ML-identified features agree with literature references of the degradation products, demonstrating that ML can be used for feature extraction with a reference-based and reference-free method. Similarly, the maps all display agreeing spatial features, further proving the efficacy of the automated methods. Although the model is not yet generalizable, as further research is needed to refine the method and mitigate some limitations (sensitivity to instrumental effects and noise, algorithm tuning, case-sensitivity, required technical knowledge), this work provides an efficient tool for the non-invasive investigation and automatic retrieval of information regarding the subsurface of materials, independently from any literature. By extending the proposed method to automatic *in situ* micro-SORS mapping of large areas, it could be possible to quickly obtain an overview of the compounds involved in their original state and environmentally altered conditions, both on and below the surface.

Conflicts of interest

There are no conflicts to declare.

Data availability

Experimental data and literature-based script is openly available in a Mendeley Data repository.

Lux, Alberto; Schlanz, Julie (2025), "*ML paper*", *Mendeley Data, V1*, DOI: https://doi.org/10.17632/mhthdcszw9.1.

The data analysis scripts of the article, including the output files, are available on GitHub at https://github.com/DimaUClab/Micro-SORS_Machine_Learning_Pigment_Degradation.git.

Supplementary information (SI) is available. All data relative to the red lead sample, and additional data regarding the orpiment sample, are presented in the Supplementary Information. See DOI: https://doi.org/10.1039/d5an01176k.

Acknowledgements

For the preparation and aging of paint mock-ups, Letizia Monico thanks the support of the Project PE00000020 CHANGES, funded by the European Union–Next Generation EU under the Italian Ministry of University and Research (MUR), NRRP, Mission 4, Component 2, Investment 1.3, CUP: B53C22003890006.

References

1 P. Matousek, I. P. Clark, E. R. C. Draper, M. D. Morris, A. E. Goodship, N. Everall, M. Towrie, W. F. Finney and A. W. Parker, *Appl. Spectrosc.*, 2005, **59**, 393–400.
2 S. Mosca, C. Conti, N. Stone and P. Matousek, *Nat. Rev. Methods Primers*, 2021, **1**, 22.
3 P. Matousek, M. D. Morris, N. Everall, I. P. Clark, M. Towrie, E. Draper, A. Goodship and A. W. Parker, *Appl. Spectrosc.*, 2005, **59**, 1485–1492.
4 P. Matousek, N. Everall, M. Towrie and A. W. Parker, *Appl. Spectrosc.*, 2005, **59**, 200–205.
5 C. Moffa, D. Francescone, A. Curcio, A. C. Felici, M. Bellaveglia, L. Piersanti, M. Migliorati and M. Petrarca, *Spectrochim. Acta, Part A*, 2025, **343**, 126510.
6 S. Mosca, P. Dey, M. Salimi, B. Gardner, F. Palombo, N. Stone and P. Matousek, *Anal. Chem.*, 2021, **93**, 6755–6762.
7 C. Conti, C. Colombo, M. Realini, G. Zerbi and P. Matousek, *Appl. Spectrosc.*, 2014, **68**, 686–691.
8 C. Conti, C. Colombo, M. Realini and P. Matousek, *J. Raman Spectrosc.*, 2015, **46**, 476–482.
9 A. Lux, C. Conti, A. Botteon, S. Mosca and P. Matousek, *Appl. Spectrosc.*, 2024, **79**(6), 1–14.
10 A. Rousaki and P. Vandenabeele, *J. Raman Spectrosc.*, 2021, **52**, 2178–2189.
11 A. Botteon, M. Vermeulen, L. Cristina, S. Bruni, P. Matousek, C. Miliani, M. Realini, L. Angelova and C. Conti, *Anal. Chem.*, 2024, **96**(11), 4535–4543.
12 M. Realini, C. Conti, A. Botteon, C. Colombo and P. Matousek, *Analyst*, 2017, **142**, 351–355.
13 C. Conti, M. Realini, C. Colombo, K. Sowoidnich, N. K. Afseth, M. Bertasa, A. Botteon and P. Matousek, *Anal. Chem.*, 2015, **87**, 5810–5815.
14 K. Buckley, C. G. Atkins, D. Chen, H. G. Schulze, D. V. Devine, M. W. Blades and R. F. B. Turner, *Analyst*, 2016, **141**, 1678–1685.
15 A. Botteon, W.-H. Kim, C. Colombo, M. Realini, C. Castiglioni, P. Matousek, B.-M. Kim, T.-H. Kwon and C. Conti, *Anal. Chem.*, 2022, **94**, 2966–2972.
16 M. Fiorucci, M. Khoroshiltseva, M. Pontil, A. Traviglia, A. Del Bue and S. James, *Pattern Recognit. Lett.*, 2020, **133**, 102–108.
17 M. B. Prados-Peña, G. Pavlidis and A. García-López, *J. Cult. Heritage Manag. Sustain. Dev.*, 2025, **15**, 664–686.
18 A. Belhi, A. Bouras, A. K. Al-Ali and S. Foufou, *J. Enterp. Inf. Manage.*, 2023, **36**, 734–746.
19 M. Cappellazzo, G. Patrucco and A. Spanò, *Heritage*, 2024, **7**, 5521–5546.
20 M. Casillo, F. Colace and M. Pellegrino, *J. comput. appl. archaeol.*, 2025, **8**, 224–241.
21 A. Kachkine, *Nature*, 2025, **642**, 343–350.
22 C. de Mecquenem, M. Eveno, M. Alfeld, T. Calligaro, E. Laval, K. Mösl and I. Reiche, *npj Heritage Sci.*, 2025, **13**, 1–11.

23 R. A. Galantucci, A. Musicco, C. Verdoscia and F. Fatiguso, *Int. J. Archit. Heritage*, 2025, **19**, 389–407.
24 S. Bruno, R. A. Galantucci and A. Musicco, *Vitruvio*, 2023, **8**, 6–17.
25 V. Daniels and B. Leach, *Stud. Conserv.*, 2003, **9**, 73–84.
26 L. Monico, S. Prati, G. Sciutto, E. Catelli, A. Romani, Q. Balbas, Z. Li, S. De Meyer, G. Nuyts, K. Janssens, M. Cotte, J. Garrevoet and G. Falkenberg, *J. Anal. At. Spectrom.*, 2022, **37**, 114–129.
27 F. T. H. Broers, K. Janssens, J. Nelson Weker, S. M. Webb, A. Mehta, F. Meirer and K. Keune, *J. Am. Chem. Soc.*, 2023, **145**, 8847–8859.
28 N. De Keyser, F. Broers, F. Vanmeert, S. De Meyer, F. Gabrieli, E. Hermens, G. Van Der Snickt, K. Janssens and K. Keune, *Sci. Adv.*, 2022, **8**, eabn6344.
29 M. Vermeulen, K. Janssens, J. Sanyova, V. Rahemi, C. McGlinchey and K. De Wael, *Microchem. J.*, 2018, **138**, 82–91.
30 F. S. Gard, D. M. Santos, M. B. Daizo, E. Freire, M. Reinoso and E. B. Halac, *Appl. Phys. A:Mater. Sci. Process.*, 2020, **126**, 1–12.
31 Y. Zhao, J. Wang, A. Pan, L. He and S. Simon, *Color Res. Appl.*, 2019, **44**, 790–797.
32 F. Vanmeert, G. Van der Snickt and K. Janssens, *Angew. Chem., Int. Ed.*, 2015, **54**, 3607–3610.
33 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.
34 A. Moez, *Pycaret: An open source, low-code machine learning library in Python*, 2020.
35 S. M. Lundberg and S. I. Lee, *Adv. Neural Inf. Process. Syst.*, 2017, 4766–4775.
36 K. Trentelman, L. Stodulski and M. Pavlosky, *Anal. Chem.*, 1996, **68**, 1755–1761.
37 A. Macchia, L. Campanella, D. Gazzoli, E. Gravagna, A. Maras, S. Nunziante, M. Rocchia and G. Roscioli, *Procedia Chem.*, 2013, **8**, 185–193.
38 E. Gliozzo and L. Burgio, *Archaeological and Anthropological Sciences*, 2022, **14**, 4.
39 P. Bonazzi, S. Menchetti, G. Pratesi, M. Muniz-Miranda and G. Sbrana, *Am. Mineral.*, 1996, **81**, 874–880.
40 A. Kyono, *J. Photochem. Photobiol., A*, 2007, **189**, 15–22.
41 E. Gliozzo and C. Ionescu, *Archaeol. Anthropol. Sci.*, 2022, **14**, 17.
42 W. Scheuermann and G. J. Ritter, *Z. Naturforsch., A*, 1969, **24**, 408–411.
43 R. L. Frost, W. N. Martens and J. T. Kloprogge, *Neues Jahrb. Mineral., Monatsh.*, 2002, **2002**, 469–480.
44 K. Keune, J. Mass, A. Mehta, J. Church and F. Meirer, *Heritage Sci.*, 2016, **4**, 1–14.