

High-resolution compound-specific mapping in works of art via data fusion of MA-XRPD with hyperspectral data (part 2: case studies)

Arthur Gestels^{a,b,*}, Francesca Gabrieli^c, Frederik Vanmeert^c, Steven De Meyer^{a,d},
Nouchka De Keyser^{a,b,c,e}, John Delaney^f, Koen Janssens^a, Gunther Steenackers^b

^a University of Antwerp, Department of Physics, AXIS Research Group, Groenenborgerlaan 171, B-2020, Antwerp, Belgium

^b University of Antwerp, Faculty of Applied Engineering, Department Electromechanics InViLab Research Group, Groenenborgerlaan 171, B-2020, Antwerp, Belgium

^c Conservation and Science Department, Rijksmuseum, Hobbemastraat 22, 1017 ZC, Amsterdam, the Netherlands

^d Paintings Laboratory, Royal Institute for Cultural Heritage (KIK-IRPA), Jubelpark 1, 1000, Brussels, Belgium

^e University of Amsterdam, Van 't Hoff Institute for Molecular Sciences, 1090GD, Amsterdam, the Netherlands

^f Scientific Research Department, National Gallery of Art, 6th and Constitution Avenue NW, Washington, DC, 20565, USA

ARTICLE INFO

Keywords:

X-ray diffraction
X-ray fluorescence
Reflectance imaging spectroscopy
Machine learning
Random forest regression
Cultural heritage

ABSTRACT

Background: This study is the second part of an investigation into enhancing high-resolution compound-specific mapping for artworks using multi-modal chemical imaging and machine learning. It builds upon the previously introduced framework, part one: method evaluation, by the authors that integrates macroscopic X-ray powder diffraction (MA-XRPD) macroscopic X-ray fluorescence (MA-XRF) or reflectance imaging spectroscopy (RIS). While MA-XRPD is time-consuming, the unique advantage of the technique lies in its ability to visualize the distributions of crystalline phases (e.g., pigments, degradation products) at the surface of artworks and estimate in a quantitative manner their relative concentrations. By leveraging data fusion and predictive modeling, this study aims to enhance the spatial resolution and efficiency of compound-specific mapping of artworks.

Results: This paper applies the previously established machine learning approach to three case studies, demonstrating its feasibility in analyzing heterogeneous, complex artworks. The first case study validates the method on a 17th-century Dutch still life painting, comparing the predicted high-resolution compound-specific maps against experimental data. The second case study examines the potential to reduce MA-XRPD scanning time through lower resolution scans and predictive modeling while maintaining accuracy, which is assessed using root mean square error (RMSE) and structural similarity index measure (SSIM). The third case study explores model transferability, evaluating whether predictive models trained on one artwork can accurately predict compound distributions in another. Results confirm the efficacy of the machine learning approach in improving spatial resolution and reducing data acquisition constraints.

Significance: This study addresses the time-intensive nature of MA-XRPD by demonstrating a scalable approach for generating high-resolution, compound-specific maps through data fusion and predictive modeling. By integrating the high specificity of MA-XRPD with the spatial efficiency of faster hyperspectral imaging techniques (MA-XRF and RIS), and applying a random forest regression framework, the method significantly enhances the efficiency and interpretability of chemical imaging in cultural heritage research, enabling more accessible and large-scale non-invasive analysis of complex artworks.

1. Introduction

Macroscopic X-ray powder diffraction (MA-XRPD) is a hyperspectral imaging technique suitable for non-invasive analysis of painted works of art. It yields highly specific distribution images of the crystalline compounds present at the surface of artefacts. Initially, MA-XRPD was only

available at synchrotron facilities or through self-built scanners, limiting its accessibility. However, the recent development of a self-built scanners has made the technique more widely available [1–14]. Despite its high specificity, MA-XRPD remains inherently time-intensive, often restricting its application to selected points or limited areas of an artwork rather than a full-object analysis. We previously introduced a machine learning

* Corresponding author. University of Antwerp, Department of Physics, AXIS Research Group, Groenenborgerlaan 171, B-2020, Antwerp, Belgium
E-mail address: koen.janssens@uantwerpen.be (A. Gestels).

<https://doi.org/10.1016/j.talanta.2025.128854>

Received 29 April 2025; Accepted 13 September 2025

Available online 23 September 2025

0039-9140/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

(ML)-based framework that combines MA-XRPD data with those from other hyperspectral imaging techniques such as macroscopic X-ray fluorescence (MA-XRF) and reflectance imaging spectroscopy (RIS) [15]. The study demonstrated how MA-XRPD in combination with complementary hyperspectral imaging techniques and random forest regression models can be used to obtain compound-specific distribution maps with improved spatial resolution and signal-to-noise content, effectively overcoming the main limitations of MA-XRPD, long acquisition times and/or a lateral resolution limited to approximately 1 mm.

In this paper the previously established machine learning framework is applied to three case studies that explore in more detail its potential for analysis of more complex works of art. The first case study involves the analysis of a 17th-century still life painting by Jacob Marrel (1613–1681), *Still Life with a Vase of Flowers and a Dead Frog*, from the collection of the Rijksmuseum. Unlike the counterfeit flower piece made in the style of a 17th-century Dutch old masters still life [16], and the 16th-century illuminated manuscript analyzed in Part 1, this old master painting provides an opportunity to qualitatively validate the ML-based methodology on the more complex multi-layered 17th-century painting technique. Furthermore, we investigate the capacity to predict compound distributions across the entire paint surface, by training models first on a small representative region of the painting. The second case involves a 15th-century South-Netherlandish panel painting, *Three Maries at the Sepulchre*, and explores the feasibility of reducing the total required MA-XRPD scanning time by employing scanning grids of increasing scanning step sizes, and thus reducing the amount of experimental data. Machine learning models are trained on these (generally) lower-resolution experimental MA-XRPD datasets to predict (higher resolution) compound distributions. The accuracy of the ML models is evaluated against the results obtained using the original (measured) data. To quantitatively assess model performance, the figures-of-merit introduced in Part 1, such as root mean square error (RMSE) and structural similarity index (SSIM) are employed. Additionally, histograms of the predicted and observed distribution maps are compared and examined to further evaluate the impact of size of the training set. The third case study examines the possibility for model transferability across different artworks. For this, two 17th-century still-life paintings by Jan Davidsz. de Heem (1606–1684), *Festoon of Fruit and Flowers* (Rijksmuseum Amsterdam, NL) and *Flowers and Insects* (Royal Museum of Fine Arts Antwerp, BE), are used to evaluate whether models trained on data from one painting can accurately predict pigment distributions in an artwork that was not included in the training data. Again, the RMSE and SSIM figures-of-merit are used to quantitatively assess the accuracy of the models.

This study builds on our earlier work by addressing key practical considerations for the routine application of MA-XRPD in the study of cultural heritage artefacts. To demonstrate the potential to upscale low spatial resolution yet highly compound-specific MA-XRPD data, we employ coarse scanning strategies combined with data-driven approaches. Through these methods, we aim to advance the feasibility of large-scale, non-destructive analysis of cultural heritage objects. These findings underscore the potential of machine learning in the field of heritage science, enabling more efficient, and interpretable chemical imaging workflows that bridge the gap between technical analysis and practical conservation needs.

2. Experimental methods

2.1. Artworks

This study examines three case studies involving four paintings to evaluate the effectiveness of the proposed methodology. The first case study focuses on *Still Life with a Vase of Flowers and a Dead Frog* (Fig. 1A) by Jacob Marrel, 40.3 × 31.0 cm, (Rijksmuseum Amsterdam, NL), dated



Fig. 1. Photographs of (A) *Still Life with a Vase of Flowers and a Dead Frog*, Jacob Marrel, 1634, Oil on Panel, Rijksmuseum Amsterdam, NL, (B) *Three Maries at the Sepulchre*, 15th century, attributed to Master of the Book of Hours of the Duc de Berry, Paper on Panel, private collection, (C) *Festoon of fruit and flowers*, Jan Davidsz. De Heem, 1660–1670, Oil on Canvas, Rijksmuseum. (D) *Flowers and Insects*, Jan Davidsz. De Heem, 1660–1670, Oil on Canvas, KMSKA. Scanned MA-XRPD (sub)areas are indicated in blue.

1634 and henceforth referred to as painting A. The second case study investigates a 15th-century religious painting (private collection), titled *Three Maries at the Sepulchre*, 40.0 × 25.0 cm, attributed to the Master of the Book of Hours of the Duc de Berry (Fig. 1B) and henceforth referred to as painting B. The third case study includes two 17th-century still life paintings by Jan Davidsz. de Heem: painting C, entitled *Festoon of Fruits and Flowers*, 74.0 × 60.0 cm, (Rijksmuseum Amsterdam, NL) and painting D, *Flowers and Insects*, 49 × 67 cm, (Royal Museum of Fine Arts Antwerp, BE). The two still life paintings by Jan Davidsz. De Heem were previously analyzed with results published elsewhere [8,17].

2.2. Measurement

2.2.1. MA-XRPD

MA-XRPD measurements were performed on the paintings using an in-house built MA-XRPD instrument. For Paintings A and B, the measurements were performed in reflection mode, employing a 50 W Cu-anode X-ray micro source. The measurements on paintings C and D were performed in transmission mode, employing a 44 W, Ag-anode X-ray micro source. Diffraction patterns were collected using a PILATUS 200 K area detector. Processing of the diffraction data was performed using the XRDUA software following a procedure that has been described elsewhere [2,4]. The measurement parameters for MA-XRPD analysis conducted on all paintings are summarized in Table 1.

2.2.2. MA-XRF

MA-XRF measurements of paintings A and C were collected with the commercially available MA-XRF scanner M6 Jetstream from Bruker Nano GmbH (Berlin, Germany). The M6 Jetstream consists of a 30-W Rh-target microfocus X-ray tube, and a 30-mm² X-Flash silicon drift detector [18]. The elemental distribution maps of the paintings A and C were collected with respectively a step size of 250 and 600 μm and dwell time of 120 and 50 ms/step. MA-XRF measurement of painting B (private collection) was collected with an inhouse built MA-XRF scanner (ATLAS) utilizing a 50-W Rh-anode X-ray tube in conjunction with four Vortex EX-90 SDD detectors to capture XRF signals. The elemental distribution maps of painting B were collected using a dwell time of 50 ms and a step size of 150 μm . MA-XRF measurement of Painting D (Royal Museum of Fine Arts Antwerp) was collected by another in-house built mobile MA-XRF scanner (AXIL scanner). It consists of a 50-W Mo-anode X-ray tube equipped with an SDD detector. The elemental distribution maps of the painting D were collected with step size 650 μm and dwell time 285 ms/step. The measurement parameters for MA-XRF analysis conducted on all paintings are summarized in Table 1.

All XRF spectral fitting were performed using the data analysis software packages PyMCA and Datamuncher [19,20].

2.2.3. RIS

Reflectance Imaging Spectroscopy (RIS) was employed to capture spectral reflectance data spanning the Visible and Near-Infrared (VNIR) and Short-Wave Infrared (SWIR) regions. Two specialized hyperspectral cameras facilitated high-resolution imaging, resulting in two RIS datasets: VNIR (Surface Optics Corp, 710E model with spectral range: 400–1000 nm) and long SWIR (M640 by Headwall with spectral range: 900–2500 nm).

2.3. Model architecture

The data registration and pre-processing procedure described in the Part 1 was followed [15]. Briefly, MA-XRF, and RIS datasets are downscaled and registered to the MA-XRPD dataset by using DatahandlerP [21]. For each of the identified compounds, a random forest

Table 1

Summary of the measurement parameters for MA-XRF and MA-XRPD analyses conducted on paintings A, B, C, and D. The table includes details on MA-XRPD acquisition mode, MA-XRF scanner instrument, step sizes and dwell times for each painting.

	MA-XRPD			MA-XRF		
	Mode	Step size (mm)	Dwell time (s pt ⁻¹)	Instrument	Step size (mm)	Dwell time (s pt ⁻¹)
Painting A	Reflection	1.2	10	M6	0.25	0.120
Painting B	Reflection	1.1	10	ATLAS	0.15	0.050
Painting C	Transmission	1.5	10	M6	0.60	0.050
Painting D	Transmission	1.0	15	AXIS	0.65	0.285

regression model was trained using preprocessed hyperspectral data and the corresponding MA-XRPD compound-specific map. The open source Python package “sci-kit learn” was used to implement the algorithm [22].

2.4. Model validation

2.4.1. Root mean square error (RMSE)

As defined in Equation (1) of Part 1 [15], the Root Mean Square Error (RMSE) continues to serve as a fundamental metric for evaluating the difference on a pixel-by-pixel basis between the predicted abundance maps and the experimentally determined MA-XRPD data.

2.4.2. Structural similarity index measure (SSIM)

To complement the RMSE, a figure-of-merit, the Structural Similarity Index Measure (SSIM) is employed. SSIM measures the similarity between two images based on luminance, contrast, and structural information, providing a perception-based assessment of image quality. Unlike the RMSE, which focuses on pixel-wise differences, SSIM accounts for spatial dependencies, making it less sensitive to minor shifts and more aligned with human visual perception. The SSIM is calculated using the following equation (2), resulting in values between −1 and 1 with 1 corresponding to a completely identical image:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

where μ_x and μ_y are the average of the image intensities x and y , σ_x^2 and σ_y^2 are the variances of the latter, σ_{xy} is the covariance of x and y . c_1 and c_2 are constants to limit the value of the ratio to a specific numerical range.

For detailed information regarding the investigated artworks, instrumentation specifications and preprocessing methods please refer to the Supplementary Information.

3. Results and discussion

3.1. Painting A: high-resolution compound mapping

As a first case study, the 17th-century still life painting, *Still Life with a Vase of Flowers and a Dead Frog* by Jacob Marrel (Fig. 2a) is investigated. In total 10 compounds could be identified by means of the MA-XRPD scans, 2 of which are shown in Fig. 2c, the others can be found in SI-2. 7 compounds commonly found in 17th century artworks were found: hydrocerussite ($2\text{PbCO}_3\cdot\text{Pb}(\text{OH})_2$), cerussite (PbCO_3), lead tin yellow type 1 (Pb_2SnO_4), calcite (CaCO_3), azurite ($2\text{CuCO}_3\cdot\text{Cu}(\text{OH})_2$), vermilion (HgS), and gypsum ($\text{CaSO}_4\cdot 2\text{H}_2\text{O}$). Together with 3 alteration products known to form as a result of degradation processes involving original compounds used in old master paintings: palmierite ($\text{K}_2\text{Pb}(\text{SO}_4)_2$), weddellite ($\text{CaC}_2\text{O}_4\cdot 2\text{H}_2\text{O}$) and lead palmitate ($\text{Pb}(\text{C}_{16})_2$, where C16 stands for $\text{C}_{15}\text{H}_{31}\text{COO}^-$). Palmierite can emerge through the interaction of lead-containing pigments with sulfur and potassium sources, the latter could originate from potash which was used as a substrate for lake pigments [23,24]. Weddellite typically forms via the oxidation of organic binders [25,26], while lead palmitate is a lead soap formed from reactions between lead-based pigments and fatty acids present in drying oils [27,28].

A resolution of 1.2 mm was used, which is sufficient for identifying the general compound distributions in this painting but lacks the detail necessary to resolve its finer features in the painting. As a result, the boundaries between different pigment areas become less distinct, leading to potential ambiguities in interpretation. To overcome this limitation, the coarse MA-XRPD maps are integrated with higher resolution MA-XRF data, using a step size of 0.25 mm, (20 maps, see SI-1, of which 3 are shown in Fig. 2b). This integration retains the high

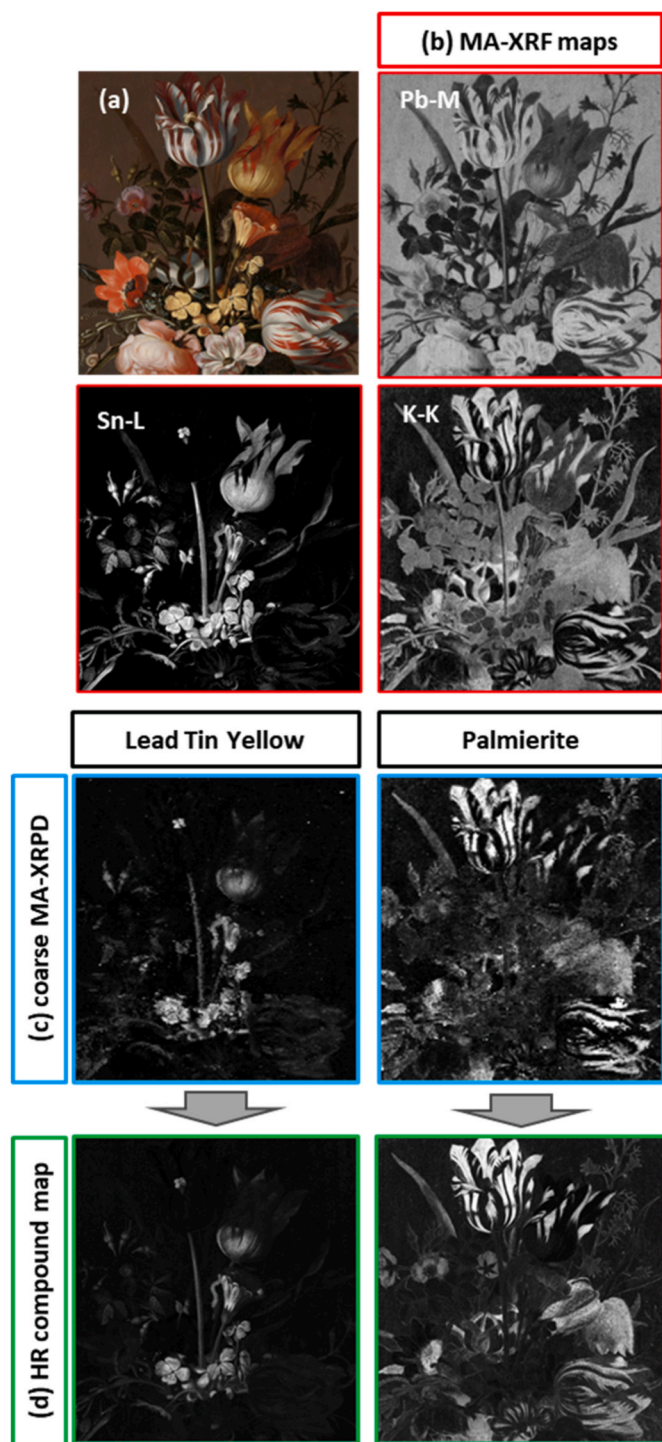


Fig. 2. (a) RGB image of the central part of the still life painting by Jacob Marrel. Grey scale images of the (b) high resolution elemental distribution from MA-XRF measurement, (c) experimentally recorded ('coarse') MAXRPD distributions and (d) predicted higher resolution ('HR') compound distribution maps. Lighter tones indicate a higher XRF intensity or XRPD abundance, corresponding respectively to a higher elemental or compound abundance in each pixel.

specificity of MA-XRPD while leveraging the high spatial resolution of MA-XRF, resulting in 10 high-resolution, compound-specific maps, of which two, lead tin yellow and palmierite, are presented in Fig. 2d.

For example, in the coarse MA-XRPD lead tin yellow map (Fig. 2c), it is difficult to clearly distinguish the different petals of the yellow flowers in the lower middle of the bouquet. In contrast, the high-resolution MA-

XRF maps allow for a much clearer visual interpretation of the flower piece.

The high-resolution (HR) compound maps in Fig. 2d demonstrate how fusing data from these two techniques preserves their individual benefits. The predicted HR lead tin yellow map accurately reflects the lower abundance of lead tin yellow in the darker areas on the surface of the painting while maintaining well-defined boundaries and finer details due to its high spatial resolution. This improvement is due to the fivefold smaller step size used during MA-XRF scanning compared to MA-XRPD. As a result, the integrated data provides a more precise distinction of the individual yellow petals in the center of the bouquet—details that are lost in the coarser MA-XRPD scans. A similar trend is observed for palmierite. While palmierite is present in the central bouquet, it is challenging to visually distinguish the separate flowers in the coarse MA-XRPD distribution. In contrast, the predicted HR palmierite map aligns more clearly with the bouquet's structure, making its distribution easier to interpret.

Furthermore, after the model is trained on a small, representative part of the painting (full blue square shown in Fig. 3k), by utilizing the full MA-XRF dataset they can be used to predict compound-specific distributions beyond the training region. For this purpose, the quality of the model was assessed by looking at the dead frog on the bottom right of the painting (dashed blue square shown in Fig. 3k). This area, while also scanned using MA-XRPD, was not part of the training data. Fig. 3 presents HR compound-specific maps (i) of the area around the dead frog and (ii) of the entire painting; here to maintain focus only the lead-containing compounds present in painting A are shown, highlighting the ability to differentiate between lead-based pigments, something that is not possible with MA-XRF alone. While the HR predictions from the models closely resemble the coarse MA-XRPD data for hydrocerussite, cerussite, lead tin yellow, and palmierite. The coarse palmierite map (Fig. 3i) contains some intense pixels which are the result of a fitting artefact during the MA-XRPD data processing, however with the HR prediction (Fig. 3j), these pixels are not present. On the other hand, the prediction for lead palmitate is less reliable. No lead palmitate could be found in the area around the dead frog using the coarse MA-XRPD data, yet the predicted high-resolution distribution map indicates its presence throughout this region. This discrepancy stems from the initial training dataset, i.e. the flower centerpiece, where lead palmitate was incorrectly identified in the background at relative concentrations around 5 %. As shown in the corresponding lead palmitate map in the training set (see SI), the fit is poor, and closer inspection reveals that lead palmitate is in fact not present in that region. Since the model was trained on this data, it learned to associate certain spectral features with lead palmitate at a concentration of 5 % or more. As a result, it falsely predicts the presence of lead palmitate in the area around the dead frog. This example illustrates a key limitation of the method: if the model is trained on flawed or unrepresentative data, it can propagate those inaccuracies and produce misleading predictions.

This first case study highlights the potential, but also limitation of the algorithm in situations for which complete scanning of the painting with MA-XRPD is not feasible, e.g. due to time restrictions.

3.2. Painting B: balancing data sparseness with prediction accuracy

The second case study examines painting B (Fig. 1B), depicting the biblical scene of the *Three Maries at the Holy Sepulchre*. Using MA-XRPD imaging, compounds consistent with 15th century South Netherlandish painting could be identified such as vermillion, lead tin yellow (Type 1), lazurite, hydrocerussite and cerussite. The maps of the lead white compounds, i.e., hydrocerussite (HCer) and cerussite (Cer), show the use of multiple types of lead white to create this panel. Previous literature has already shown that artists such as Matthias Grünewald, Leonardo Da Vinci and Johannes Vermeer used different subtypes of lead white in their paintings [6,29–31]. Previously, De Meyer et al. employed MA-XRPD to investigate the different types of lead white in *Girl With the*

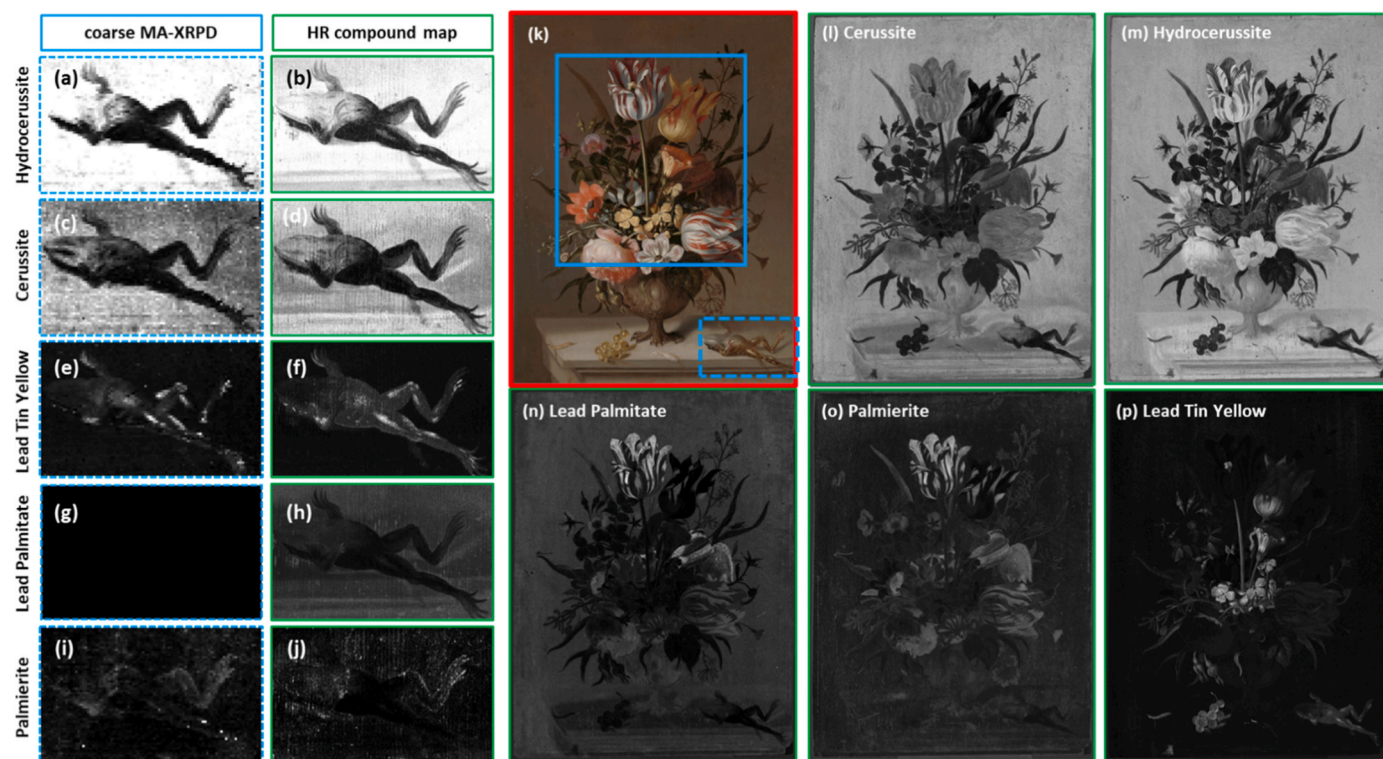


Fig. 3. (a)–(e) Predicted higher resolution (‘HR’) compound distribution maps of the entire paint surface resulting from the trained algorithm. (f) the entire painting, red the MA-XRF scanned area, and full blue square indicates MA-XRPD scanned area used for training, dashed blue square indicates MA-XRPD scanned area used for testing.

Pearl Earring by Johannes Vermeer [6]. The different types can be distinguished from each other by calculating the relative abundance r_{HCer} of HCer within the lead white as $r_{\text{HCer}} = m_{\text{HCer}} / (m_{\text{HCer}} + m_{\text{Cer}})$ by using the relative mass fraction (m) of HCer and Cer obtained from the MA-XRPD fitting procedure. The same method was used here to gain a better understanding of the lead white types present in painting B; r_{HCer} was calculated in every pixel, excluding the pixels showing a low abundance of lead white (i.e., when m_{HCer} and m_{Cer} are lower than 3 %). The resulting ratio distribution image and its histogram are shown in Fig. 4a and e. Fig. 4e indicates very clearly that six distinct lead white subtypes are present in this painting, characterized by HCer contents between 10 and 100 % (see below).

While MA-XRPD provides highly specific compound distributions, the fairly long acquisition time (typically 10 s) per measurement position can be limiting when large-scale areas need to be scanned. To address this, we investigate whether reducing spatial resolution, i.e., by increasing the step size during scanning, can still yield reliable predicted compound maps. By systematically down sampling of a single, experimentally recorded, MA-XRPD dataset, we assess how much detail can be sacrificed while maintaining accurate compound distribution predictions. Down sampling was performed using kernel sizes of 2×2 , 5×5 , 10×10 , and 20×20 pixels and randomly selecting one pixel from inside the kernel. This corresponds to MA-XRPD maps recorded with a step size of resp. 2.2, 5.5, 11 and 22 mm. The models are trained separately for each increase in data sparseness using MA-XRF, VNIR, or SWIR data and subsequently used to predict the HCer and Cer distributions of the entire painting. The predictions were generated at the original (1.1 mm) resolution of the experimentally collected MA-XRPD data set. In the following paragraphs the experimentally collected MA-XRPD dataset will be referred to as ground truth. To evaluate and compare the prediction accuracy of all models, RMSE and SSIM values were calculated for the HCer and Cer distributions (Table 2).

Several observations can be made: first, compound distributions maps predicted using MA-XRF data yields better results than using RIS

data in both the VNIR or SWIR range: RMSE values are ± 2 % higher for both RIS datasets compared to the MA-XRF dataset. Furthermore, the SSIM values for the RIS datasets are close to 0.5 or lower during down sampling, indicating that RIS datasets do not represent the original ground truth, i.e., the experimentally collected MA-XRPD data set, as effectively as the MA-XRF equivalent. For example, the 2×2 down sampled MA-XRPD fused to VNIR or SWIR data yields comparable SSIM values to 10×10 down sampled MA-XRPD fused to MA-XRF data for the HCer distributions, and to the 20×20 down sampled MA-XRPD + MA-XRF for the Cer distributions. Similar to part 1 of this study, the VNIR dataset yields better results than the SWIR dataset.

Because the models trained with MA-XRF data yield better results, only their predicted r_{HCer} distribution images and corresponding histograms are calculated and shown in Fig. 4. In the top row the ratio distribution images and their histogram of the different MA-XRPD sparse training sets are shown. In the bottom row, the predicted ratio distribution images and their histograms are presented. Here the histograms are plotted on a consistent scale to ease interpretation. Each column displays the histogram of the ground truth r_{HCer} values (in grey), overlaid with a semi-transparent histogram (in red) representing the predicted r_{HCer} distribution.

In the original resolution MA-XRPD training data (Fig. 4a and e), six distinct lead white subtypes were identified. The lead white subtypes include three low-HCer varieties (10 %, 20 %, 35 % HCer) and three high-HCer varieties (55 %, 87 %, 100 % HCer). The 35 % HCer subtype is primarily found in the sky and area of the sepulchre, while the 20 % HCer variety is present in the grassland surrounding the grave. The 10 % HCer subtype was applied in the shadows of Jesus’ clothing, appearing in the histogram (Fig. 4e) as a shoulder on the 20 % HCer peak. Among the high-HCer varieties, the 55 % HCer subtype is located in the hills in the background and facial features of the Three Maries, while the 87 % HCer subtype is observed in the red clothing of the kneeling Mary. The 100 % HCer subtype is found in retouched areas scattered throughout the painting.

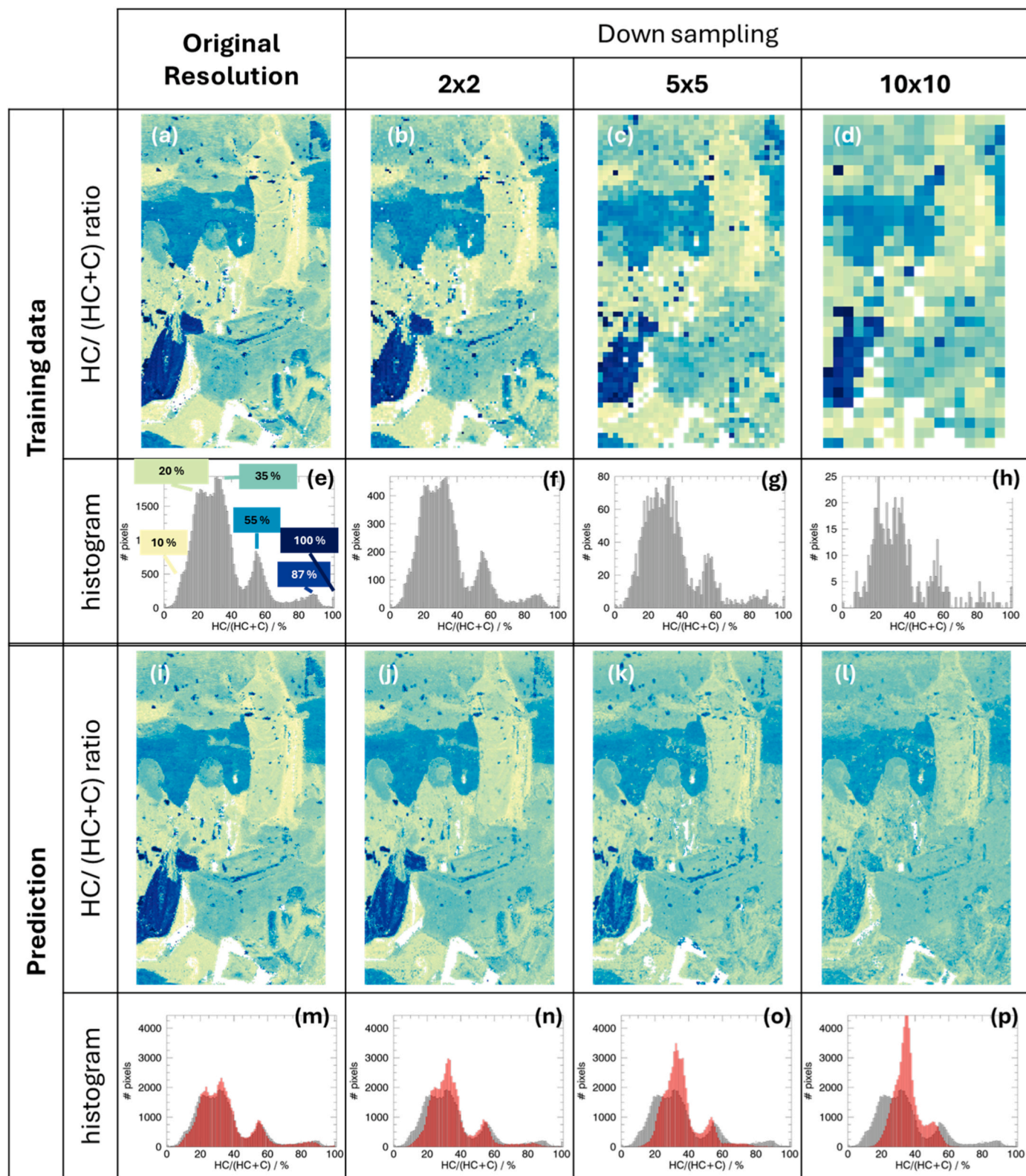


Fig. 4. (a to d) $m_{HCer}/(m_{HCer} + m_{Cer})$ ratios distribution images and (e–h) corresponding histograms with different spatial resolution obtained with reflection MA-XRPD: (a, e) original resolution, (b, f) 2×2 down scaling, (c, g) 5×5 down scaling, (d, h) 10×10 down scaling. (i to l) $m_{HCer}/(m_{HCer} + m_{Cer})$ ratios distribution images and (m–p) corresponding histograms (red) generated from predictions based on different spatial resolution datasets as training sets are plotted of the original/ground truth resolution histogram (grey): (i, m) original resolution, (j, n) 2×2 down scaling, (k, o) 5×5 down scaling, (l, p) 10×10 down scaling.

Table 2

Values of the RMSE and SSIM figures-of-merit obtained for Painting B at different levels of downsampling of the MA-XRPD data and after fusion with either MA-XRF, VNIR or SWIR data.

			Original resolution	Downsampling					
				2 × 2	5 × 5	10 × 10	20 × 20		
		# pixels	61710	15345	2442	594	144		
RMSE	HC	MA-XRF	5.27	9.03	13.25	15.96	16.90		
		VNIR	8.35	11.22	14.03	17.01	19.21		
		SWIR	8.77	11.57	14.99	17.96	17.12		
		MA-XRF	3.15	5.34	6.70	7.87	8.98		
	C	VNIR	5.74	7.53	9.06	10.35	11.12		
		SWIR	6.65	8.65	10.62	11.62	12.54		
		SSIM	HC	MA-XRF	0.93	0.78	0.68	0.58	0.46
				VNIR	0.73	0.59	0.47	0.40	0.21
SWIR	0.71			0.53	0.38	0.25	0.19		
C	MA-XRF			0.93	0.79	0.70	0.64	0.55	
	VNIR		0.74	0.58	0.45	0.33	0.26		
	SWIR		0.68	0.48	0.33	0.26	0.20		

The r_{HCer} histogram of the predicted maps at original resolution (Fig. 4m) closely mirrors the ground truth, with all six lead white types clearly distinguishable. However, the 100 % HCer subtype appears slightly reduced in HCer, with corresponding pixels displaying an average r_{HCer} value of ~95 % instead of 100 %, showing a slight over-estimation of cerussite in these regions. As expected, while the data becomes more sparse, subtle pigment distinctions are lost, and a gradual loss in detail can be observed when looking at the r_{HCer} distribution of the training datasets (Fig. 4b–d). At the 2 × 2 down sampling level, the main figures and retouched areas remain distinguishable in the MA-XRPD training dataset (Fig. 4b and f). Consequently, using the 2 × 2 down sampling trained model, all lead white types are predicted and detectable except for the 100 % HCer subtype, probably as it is under-represented in the training dataset. In Fig. 4f, it can be seen that the 100 % HCer subtype of the 2 × 2 training dataset contains fewer than 50 pixels, resulting in the loss of the model's prediction accuracy. Additionally, in the histogram of the 2 × 2 prediction (Fig. 4n) the contribution of the most abundant lead white type (with a r_{HCer} around 35 %) increases, and all the other lead white compositions shift slightly towards this 35 % value. For example, the histogram peak originally at $r_{\text{HCer}} = 87$ % is now slightly repositioned around $r_{\text{HCer}} = 82$ %. As mentioned above, prediction accuracy decreases with increased sparseness. This effect is evident in the predicted r_{HCer} map for the 5 × 5 down sampling model (Fig. 4k), where high- r_{HCer} regions, such as the flag, start to blend into the background. The loss of accuracy is further reflected in the histogram of the 5 × 5 prediction (Fig. 4o), where only four of the original six distinct lead white subtypes can still be identified. The 10 % HCer subtype is now merged with the 20 % subtype, appearing as a shoulder on the main peak at 35 %. The 55 % HCer subtype remains well represented, while the high-HCer subtypes have merged into a single broader distribution, resulting in a predicted group with a wide range of HCer concentrations centered around 70 %.

At 10 × 10 downsampling predictions, key details are lost, retouches are no longer visible, and distinctions between lead white types in Jesus' clothing vanish (Fig. 4l). In the histogram of the 10 × 10 prediction (Fig. 4p) only two lead white types remain identifiable (~37 % and ~50 % HCer), while the 10, 20, 87, and 100 % HCer subtypes are no longer reliably predicted due to their limited representation in the training set, with each subtype being represented by fewer than 15 pixels. As a result, the models struggle to accurately distinguish these subtypes. Given the limited pixel count (<15), the models try to minimize the errors by

predominantly predicting values close to the overall average.

This comparison emphasizes the dependency of the algorithm on having representative data for the different compound types to maintain prediction accuracy and variability. More (representative) data results in better fidelity to the original scan, while less data can lead to over-simplified, mean-centered predictions.

3.3. Painting C and D: assessing model transferability across different artworks

As a third case study, we investigated the ability of our ML model to predict compound distributions of an artwork not previously used to train the model. For this purpose, the two still life paintings by Jan Davidsz. de Heem shown in Fig. 1c and d were selected. Relatively small subareas of both paintings were analyzed by means of MA-XRPD in transmission mode. MA-XRF scans of both entire paintings were acquired using different dwell times per pixel and scanner configurations, each with a different detector solid angle, X-ray source and source parameters. The different parameters employed for MA-XRPD and MA-XRF data collection are summarized in Table 1. To predict high-resolution compound-specific maps for the totality of painting C, a predictive model was first trained using MA-XRF and MA-XRPD data from painting D. This model, built on MA-XRF and MA-XRPD data from the scanned subarea 1D (marked in blue in Fig. 1D), was then applied to Painting C. The process was also reversed: a model trained on MA-XRF and MA-XRPD data from subarea 2C of painting C was used to predict compound distributions across the entire surface of painting D. To allow for these operations, both MA-XRF datasets were normalized to remove (as much as possible) the effect of differences in XRF excitation and data-collection parameters. This was accomplished by thresholding the most intense outliers for each elemental MA-XRF distribution image and subsequently scaling the intensities for each elemental MA-XRF distribution image between 0 and 1.

Fig. 5, the top row shows 4 MA-XRF maps of the subareas of paintings C (Fig. 5a–e) and D (Fig. 5f–j) that were used to train the models. In the middle row a selection of the corresponding MA-XRPD maps are shown for C (Fig. 5k–o) and D (Fig. 5p–t). In the bottom row, the predicted compounds for 2C are shown in Fig. 5u–y with the model trained on the dataset from subarea 1D and the predicted compounds for 1D are shown in Fig. 5z–8 with the model trained on the dataset from subarea 2C. Additional MA-XRF and MA-XRPD maps of both paintings can be found in the SI.

The azurite distribution is well predicted by both models (Fig. 5u and z) as the ground truth azurite distributions (Fig. 5k and p) visually correlate strongly with the Cu–K distributions (Fig. 5a and f) in both maps. However, model D outperforms model C, achieving a lower RMSE, indicating a more accurate prediction overall in subarea 2C. Despite this, the SSIM score of model D is lower (0.31) than that of model C (0.56), likely due to the presence of significant noise in the ground truth azurite data of subarea 2C. Since the predicted azurite distribution is smoother, it appears less similar to the noisy ground truth, leading to a lower SSIM score despite improved accuracy, as indicated by the lower RMSE. On the other hand, as model C is trained with noisy azurite MA-XRPD data from subarea 2C (Fig. 5k), it struggles to accurately predict regions with high azurite abundance in subarea 1D and instead predicts a more averaged distribution, as seen in Fig. 5z.

A qualitative comparison of the predicted hydrocerussite distributions in subareas 2C and 1D with the training data shows that the former (Fig. 5v and a) were (mainly) obtained based on both the Pb–L and Pb–M MA-XRF distributions. They exhibit the strongest correlation with the Pb–L XRF maps (Fig. 5b and g), particularly in the areas where low As–K signals (Fig. 5d and i) are observed. As previously noted by Vanmeert et al., hydrocerussite is depleted in regions containing arsenic compounds [8]. This trend, which is present in both paintings, is also correctly captured in the predictions of both models. However, differences in the Pb–L distributions between the paintings introduce further

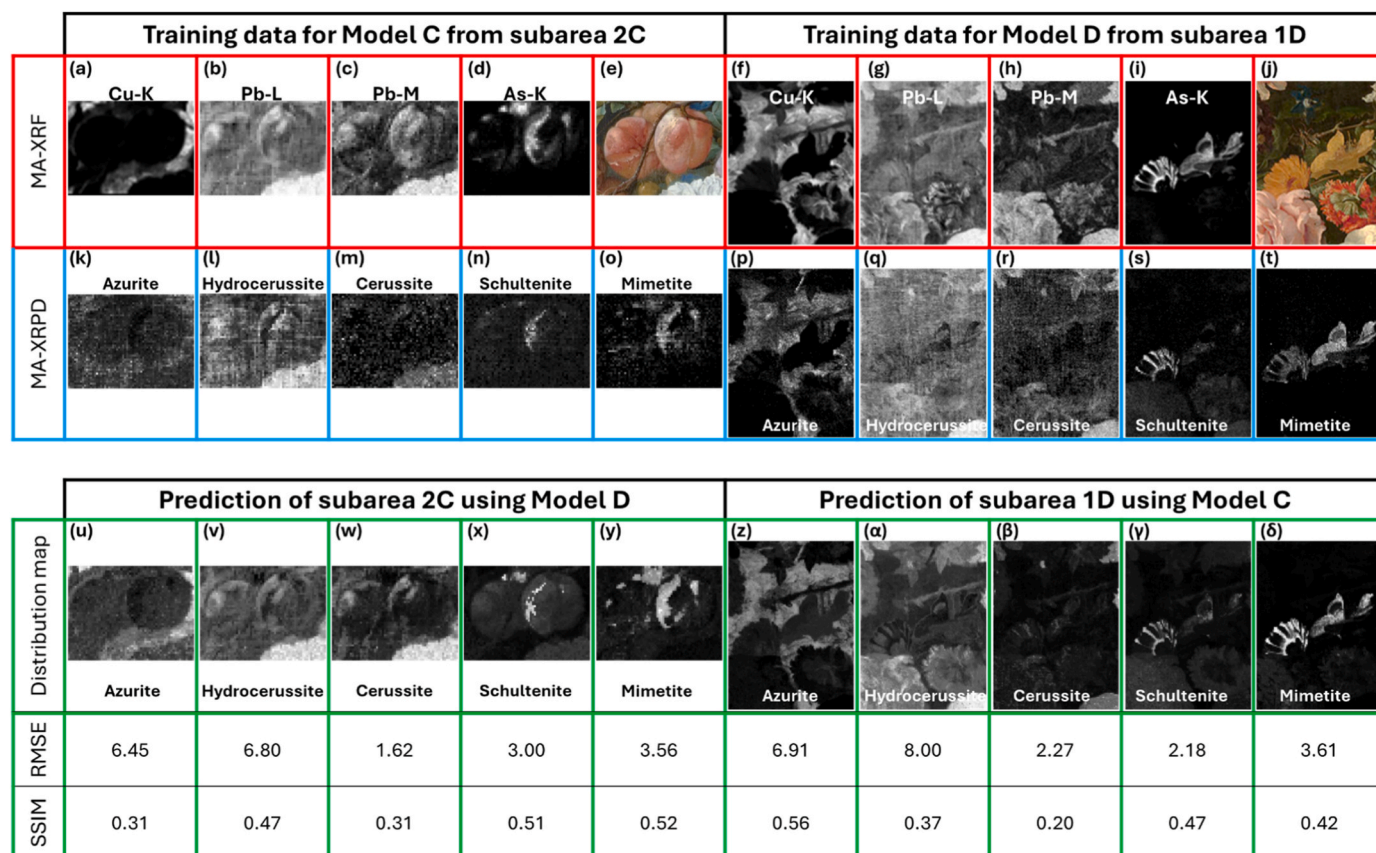


Fig. 5. Scanned subareas for painting C and D and their elemental distribution maps (red), ground truth XRPD data (blue), and predicted distribution maps (green). Top panel: training data for the models. Bottom panel: resulting predictions of Model D for subarea 2C (left) and of Model C for subarea 1D (right). The predicted distributions of azurite, hydrocerussite, cerussite, schultenite, and mimetite are shown with the corresponding SSIM and RMSE values listed below each predicted map.

difficulties. For example, in painting C it can be seen that the high Pb-L signals correspond to high hydrocerussite signals, whereas this trend is not always observed in Painting D. The dark purple-bordeaux flowers (left-top) and the dark blue flower (top-middle) and the branch in subarea 1D (Fig. 5j) are high in Pb-L signals, but low in hydrocerussite abundance. Thus, these regions are wrongly predicted to have high hydrocerussite content in the hydrocerussite prediction (Fig. 5g) made by model C. Another difference is that in painting D, the Pb-L signals (Fig. 5g) are less intense in areas with green foliage, i.e., regions with higher Cu-K intensities (Fig. 5f), whereas this trend is not observed in Painting C. Consequently, model C fails to accurately predict the hydrocerussite distribution (Fig. 5α) in these parts of subarea 1D: it predicts lower hydrocerussite abundance in regions where lower Pb-L signals coincide with higher Cu-K signals. Additionally, the presence of vermilion, a red mercury sulfide pigment, in subarea 2C (see SI-7), used for the red hues of the fruits, introduces misleading correlations in model C. Since these red areas are present throughout the entire training set, the model erroneously associates (high) mercury MA-XRF signals with (high) hydrocerussite MA-XRPD signals. This leads to an overestimation of the hydrocerussite distribution (Fig. 5α) in the red-colored areas, such as the red flower in subarea 1D. By contrast, model D's prediction is noticeably more accurate, achieving a higher SSIM (0.47) and a lower RMSE (6.80) score compared to model C. However, the predicted distribution (Fig. 5v) is smoother and less noisy than the ground truth (Fig. 5l), which may explain why the SSIM score is still relatively low. This discrepancy could indicate that the model does not capture finer structural details, such as canvas texture or subtle variations in training set.

The predicted maps for cerussite in both subareas do not align well

with the ground truth, which is also reflected in the low SSIM scores (0.31 for model C/subarea 1D and 0.20 for model D/subarea 2C). Examining the training data provides insight into these discrepancies. Model C was trained on data where cerussite (Fig. 5m) is primarily detected in regions with high Pb-L (and to a lesser extent, Pb-M) intensity (Fig. 5b and c). Additionally, cerussite was identified throughout the entire map of subarea 2C, however its distribution contains a lot of noise. Consequently, the model incorrectly only associates the abundance of cerussite with high Pb-L intensity, rather than accurately capturing its true distribution in subarea 1D (Fig. 5r), where lower abundance of cerussite is more uniformly present alongside small regions with higher cerussite abundance. The inverse effect is observed in model D, as in subarea 1D the cerussite distribution (Fig. 5r) closely resembles the Pb-M map (and to a lesser extent that of Pb-L) (Fig. 5g and h) in the parts with low As-K signal (Fig. 5i). This highlights again the sensitivity of the model to accidental correlations in the training data that have no physical meaning.

Finally, the distributions of the arsenic-containing compounds are well predicted (Fig. 5x, y, γ and δ), demonstrating the models' capability to accurately infer compound distributions. Also here, model D performs better than C, as seen in its higher SSIM scores for mimetite and schultenite. For schultenite, the SSIM values are close to each other for both models, however the RMSE of model C (2.18) is also lower than for model D (3.00). For the prediction of the mimetite distribution in subarea 1D using model C, the resulting SSIM value (0.42) is slightly lower than that obtained using model D in subarea 2C (0.52). This can be explained by the fact that in subarea 2C, the most intense As-K pixels (Fig. 5d) are directly associated with the most intense mimetite pixels (Fig. 5o), which is not the case for subarea 1D (Fig. 5i and t). This further

emphasizes the importance of ensuring that training datasets are representative of the target painting.

The superior performance of model D across multiple pigment distributions can be attributed, in part, to its significantly larger training set (13,230 pixels vs. 2960 pixels in model C). The increased dataset size allows for better generalization, improved accuracy and reduced sensitivity to noise-related errors. These findings emphasize the importance of high-quality, large, and representative training datasets. Expanding the dataset and incorporating more diverse training regions will be crucial for improving the predictive accuracy and model robustness across different artworks.

Painting C was completely scanned using MA-XRF which allows us to apply the model trained on subarea 1D for predicting the MA-XRPD distributions across the entire paintings. The model trained on subarea 1D produced the most accurate results; therefore, the following discussion focuses on full-painting predictions based on this model. As noted earlier, the predictions for hydrocerussite and cerussite distributions were inaccurate, and therefore, they are not included in Fig. 6, which shows the predicted compound distribution maps for painting C as a whole (Fig. 6f, g and h). As expected, the predicted azurite distribution (Fig. 6h) closely follows the Cu-K distribution (Fig. 6d), consistent with the findings from the previously scanned and analyzed subarea. For the arsenic containing degradation compounds, several additional regions exhibit a high intensity in the predicted compound maps. This raises the possibility that mimetite and schultenite may be present in these areas, suggesting that a re-examination of both paintings with MA-XRPD would be valuable to confirm these predictions. In painting C, the sub-area of the lemon (subarea 1C in Fig. 1) was additionally scanned by means of MA-XRPD, revealing that the distributions of mimetite and schultenite closely align, further demonstrating the potential of this predictive approach.

4. Conclusions

This study builds upon previous work by evaluating the applicability of a machine learning-assisted MA-XRPD methodology across three distinct case studies. The results highlight both the strengths and limitations of the approach, providing insights into its scalability, predictive accuracy, and practical implementation for non-invasive chemical imaging in heritage science.

The first case study demonstrates that by integrating MA-XRPD with

the MA-XRF dataset, compound distribution maps can be extrapolated beyond MA-XRPD scanned areas. This finding underscores the potential of predictive models to extend material analysis across entire artworks, for example in cases where full scanning is impractical due to time limits or physical constraints, such as the presence of a painting's frame.

The results of the second case study show that while it is possible to reduce MA-XRPD data acquisition time by employing a large step size during scanning, this comes at the cost of prediction accuracy. This highlights a clear trade-off: while finer and longer scans offer better accuracy, coarser and shorter scans sacrifice the level of quantitative information present in the data set. It should be noted, however, that the trade-off strongly depends on the representativeness of the limited MA-XRPD data with respect to the diversity of compounds present throughout the entire artwork. Simply increasing the step size uniformly may significantly degrade prediction quality, but if data collection were adapted in a more targeted way, such as by focusing measurements in areas of greater chemical variability and avoiding redundant sampling in more homogeneous regions, it might be possible to reduce the amount of data collected without severely impacting model accuracy.

The third case study serves as a first steppingstone toward evaluating the external predictive capabilities of our models, assessing their ability to predict compound distributions in paintings that were not included in the training dataset. The results indicate that increasing the amount of training data improves prediction accuracy, but diversity in training samples is more beneficial than merely adding more data from similar regions of the training painting. Systematic prediction errors were observed for certain compounds, such as hydrocerussite and cerussite, due to the lack of variation in the training set of the model. These findings highlight the need for expanded and representative training data sets to improve model robustness or for an equivalent simulation model to generate training data with a high level of representativeness.

Despite these challenges, the integration of machine learning with MA-XRPD and other hyperspectral techniques presents a significant advancement in non-invasive heritage science methodologies. By reducing reliance on time-intensive MA-XRPD scans, this approach could offer a practical solution for large-scale compound mapping while maintaining analytical precision, i.e., when data from multiple works from a single artist is available, or when multiple datasets from a single large work are available. Additional research, employing more advanced machine learning techniques such as convolutional neural networks could further enhance prediction quality.

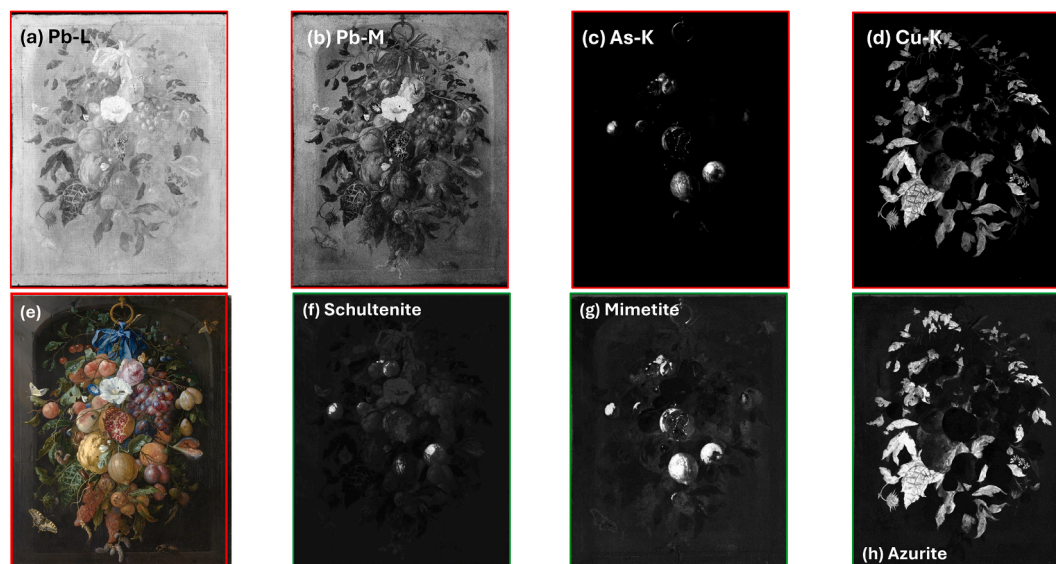


Fig. 6. (a–d) MA-XRF elemental distributions (red) of painting C, (e) RGB image of painting C and (f–h) the predicted higher resolution ('HR') compound distribution maps (green) of the totality of painting C based on models D.

CRedit authorship contribution statement

Arthur Gestels: Formal analysis, Methodology, Visualization, Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. **Francesca Gabrieli:** Formal analysis. **Frederik Vanmeert:** Formal analysis, Data curation, Writing – review & editing, Conceptualization, Resources. **Steven De Meyer:** Writing – review & editing, Formal analysis. **Nouchka De Keyser:** Formal analysis, Writing – review & editing. **John Delaney:** Supervision. **Koen Janssens:** Project administration, Writing – review & editing, Resources, Funding acquisition, Supervision. **Gunther Steenackers:** Funding acquisition, Resources, Project administration, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Arthur Gestels reports equipment, drugs, or supplies was provided by Netherlands National Museum. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors acknowledge financial support from FWO Brussels through grant from FWO project: A combined IR, NIR and MA-XRF material inspection method with FWO ID G056619 N and from Interreg Vlaanderen-Nederland (Smart*Light 2.0 project). Special thanks to the Rijksmuseum for this fruitful collaboration. We thank Olivier Deen and Frédérique Broers for insightful input regarding the topic.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.talanta.2025.128854>.

Data availability

Data will be made available on request.

References

- [1] F. Vanmeert, W. De Nolf, S. De Meyer, J. Dik, K. Janssens, Macroscopic X-ray powder diffraction scanning, a New method for highly selective chemical imaging of works of art: instrument Optimization, *Anal. Chem.* 90 (11) (2018) 6436–6444.
- [2] W. De Nolf, F. Vanmeert, K. Janssens, XRDUA: crystalline phase distribution maps by two-dimensional scanning and tomographic (micro) X-ray powder diffraction, *J. Appl. Crystallogr.* 47 (3) (2014) 1107–1117.
- [3] S. Legrand, F. Vanmeert, G. Van der Snickt, M. Alfeld, W. De Nolf, J. Dik, K. Janssens, Examination of historical paintings by state-of-the-art hyperspectral imaging methods: from scanning infra-red spectroscopy to computed X-ray laminography, *Heritage Science* 2 (1) (2014) 13.
- [4] F. Vanmeert, W. De Nolf, J. Dik, K. Janssens, Macroscopic X-ray powder diffraction scanning: possibilities for quantitative and depth-selective parchment analysis, *Anal. Chem.* 90 (11) (2018) 6445–6452.
- [5] F. Vanmeert, E. Hendriks, G. Van der Snickt, L. Monico, J. Dik, K. Janssens, Chemical mapping by macroscopic X-ray powder diffraction (MA-XRPD) of van Gogh's Sunflowers: Identification of areas with higher degradation Risk, *Angew. Chem. Int. Ed.* 57 (25) (2018) 7418–7422.
- [6] S. De Meyer, F. Vanmeert, R. Vertongen, A. Van Loon, V. Gonzalez, J. Delaney, K. Dooley, J. Dik, G. Van der Snickt, A. Vandivere, K. Janssens, Macroscopic x-ray powder diffraction imaging reveals Vermeers discriminating use of lead white pigments in Girl with a Pearl Earring, *Sci. Adv.* 5 (8) (2019) eaax1975.
- [7] J. Simoen, S. De Meyer, F. Vanmeert, N. de Keyser, E. Avranovich, G. Van der Snickt, A. Van Loon, K. Keune, K. Janssens, Combined Micro- and Macro scale X-ray powder diffraction mapping of degraded Orpiment paint in a 17th century still life painting by Martinus Nelli, *Heritage Science* 7 (1) (2019) 83.
- [8] F. Vanmeert, N. de Keyser, A. van Loon, L. Klaassen, P. Noble, K. Janssens, Transmission and Reflection Mode Macroscopic X-ray Powder Diffraction Imaging for the Noninvasive Visualization of Paint Degradation in Still Life Paintings by Jan Davidsz. de Heem, *Anal. Chem.* 91 (11) (2019) 7153–7161.
- [9] A. Gestels, G. Van der Snickt, J. Caen, G. Nuyts, S. Legrand, F. Vanmeert, F. Detry, K. Janssens, G. Steenackers, Combined MA-XRF, MA-XRPD and SEM-EDX analysis of a medieval stained-glass panel formerly from Notre Dame, Paris reveals its material history, *Microchem. J.* 177 (2022) 107304.
- [10] V. Gonzalez, I. Fazlic, M. Cotte, F. Vanmeert, A. Gestels, S. De Meyer, F. Broers, J. Hermans, A. van Loon, K. Janssens, Lead (II) formate in Rembrandt's Night watch: detection and distribution from the macro- to the micro-scale, *Angew. Chem.* (2023).
- [11] G. Van der Snickt, K. Janssens, J. Dik, W. De Nolf, F. Vanmeert, J. Jaroszewicz, M. Cotte, G. Falkenberg, L. Van der Loeff, Combined use of synchrotron radiation based micro-X-ray fluorescence, micro-X-ray diffraction, micro-X-ray absorption near-edge, and micro-fourier transform infrared spectroscopies for revealing an alternative degradation pathway of the pigment cadmium yellow in a painting by Van Gogh, *Anal. Chem.* 84 (23) (2012) 10221–10228.
- [12] F. Vanmeert, G. Van der Snickt, K. Janssens, Plumbonacrite identified by X-ray powder diffraction tomography as a missing link during degradation of red lead in a Van Gogh painting, *Angew. Chem. Int. Ed.* 54 (12) (2015) 3607–3610.
- [13] G. Festa, C. Caliri, M. Botticelli, C.G. Fatuzzo, E. Ferraris, J. Auenmüller, D. P. Pavone, G. Privitera, C. Scatigno, C. Miliani, F.P. Romano, Studying ancient Egyptian copper-alloy objects via X-ray diffraction and Machine Learning, *J. Cult. Herit.* 72 (2025) 48–58.
- [14] I.N. Cimento, The New Version of the Portable XRD System of the LANDIS Laboratory and its Application for the Non-destructive Characterisation of Pigments in Ancient Roman Frescoes, 2007.
- [15] A. Gestels, F. Gabrieli, T. De Kerf, F. Vanmeert, H.F. García, J. Delaney, K. Janssens, G. Steenackers, S. Vanlanduit, High-resolution compound-specific mapping in works of art via data fusion of MA-XRPD with hyperspectral data (part 1: method evaluation), *Talanta* 280 (2024) 126731.
- [16] F. Vanmeert, S. De Meyer, A. Gestels, E.A. Clerici, N. Deleu, S. Legrand, P. Van Espen, G. Van der Snickt, M. Alfeld, J. Dik, Non-invasive and non-destructive examination of artists' pigments, paints and paintings by means of X-ray imaging methods, in: *Analytical Chemistry for the Study of Paintings and the Detection of Forgeries*, Springer, 2022, pp. 317–357.
- [17] N. De Keyser, G. Van der Snickt, A. Van Loon, S. Legrand, A. Wallert, K. Janssens, Jan Davidsz. de Heem (1606–1684): a technical examination of fruit and flower still lifes combining MA-XRF scanning, cross-section analysis and technical historical sources, *Heritage Science* 5 (1) (2017) 38.
- [18] M. Alfeld, J.V. Pedrosa, M. van Eikema Hommes, G. Van der Snickt, G. Tauber, J. Blaas, M. Haschke, K. Erler, J. Dik, K. Janssens, A mobile instrument for in situ scanning macro-XRF investigation of historical paintings, *J. Anal. Atomic Spectrom.* 28 (5) (2013) 760–767.
- [19] V.A. Solé, E. Papillon, M. Cotte, P. Walter, J. Susini, A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra, *Spectrochim. Acta B Atom Spectrosc.* 62 (1) (2007) 63–68.
- [20] M. Alfeld, K. Janssens, Strategies for processing mega-pixel X-ray fluorescence hyperspectral data: a case study on a version of Caravaggio's painting *Supper at Emmaus*, *J. Anal. Atomic Spectrom.* 30 (3) (2015) 777–789.
- [21] M. Alfeld, S. Pedetti, P. Martinez, P. Walter, Joint data treatment for Vis–NIR reflectance imaging spectroscopy and XRF imaging acquired in the Theban Necropolis in Egypt by data fusion and t-SNE, *C. R. Phys.* 19 (7) (2018) 625–635.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [23] V. Gonzalez, A. van Loon, S.W. Price, P. Noble, K. Keune, Synchrotron micro-XRD and micro-XRD-CT reveal newly formed lead-sulfur compounds in Old Master paintings, *J. Anal. Atomic Spectrom.* 35 (10) (2020) 2267–2273.
- [24] A. Van Loon, A. Vandivere, J.K. Delaney, K.A. Dooley, S. De Meyer, F. Vanmeert, V. Gonzalez, K. Janssens, E. Leonhardt, R. Haswell, Beauty is skin deep: the skin tones of Vermeer's Girl with a Pearl Earring, *Heritage Science* 7 (2019) 1–20.
- [25] F. Ciarati, L. Rampazzi, L. Toniolo, A. Pozzi, Calcium oxalate films on stone surfaces: experimental assessment of the chemical formation, *Stud. Conserv.* 45 (3) (2000) 180–188.
- [26] A. Zoppi, C. Lofrumento, N. Mendes, E. Castellucci, Metal oxalates in paints: a Raman investigation on the relative reactivities of different pigments to oxalic acid solutions, *Anal. Bioanal. Chem.* 397 (2010) 841–849.
- [27] M.J. Plater, B. De Silva, T. Gelbrich, M.B. Hursthouse, C.L. Higgitt, D.R. Saunders, The characterisation of lead fatty acid soaps in 'protrusions' in aged traditional oil paint, *Polyhedron* 22 (24) (2003) 3171–3179.
- [28] M. Cotte, C. Emilie, D.N. Wout, T. Yokoi, D.V. Laurence, B. Manfred, W. Philippe, R. Camille, S. Murielle, J. Koen, J. Susini, Lead soaps in paintings: Friends or foes? *Stud. Conserv.* 62 (1) (2017) 2–23.
- [29] V. Gonzalez, S. Hageraats, G. Wallez, M. Eveno, E. Ravaut, M. Réfrégiers, M. Thoury, M. Menu, D. Gourier, Microchemical analysis of Leonardo da Vinci's lead white paints reveals knowledge and control over pigment scattering properties, *Sci. Rep.* 10 (1) (2020) 1–10.
- [30] E. Welcomme, P. Walter, P. Bleuet, J.-L. Hodeau, E. Dooryhee, P. Martinetto, M. Menu, Classification of lead white pigments using synchrotron radiation micro X-ray diffraction, *Appl. Phys. A* 89 (4) (2007) 825–832.
- [31] V. Gonzalez, G. Wallez, E. Ravaut, M. Eveno, I. Fazlic, T. Fabris, A. Nevin, T. Calligaro, M. Menu, V. Delieuvin, M. Cotte, X-ray and Infrared Microanalyses of Mona Lisa's Ground Layer and Significance Regarding Leonardo da Vinci's Palette, *J. Am. Chem. Soc.* 145 (42) (2023) 23205–23213.