A new approach to the interpretation of XRF spectral imaging data using neural networks

Sotiria Kogou1 | Lynn Lee2 | Golnaz Shahtahmassebi1 | Haida Liang1

1School of Science and Technology, Nottingham Trent University, Nottingham, UK
2The Getty Conservation Institute, Los Angeles, California

Correspondence
Haida Liang, School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK.
Email: haida.liang@ntu.ac.uk

Self-organising map (SOM), an unsupervised machine learning algorithm based on neural networks, is applied to introduce a novel approach for the analysis of XRF spectral imaging data. This method automatically reduced hundreds of thousands of XRF spectra in a spectral image dataset to a handful of distinct clusters that share similar spectra. In this study, we show how clustering and the combination of spatial and spectral information can be used to aid materials identification and deduce the paint sequence. The efficiency and accuracy of the method is presented through the analysis of a Peruvian watercolour painting from the Getty Research Institute collection. Confirmation of the interpretation was provided by complementary non-invasive techniques, such as optical microscopy, reflectance and Raman spectroscopies.

1 INTRODUCTION

X-ray fluorescence (XRF) spectrometers are commonly used for non-invasive elemental analysis of artworks. Mobile instruments for in situ analysis typically detect elements with atomic number \( Z \) > 14 if operating in ambient conditions. However, they collect only one spectrum at a time at single isolated spots that are often chosen visually and may not be fully representative of the whole object. Recent developments in instrumentation allowed macro XRF scanning (MA-XRF) to collect 3D spectral image cubes (two spatial axes and one spectral axis) through raster scans across 2D areas (Figure 1), which enables visualisation of single-element distributions across the surface of an object.\(^1\) Efficient data processing methods have been developed to optimise the selection of single-element maps.\(^2-4\) As the technique gives both spatial and spectral information, it is increasingly utilised for the analysis of various types of artworks, from easel paintings\(^5-7\) to murals\(^8\) and stained-glass windows.\(^9\) MA-XRF datasets are typically analysed through examination of single-element XRF maps.\(^1,5\)

Many studies have presented applications in examinations of the painting sequence,\(^10\) investigations of possible modifications,\(^11\) and even in the revealing of entire underlying paintings.\(^4,5,9,12-14\) These applications are made possible through the imaging capability of MA-XRF; maps or images in the spatial domain are much more intuitive than spectra and can be easily appreciated by non-experts in XRF, such as curators and conservators.

For material identification, superposition of selected single-element maps\(^15-18\) as well as the production of correlation/co-occurrence maps\(^9,13\) are routinely used. However, these methods are mostly effective when dealing with a small number of elements. For complex material compositions that involve a large number of elements, such methods may become cumbersome.

In this study, we present a method that utilises the full information content in a MA-XRF image cube, that is, both spatial and spectral information, in the data analysis. MA-XRF image cubes contain a huge number of spectra (typically ca. 1 million spectra in a 1,000 x 1,000 pixels image cube) which makes full scale data...
FIGURE 1  ‘Peru type’ #478, ca. 1860, Francisco Pancho Fierro, Getty Research Institute, Los Angeles (P840001): (a) colour image; (b) MA-XRF spectral imaging data cube; (c) XRF spectrum extracted at a single spatial pixel. MA-XRF, macro X-ray fluorescence [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 2  ‘Peru type’ #478, ca. 1860, Francisco Pancho Fierro, Getty Research Institute, Los Angeles (P840001): (a) colour image; (b) single-element maps extracted from the MA-XRF spectral image cube (only the main elements). MA-XRF, macro X-ray fluorescence [Colour figure can be viewed at wileyonlinelibrary.com]
interpretation challenging. This article will outline an automated method to efficiently cluster the spectra into distinct groups and objectively assess material variations and map the distribution across an artwork.

The approach is illustrated through the analysis of a MA-XRF spectral imaging data cube of a watercolour painting by the Afro-Peruvian painter Pancho Fierro in the collection of the Getty Research Institute (GRI).

2 | MATERIAL AND METHODS

The MA-XRF spectral imaging data were obtained using a Bruker M6 Jetstream XRF scanner. The M6 uses a 30 W Rh X-ray tube with polycapillary focussing optics and a 30 mm² silicon drift detector (SDD); the voltage and current settings are 50 kV at 600 mA. The spot size was 230 μm, the sampling step was 200 μm at an integration time of 25 ms with a total of 289 × 715 pixels collected. The spectral calibration was performed by identifying unblended lines in the sum spectrum and fitting a straight line to the energy of each peak versus the corresponding channel number. The $rms$ of the residual of the fit was 0.014 keV and the spectrum was sampled at 0.01 keV per channel. The spectral resolution given by the FWHM of a single spectral line typically increases from ~0.1 keV at ~2 keV to ~0.2 keV at ~12 keV. The spectral range spans 0.35–40 keV sampled by 4,096 spectral channels. Pre-treatment of MA-XRF dataset involves spatial median filtering using a $3 \times 3$ pixels kernel and a reduction of the number of spectral channels by including only a sum of five channels around each spectral line detected. This is necessary not only for reducing the amount of data to improve processing efficiency, but also for improving the signal to noise ratio and removing channels that are noise dominated or with unwanted signals so as not to alter the clustering results.

The new approach described here utilises each spectrum within the MA-XRF spectral imaging dataset after pre-treatment. A machine learning algorithm is used to automatically group the pixels that share similar spectra, thus narrowing down the number of spectra that needs to be analysed for material identification to a handful of mean cluster spectra. Analysis of the mean spectra per cluster gives an indication of the materials in each cluster. Visualisation of the spatial distribution of the clusters

![Figure 3](image-url)
gives an overview of the material variations across an object. In heritage science, clustering has been applied to the analysis of spectral imaging datasets in different modalities, from the visible/near-infrared (VIS/NIR) to short-wave Infrared (SWIR) reflectance spectral imaging.\[^{[21,22]}\] The commonly used clustering methods such as the ‘Spectral Hourglass Wizard’ in ENVI require operator inputs in many stages leading to cumbersome and potentially subjective results.

Here we present a new clustering method that is based on the self-organising map (SOM) algorithm, a machine learning method that naturally allows the visualisation of the clusters in 2D space and enables unsupervised clustering, since it is able to learn directly from the input dataset.\[^{[23–25]}\] In unsupervised training, there is no need to have a labelled reference database. One of the most laborious and time-consuming part of machine learning is often the process of manual labelling of data to generate a reference dataset for training. SOM, in contrast, provides a means for fully automated clustering. The ‘kohonen’ function from the built-in R stats package was used.\[^{[24]}\] SOM belongs to the class of shallow artificial neural networks, which consists of an input layer, here represented by the pixel-level XRF spectra within a spectral image cube, mapped onto an output layer with nodes represented by the clusters. The input and output layers are connected through weight vectors (mean cluster spectra). The weight vectors are updated during the learning process where each output node competes for the input data. The only input parameter is the number of clusters, which should be set to a number larger than the actual possible number of clusters. The mean spectra of the output clusters along with the associated standard deviations (SD) per channel are then compared to make sure that the clusters are unique and does not require merging. Uniqueness is defined to be a difference in at least one channel that exceeds 3 SD.

Micro Raman spectroscopy and optical microscopy were also carried out in situ on the Fierro paintings. Raman spectroscopy was performed using a Renishaw inVia Raman microscope equipped with a Leica DM microscope. All spectra were collected using a long focal length ×50 objective (8 mm working distance, N.A. 0.50, sampling size of ca. 2 × 20 μm²) and a 785 nm diode excitation laser. A Keyence VHX-6000 digital microscope was used to capture macro images of the painting; depending on the details, ×20 to ×100 objectives were used to image the painting. Reflectance spectroscopy was performed

### Table 1

Elemental composition of the XRF clusters

| Cluster | Elemental composition | Material composition |
|---------|-----------------------|----------------------|
| 1       | Ca, Zn, Fe, Pb, Cu, S, Hg, Mn, Ba | Sizing/filling materials + Pb-containing ink + vermilion + barium white |
| 2       | Ca, Zn, Fe, Pb, Hg, Cu, K, Ba, Mn, S | Sizing/filling materials + Pb-containing ink + vermilion + barium white |
| 3       | Pb, Ca, Hg, Zn, Fe, Cu, Mn | Pb-containing ink + vermilion + sizing/filling materials |
| 4       | Pb, Ca, Hg, Zn, Fe, Cu, Mn | Pb-containing ink + vermilion + sizing/filling materials |
| 5       | Pb, Ca, Hg, Zn, Fe, Cu, Mn | Pb-containing ink + vermilion + sizing/filling materials |
| 6       | Fe, Ca, Pb, Hg, Zn, Mn, K, Ba, S, Cu | Yellow ochre + Pb-containing ink + vermilion + barium white + sizing/filling materials |
| 7       | Hg, Ca, Pb, Zn, Fe, Mn, Cu, K | Vermilion + Pb-containing ink + sizing/filling materials |
| 8       | Hg, Ca, Pb, Zn, Fe, Ba, S, Mn, K, Cu | Vermilion + Pb-containing ink + barium white + sizing/filling materials |
| 9       | Hg, Ca, Pb, Fe, Ba, Zn, Cu, Mn, S, K | Vermilion + Pb-containing ink + barium white + Prussian blue + sizing/filling materials |
| 10      | Ca, Pb, Ba, Fe, Hg, Zn, K, Mn, S, Cu | Pb-containing ink + vermilion + barium white + Prussian blue + sizing/filling materials |
| 11      | Pb, Ca, Fe, Cr, Hg, Ba, Zn, Cu, K | Chrome yellow + Prussian blue + barium white + Pb-containing ink + vermilion + sizing/filling materials |
| 12      | Pb, Ca, Fe, Cr, Ba, Hg, Zn, K, Cu | Chrome yellow + Prussian blue + barium white + Pb-containing ink + vermilion + sizing/filling materials |
| 13      | Pb, Ca, Hg, Fe, Cr, Ba, Zn, Cu, K | Pb-containing ink + vermilion + chrome yellow + Prussian blue + barium white + sizing/filling materials |

Note: Bold font denotes main elements (net intensity of strongest peak ≥1 counts/s), normal font denotes minor elements (net intensity of strongest peak between 0.1 and 1 counts/s) and italics font denotes trace elements (net intensity of strongest peak ≤0.1 counts/s). The elements are ordered in decreasing intensity.

Abbreviation: XRF, X-ray fluorescence.
using an ASD FieldSpec spectrometer, composed of three detectors that cover the spectral range from 350 to 2,500 nm. The spectral resolution is 3 nm in the UV/VIS regime and 10 nm in the NIR.

3 | RESULTS AND DISCUSSION

Figure 2 shows an overview of the single-element maps extracted from the MA-XRF spectral cube of the Fierro painting following the prescribed procedure: energy-channel calibration and fundamental parameters based fitting using PyMCA\textsuperscript{[26]} and examination of elemental correlations using Datamuncher.\textsuperscript{[27]} The usual approach to MA-XRF data interpretation are based on the single-element maps and spatial correlation maps between pairs of elements or co-location maps.

In our method, single-element maps are not needed as the clustering analysis is performed directly on the pre-treated MA-XRF cube. Procedures such as spectral deconvolution are not needed, until the interpretation stage when individual mean cluster spectra are examined. SOM clustering on the MA-XRF spectral cube narrows down the hundreds of thousands (289 × 715) pixel-level spectra of the cube to 13 distinct clusters/groups (Figure 3), showing variations in the material composition. Table 1 (second column) summarises the elemental content of each of the 13 clusters. Detailed analysis of the mean spectra of these clusters enabled the identification of the material composition (Table 1, third column). For watercolour paintings on paper, where the surfaces are relatively flat, assumptions are made that the matrix effects are minimal, even in mixtures, due to the thin amounts of materials applied. The paint mixture is usually well mixed with the paper fibres rather than forming a well-defined layer above the paper. The clustering method takes into account variations in the peak intensity, allowing the distinction of differences in the column density of the paint and the relative concentration of the inorganic materials. This is reflected in Table 1 where there are clusters that contain the same elemental mix but different relative strength.

![Figure 4](https://wileyonlinelibrary.com)
Visualisation of the SOM results through a cluster map (Figure 3b) offers an overview of the material diversity at a glance. With the proposed approach, the analysis of the MA-XRF spectral cube is no longer based on the single-element maps, but instead, on the single-cluster maps (Figure 3c), with each cluster corresponding to a distinct XRF spectrum and therefore a distinct material composition.

In the following, we will show how clustering can be used to deduce the paint sequence and aid with material identification.

The black scarf in the painting is described by three clusters (clusters 3, 4 and 5) as shown in Figure 4. The mean spectra of the three clusters all show the same elemental combination with the only difference being the intensity of the spectral lines, which is most likely related...

**Figure 5**  ‘Peru type’ #478, ca. 1860, Francisco Pancho Fierro, Getty Research Institute, Los Angeles (P840001): (a) colour image of the apron with a blue edge; (b) combined cluster map of clusters 8, 9 and 10; the corresponding single-cluster maps along with their mean spectra of clusters: (c,d) cluster 8, (e,f) cluster 9 and (g,h) cluster 10 (the mean spectra are colour coded in the same colour as the corresponding cluster maps) [Colour figure can be viewed at wileyonlinelibrary.com]
to differences in the column density of the black paint. All three spectra have prominent Pb lines and minor contributions from Hg and trace elements common to those found on the ‘blank’ substrate. Raman spectroscopy detected carbon black (Figure 4d) and a weak signal of vermilion which suggests that the black paint is carbon black ink with traces of vermilion. Pb-containing pigments, such as lead white and red lead, were not detected by Raman. Reflectance spectroscopy on the same area eliminated the possibility of galena (PbS) or plattnerite (PbO₂) (Figure 4e). Pb is, therefore, more likely to be associated with the black ink itself. This is further backed up by the observation that the intensity of the Pb lines increases with the increased darkness of the black paint. Pb is known to be found in various historic carbon black inks.⁴²

Figure 5 demonstrates a case where the combined analysis of the spatial distribution of the clusters, along with their mean spectra, reveals information about the painting sequence and materials. The contribution of the substrate (cluster 1 in Table 1) is present in all the three clusters. The blue edges of the apron are represented by cluster 10 (Figure 5g), where the mean spectrum (Figure 5h) has strong peaks of Fe and Ba as well as elements of the Pb-containing carbon black ink and vermilion mixture. The Fe peak is more prominent here than that of the substrate (cluster 1 in Table 1), which makes it most likely to be associated with the blue pigment Prussian blue (M₁Fe²⁺Fe⁺[CN]₆ₙH₂O, where M can be K, NH₄ or Na depending on the method of manufacture). The strong presence of Ba can be attributed to barium white (BaSO₄) which gives Prussian blue a lighter blue tone. Pure Prussian blue is nearly black. The mean spectrum of cluster 8 (Figure 5d), that corresponds to the dark outline of the blue edge of the apron (Figure 5a,c), shows the combination of a Hg-containing pigment (i.e. the red pigment vermilion) with the Pb-containing carbon black ink and traces of barium white. Micro Raman spectroscopy confirmed the identification of Prussian blue, carbon black and vermilion (Figure 6). Cluster 9 (Figure 5b,e) corresponds to areas spatially between clusters 8 (dark outline of the blue edge of the apron) and
cluster 10 (blue edge of the apron). Its mean cluster spectrum (Figure 5f) shows that it is a combination of clusters 8 and 10. The combined interpretation of the three clusters that describe the apron’s edges suggests that there is an overlap between the blue edge and the dark outline. Optical microscopy on this area confirmed this interpretation, showing that the blue was painted first and the dark outline was painted afterwards (Figure 6a).

Figure 7 illustrates a case where clustering facilitates material identification. Clusters 11 and 12 describe the green skirt. They share the same elemental content but with a slight difference in the intensity of their mean spectra which corresponds to the lighter and darker stripes in the skirt (Figure 7b). The very dark shadows of the skirt are in cluster 13 (Figure 3c). The elemental complexity could lead to various interpretations of its material composition. The detection of both Ba and Cr could mean the presence of barium chromate (BaCrO$_4$), also known as lemon yellow. However, the mean spectra of clusters 11 and 12 (Figure 7b) appear to be a spectral combination of clusters 8 (dark outline of the apron) and cluster 10 (blue edge of the apron) with the addition of Cr and much more prominent Pb lines. This suggests that the pigment composition is more likely to be chrome yellow (PbCrO$_4$), Prussian blue, barium white and a small amount of vermilion and Pb-containing carbon black ink. Raman spectroscopy confirmed the presence of chrome yellow, Prussian blue and carbon black (Figure 7c). Given that this area has a complex elemental composition (nine elements), past methods would require manual comparison of many co-location maps. This could result in potentially incomplete results, unless all combinations of bi-scatter plots and co-location maps are examined together in detail which is cumbersome and inefficient.

In single-element maps, the large dynamic range required for display of both the high and low intensity regions of an element often results in the omission of low intensity areas. In general, it is often difficult to see trace elements even with careful adjustment of contrast. The examination of the arm area illustrates this issue (Figure 8). Figure 8b show that the mean spectrum of cluster 6 (arm area) is dominated by an Fe-containing pigment (most likely yellow ochre) with a thin wash of
the vermilion and Pb-containing black ink mixture along with a trace amount of barium white. Optical microscopy image was consistent with this suggestion (Figure 8c). In the single-element maps (Figure 2), the arm area only shows up clearly in the Fe and K maps, but not in the Pb, Hg and Ba maps. The material identification in the arm region would consequently be incomplete. Clustering by spectra followed by a close examination of the mean XRF spectra avoids this problem and allows a more complete material identification. An added advantage is that clustering allows all pixels of the same material group to be averaged to produce a mean cluster spectrum which is of much higher signal to noise ratio than a simple average over a small region of interest. This is particularly powerful in needles in a haystack situation where pixels of similar material content are sparsely dispersed over a region dominated by a very different material.

4 | CONCLUSIONS

In this article, we present a new approach to the interpretation of MA-XRF spectral imaging data that uses all the spectral and spatial information thus taking full advantage of the data gathered. The new approach is based on clustering of the pixel-level spectra of the entire MA-XRF spectral image cube using an unsupervised neural network-based machine learning method, SOM, which enables automatic, accurate and time-efficient processing of a spectral imaging dataset. Visualisation of the clustering results gives an overview of the material variations across the painting. By using the full spectra of each cluster and their spatial distribution and spectral relationship, we have demonstrated how clustering the pixel-level spectra in a MA-XRF spectral imaging data cube allows not only the grouping of similar materials (inorganic components only) into distinct clusters, but also efficient pigment identification through the analysis of a cluster mean spectrum and its relationship with the various clusters across a painting. Future work will explore clustering of complementary spectral imaging data that includes both elemental and molecular information to provide more definitive material identification.

ACKNOWLEDGEMENTS

The authors would like to thank Douglas MacLennan, Nathan Daly and Catherine Patterson of Getty...
Conservation Institute (GCI) for their help in the MA-XRF data collection and insightful discussions; Karen Trentelman (GCI) for reading the manuscript and giving helpful, comprehensive feedback; Idurre Alonso and Lisa Forman at the Getty Research Institute for their help in accessing the Fierro works and for discussions on the Fierro works. Part of this project was funded by the UK Arts and Humanities Research Council (AHRC grant reference AH/T013184/1).

ORCID
Haida Liang https://orcid.org/0000-0001-9496-406X

REFERENCES
[1] M. Alfeld, K. Janssens, J. Dik, W. De Nolf, G. Van Der Snickt, J. Anal. At. Spectrom. 2011, 26, 899.
[2] M. Alfeld, K. Janssens, J. Anal. At. Spectrom. 2015, 30, 3.
[3] F. P. Romano, C. Caliri, P. Nicotra, S. Di Martino, L. Pappalardo, F. Rizzo, H. C. Santos, J. Anal. At. Spectrom. 2017, 32, 773.
[4] M. Alfeld, W. De Nolf, S. Cagno, K. Appel, D. Peter Siddons, A. Kuczewski, K. Janssens, J. Dik, K. Trentelman, M. Waltone, A. Sartorius, J. Anal. At. Spectrom. 2013, 28, 40.
[5] G. Van Der Snickt, A. Martins, J. Delaney, K. Janseens, J. Zeibeil, M. Duffy, C. Mc Glinchey, B. Van Driel, J. Dik, Appl. Spectrosc. 2016, 70, 57.
[6] C. Ruberto, A. Mazzinghi, M. Massi, L. Castelli, C. Czelusniak, L. Palla, N. Gelli, M. Betuzzi, A. Impallaria, R. Branccio, E. Peccenini, M. Raffaelli, Microchem. J. 2016, 126, 63.
[7] K. A. Dooley, D. M. Conover, L. D. Glinsman, J. K. Delaney, Angew. Chem. Int. Ed. 2014, 53, 50.
[8] M. Alfeld, S. Pedetti, P. Martinez, P. Walter, Compt. Rend. Phys. 2018, 19, 7.
[9] G. Van der Snickt, S. Legrand, J. Caen, F. Vanmeert, M. Alfeld, K. Janssens, Microchem. J. 2016, 124, 615.
[10] A. Martins, C. Albertson, C. McGlinchey, J. Dik, Herit. Sci. 2016, 4, 1.
[11] M. Iwanicka, P. Targowski, Ł. Ćwikliński, Proceedings of 2nd International Conference on Art & Archaeology. Jerusalem: ISAS, 2016.
[12] A. Anitha, A. Brasoveanu, M. F. Duarte, S. M. Hughes, I. Daubechies, J. Dik, K. Janssens, M. Alfeld, European Signal Processing Conference. Barcelona: IEEE, 2011.
[13] P. Walter, P. Sarrazin, M. Gailhanou, D. Hérouard, A. Verney, D. Blake, X-Ray Spectrom. 2019, 48, 4.
[14] A. Galli, M. Caccia, R. Alberti, L. Bonizzoni, N. Aresi, T. Frizzi, L. Bombelli, M. Gironia, M. Martini, X-Ray Spectrom. 2017, 46, 5.
[15] P. Ricciardi, S. Legrand, G. Bertolotti, K. Janssens, Microchem. J. 2016, 124, 5.
[16] P. Targowski, M. Pronobis-Gajdzis, A. Surmak, M. Iwanicka, E. A. Kaszewska, M. Sylwestrzak, Stud. Conserv. 2015, 60, S167.
[17] L. de Viguerie, S. Rochut, M. Alfeld, P. Walter, S. Astier, V. Gontero, F. Boulic’h, Herit. Sci. 2018, 6, 1.
[18] S. Mosca, T. Frizzi, M. Pontone, R. Alberti, L. Bombelli, V. Capogroso, A. Nevin, G. Valentini, D. Comella, Microchem. J. 2016, 124, 775.
[19] B. Anderson, El Palacio, 2011, p. 44. https://www.yumpu.com/en/document/read/51020413/from-lima-to-canton-and-back-nineteenth-century-watercolors-for-

How to cite this article: Kogou S, Lee L, Shahtahmassebi G, Liang H. A new approach to the interpretation of XRF spectral imaging data using neural networks. X-Ray Spectrom. 2021;50: 310–319. https://doi.org/10.1002/xrs.3188