

Enhanced Visualization and Interpretation of XMCD-PEEM Data Using SOM-RPM Machine Learning

See Yoong Wong, Sarah L. Harmer, Wil Gardner, Alex K. Schenk, Davide Ballabio, and Paul J. Pigram*

Photoemission electron microscopy (PEEM) is a powerful technique for surface characterization that provides detailed information on the chemical and structural properties of materials at the nanoscale. In this study, the potential is explored using a machine learning algorithm called self-organizing map with a relational perspective map (SOM-RPM) for visualizing and analyzing complex PEEM-generated datasets. The application of SOM-RPM is demonstrated using synchrotron-based X-ray magnetic circular dichroism (XMCD)-PEEM data acquired from a pyrrhotite sample. Traditional visualization approaches for XMCD-PEEM data may not fully capture the complexity of the sample, especially in the case of heterogeneous materials. By applying SOM-RPM to the XMCD-PEEM data, a colored topographic map is created that represents the spectral similarities and dissimilarities among the pixels. This approach allows for a more intuitive and easily interpretable representation of the data without the need of data binning or spectral smoothing. The results of the SOM-RPM analysis are compared to the conventional visualization approach, highlighting the advantages of SOM-RPM in revealing features that are not readily observable in the conventional method. This study suggests that the SOM-RPM approach can be used complementarily for other PEEM-based measurements, such as core level and valence band X-ray photoelectron spectroscopy.

nanoscale by measuring the electrons emitted by a sample under specific illumination conditions.^[1] One way in which PEEM can be used is to analyze the secondary electrons emitted from each pixel – a measure of the photon absorption rate – as the exciting photon energy is scanned in the X-ray absorption near edge structure (XANES) region which, among other details, reveals information on the distribution of chemical states and bonding of each element. This technique can be extended by acquiring XANES spectra with left- (LHS) and right-circularly (RHS) polarized beams as the excitation source and exploring the difference in electron yield with the changed polarization, known as X-ray Magnetic Circular Dichroism PEEM (XMCD-PEEM). As XMCD absorption is strongly dependent on the magnetic properties of a sample, XMCD-PEEM can be used to explore the complex relationship between magnetic domains and chemical composition in a material in a spatially resolved fashion. As a result, this technique has been

1. Introduction


Photoemission Electron Microscopy (PEEM) is an advanced surface characterization technique, operable in multiple modes. It can provide detailed, spatially resolved information on the chemical and structural properties of a materials surface at the

used to study a wide range of material systems, including thin films,^[2] multilayers,^[3] and nanoparticles.^[4]

While XMCD-PEEM can provide valuable insights into the elemental distribution, chemical states, magnetization, and limited information on molecular geometry, effectively analyzing all of this information and forming correlations between features in

S. Y. Wong, W. Gardner, P. J. Pigram
Centre for Materials and Surface Science
La Trobe University
Melbourne, Victoria 3086, Australia
E-mail: p.pigram@latrobe.edu.au

S. Y. Wong, W. Gardner, A. K. Schenk, P. J. Pigram
Department of Mathematical and Physical Sciences
La Trobe University
Melbourne, Victoria 3086, Australia
S. L. Harmer
Flinders Microscopy and Microanalysis
College of Science and Engineering
Flinders University
Adelaide, South Australia 5001, Australia
D. Ballabio
Milano Chemometrics and QSAR Research Group
Department of Earth and Environmental Sciences
University of Milano-Bicocca
Milano 20126, Italy

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/admi.202300581>

© 2023 The Authors. Advanced Materials Interfaces published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/admi.202300581

images can be challenging. Conventionally, researchers use the relative electron flux captured at specific energy levels to represent the distribution of elemental and chemical states,^[5,6] and the contrast in emitted electrons between left- and right-circularly polarized beams at that energy level is displayed separately as an XMCD map.^[7,8] While this is suitable for simple materials where there are few distinct phases, these methods may be less informative when examining complex samples that have a more heterogeneous nature. To address these challenges, new approaches to data reduction and visualization need to be developed, in order to better exploit the full potential of XMCD-PEEM for characterization of complex material phases.

The use of machine learning (ML) as a tool for aiding spectroscopy data analysis is an emerging field, and there are many algorithms and approaches that can be used, as well as many workflows in which it can play a role in. For example, a ML model known as Gaussian process modeling has been used to reduce the size of XMCD dataset for accurate magnetic moment evaluation.^[9] ML algorithms, such as the unsupervised clustering algorithms (K-means and fuzzy-c-means),^[10] can also reveal spatial distribution patterns and identify molecular heterogeneity within samples. While these algorithms are proven useful for data classification and prediction, a natural concern of spectroscopists in every case is the selection of an appropriate tool which is used in a scientifically rigorous fashion, with an understanding of the limitations of the selected tool. Hence, it is important to explore the application of various ML algorithm and their respective limitations.

Self-organizing maps (SOM), which is another type of unsupervised ML model, were successfully applied in the field of mass spectrometry imaging (MSI).^[11,12] The term unsupervised refers to a category of algorithms and techniques that involve training a model on a dataset without explicit supervision via labels. By processing high-dimensional hyperspectral data, the SOM creates a topological map that reflects the relationships between different molecular species. This empowers researchers to visually explore the complex data space of MSI, facilitating the identification of regions with similar molecular composition and spatial correlations. In our previous works, we have further enhanced the traditional SOM by integrating it with a secondary algorithm known as the relational perspective map (RPM), resulting in a method called SOM-RPM.^[13–15] Our previous studies have demonstrated the effectiveness of SOM-RPM in visualizing MSI datasets, providing an intuitive and interpretable low-dimensional representation of the complex data. Given that the MSI dataset and the XMCD-PEEM dataset share the same hyperspectral format, the utilization of SOM-RPM as a visualization technique for XMCD-PEEM is highly promising.

In this study, we use the iron sulphide mineral pyrrhotite and synchrotron-based XMCD-PEEM acquired at the Fe L-edge to demonstrate the potential of SOM-RPM for exploring complex PEEM-generated datasets. Pyrrhotite has a chemical composition of $\text{Fe}_{(1-x)}\text{S}$, where x can range from 0 to 0.125, reflecting its iron deficient nature,^[16] and is composed of both monoclinic (Fe_7S_8) and hexagonal domains (FeS), which exhibit magnetic and antiferromagnetic properties, respectively. Furthermore, pyrrhotite has a high degree of reactivity with oxygen, resulting in the formation of iron oxides and oxyhydroxides.^[6,17] While pyrrhotite is commonly considered as mining waste, it is widely studied

in the paleomagnetic and geomagnetic fields since it is a common rock-forming mineral with magnetic properties.^[18] However, pyrrhotite is often confused with other iron sulphide minerals such as greigite.^[19] More advanced techniques are also required to identify the origin of pyrrhotite. Its high metallic-type conductivity^[18] compared to other sulfides also makes it an attractive candidate as the cathodic material for lithium-sulphur batteries.^[20] The fact that the mineral includes coexisting phases with differing chemical composition and magnetic properties – which both influence the shape of Fe L-edge spectra acquired with different rotational polarizations – make this mineral a useful demonstration of what ML and, more specifically, SOM-RPM, can offer for analysis of PEEM-based measurements.

To compare the visualization ability of SOM-RPM in interpreting XMCD-PEEM data, we first display the typical PEEM and XMCD images separately in the conventional way. We then analyze and compare different regions on the similarity map produced by SOM-RPM to study the properties of the pyrrhotite sample. We show that SOM-RPM provides an intuitive yet comprehensive description of the underlying relationships and key features in the data. We therefore propose that this method can effectively enhance our understanding of surface characterization techniques and the properties of pyrrhotite, contributing to the current understanding of materials science.

2. Results and Discussion

2.1. Synchrotron XMCD-PEEM of Pyrrhotite – Traditional Visualization Approach

The synchrotron XMCD-PEEM images of a freshly polished pyrrhotite sample are presented in **Figure 1**. After the baseline removal explained above, all maps in Figure 1 are normalized by their respective maximum values for better visualization.

Figure 1a,b display the normalized electron count acquired at 707.8 eV, which, as demonstrated by Mikhlin & Tomashevich,^[21] is associated with electronic transitions in iron sulfides.^[22] Figure 1c,d show the normalized electron count acquired at 709.4 eV for LHS and RHS polarization respectively, revealing the iron oxidation products,^[6] which is likely to be amorphous.^[17] Figure 1c,d indicate that the sample contains mostly unoxidized pyrrhotite (blue), with a few highly oxidized areas indicated by the yellow domains. This observation is consistent with the expected composition of the sample, which contains fewer reactive monoclinic regions and more reactive hexagonal regions.^[17] Figure 1e,f show the XMCD magnetic contrast image at 707.8 eV and 709.4 eV respectively. The asymmetry value is calculated by taking the difference of electron intensity between left and right polarization at specific energy level and dividing by their sum, as $\frac{I_{\text{RHS}} - I_{\text{LHS}}}{I_{\text{RHS}} + I_{\text{LHS}}}$, to represent the magnetic contrast across the sample.^[8,23]

Pyrrhotite in the iron-rich region has a hexagonal structure and a chemical composition of FeS, similar to troilite, which is a non-magnetic mineral.^[24] As the level of iron vacancy increases, the electrical charge of the iron changes, leading to a corresponding increase in its magnetic response. In the iron-deficient region, pyrrhotite has a monoclinic structure and a chemical

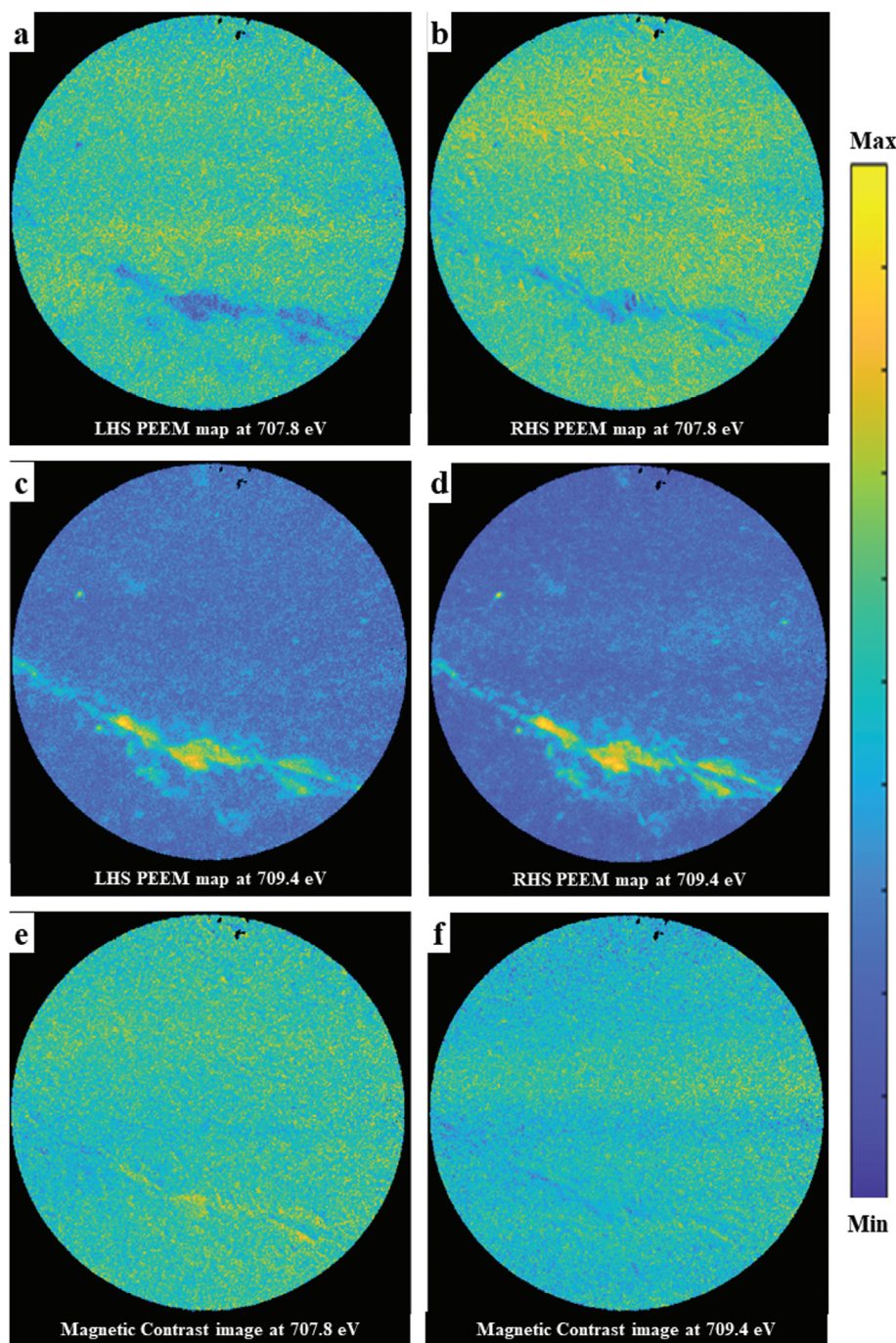


Figure 1. Visualization of XMCD-PEEM data using the conventional visualization approach. a) LHS and b) RHS PEEM map at 707.8 eV, representing the iron sulfide species. c) LHS and d) RHS PEEM map at 709.4 eV, representing the iron oxidation products. XMCD-PEEM asymmetry images at e) 707.8 eV and f) 709.4 eV, displaying the degree of magnetism across the sample for the iron sulfide and iron oxide species respectively. The images show the same sample area, and they are displayed in a normalized Parula color scale.

composition of Fe_7S_8 . Therefore, the vacancy of iron on the pyrrhotite sample could be revealed by comparing the unoxidized region (yellow in Figure 1a,b; blue in Figure 1c,d) with Figure 1e, which is the magnetic contrast at 707.8 eV. This comparison indicated that the oxidized region (yellow in Figure 1c,d) is surrounded by non-magnetic iron sulfide (blue in Figure 1e).

These findings align with the reactivity of pyrrhotite, as monoclinic pyrrhotite, which has an activation energy of $50.21 \text{ kJ mol}^{-1}$, is less reactive than hexagonal pyrrhotite, which has a lower activation energy of $46.23 \text{ kJ mol}^{-1}$.^[17]

While it is apparent that a traditional visualization approach allows us to examine pyrrhotite and already draw some

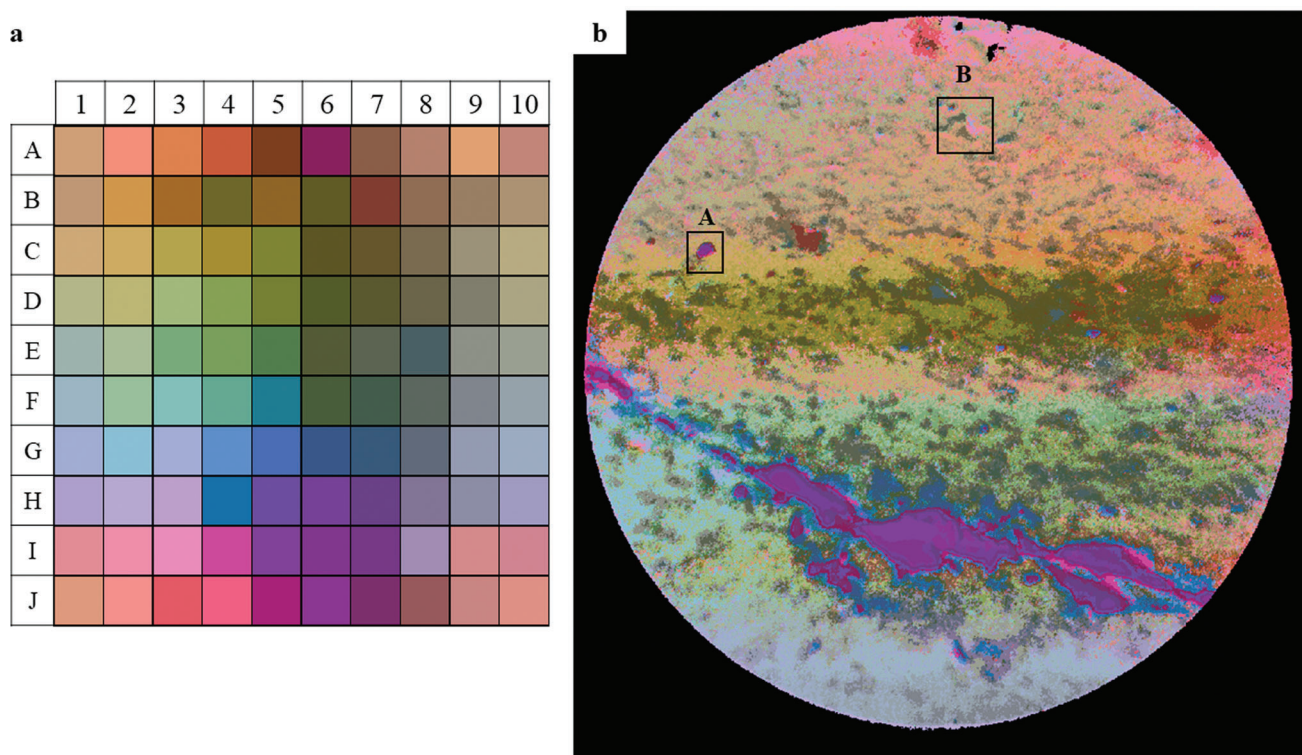


Figure 2. SOM-RPM image of XMCD-PEEM of pyrrhotite, 10×10 neurons and 10 000 epochs. a) Colored SOM topographic map. The neurons are labelled for further discussion. Note that the SOM is unfolded from a toroidal topology. b) Similarity map of pyrrhotite. Region A and B are highlighted for further comparisons.

conclusions, we now turn to exploring the possibilities offered by the SOM-RPM.

2.2. SOM-RPM Analysis of XMCD-PEEM Data

To consider all acquired data and enable a more detailed characterization, we concatenated the Fe L-edge XANES spectra obtained from both polarities and the data was used to train the SOM-RPM model. By doing so, we simultaneously consider both the oxidation state and magnetic characteristics to produce a single map of the data product. **Figure 2a** shows the colored SOM (10×10 neurons, 10 000 epochs) and **Figure 2b** shows the similarity map of the same dataset shown in **Figure 1**.

The similarity maps trained with 6×6 and 14×14 neurons are shown in **Figure S2** (Supporting Information). While all three maps are similar visually, some minor features revealed in **Figure S2b** (Supporting Information) (10×10 neurons) and **Figure S2c** (Supporting Information) (14×14 neurons) are missing in **Figure S2a** (Supporting Information) (6×6 neurons). These observations highlight the impact of network size on the hierarchy of clustering. A larger network size enables better pixel allocation, leading to improved clustering quality and reduced quantization error. However, it remains crucial for analysts to interpret the data clustering. Individual neurons or neuron clusters may represent distinct crystal phases in the sample. Here, we use the model trained using 10×10 neurons to study the convergence of the SOM model for training since **Figure S2c** (Supporting In-

formation) and **Figure S2b** (Supporting Information) are visually similar. Larger SOM sizes may have revealed more subtle differences between similar groups of pixels, however for the purposes of this work 10×10 neurons was sufficient. Again, we note that this decision must be based upon analyst expertise and the requirements of the study, as there is no objective measure to indicate overfitting (unlike in supervised learning).

The quantization error for SOM with 10×10 neurons and various number of epochs are tabulated in **Table S1** (Supporting Information) and plotted in **Figure S3** (Supporting Information). The quantization error dropped from 9.6 to 0.3 just after five iterations, showing the efficiency of SOM in clustering XMCD-PEEM spectra. We proceed with the model that exhibited the lowest quantization error, achieved through a configuration of 10×10 neurons and trained over 10 000 epochs, for further discussions. The location of pixels assigned to each neuron in **Figure 2a**, and their respective mean intensity and standard deviation are presented in **Figures S4–S13** (Supporting Information). Low standard deviation values calculated for all neurons further suggest that sufficient size and iteration are provided for SOM training.

A comparison of the conventional approach (**Figure 1**) to the similarity map generated by the SOM-RPM model (**Figure 2**) reveals that the similarity map in **Figure 2** provides a more comprehensive view of the sample. By comparing conventional PEEM-XMCD map (**Figure 1**) to the similarity map (**Figure 2b**), the color scheme in **Figure 2b** presents iron oxidation products (yellow in **Figure 1c,d**) as dark purple (neuron H5 to J7), FeS regions (non-magnetic, blue in **Figure 1e**) as pink (neuron I3), and Fe_7S_8

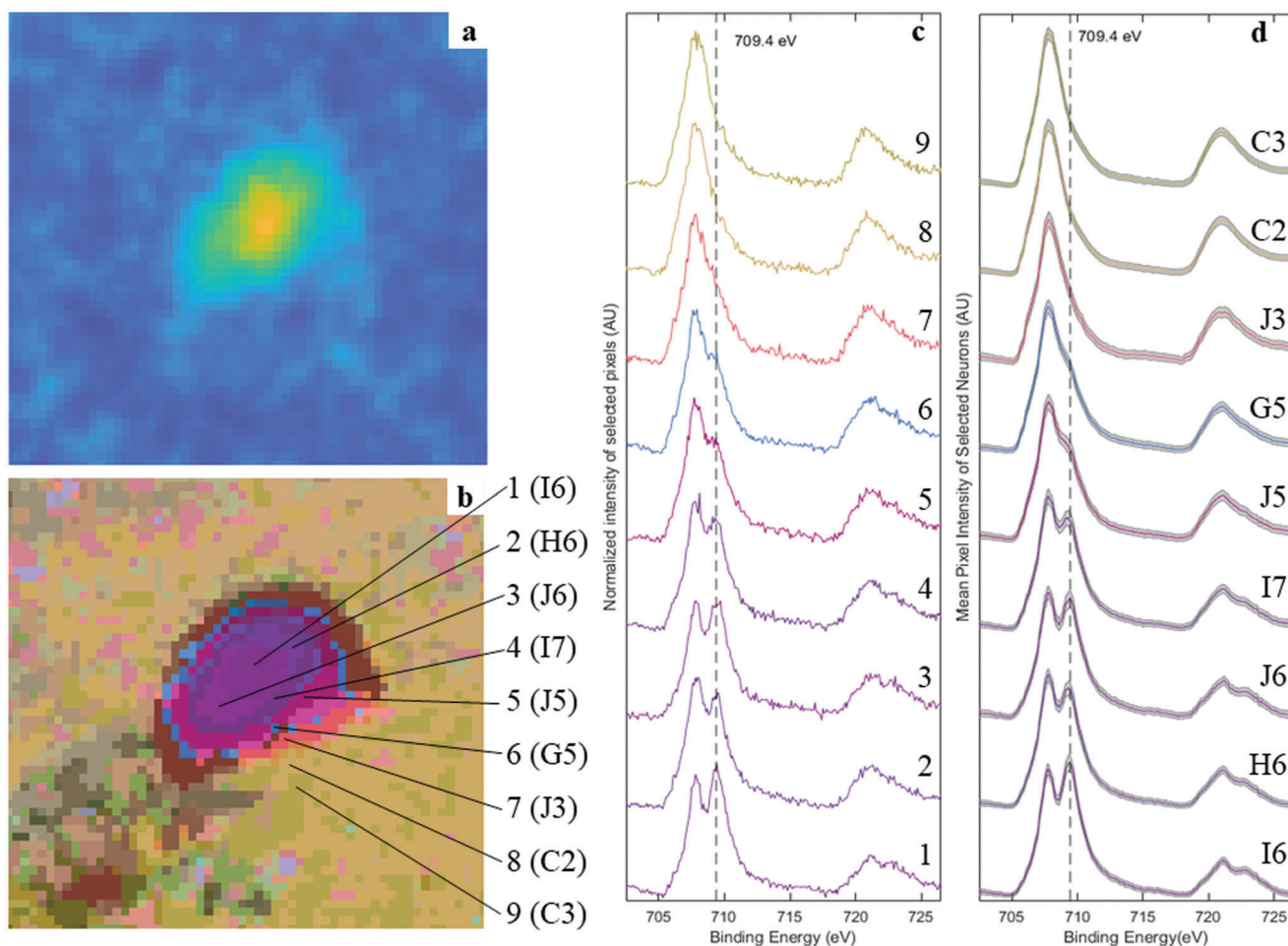


Figure 3. Comparison between SOM-RPM and conventional approach in the analysis of iron oxidation products (Region A highlighted in Figure 2b). a) Magnified PEEM map on Region A in Figure 1c. The image is reproduced here to allow better comparison. b) Magnified SOM-RPM map on Region A in Figure 2b, nine pixels are selected for discussion, numbers in brackets show the assigned neurons of selected pixels. c) Waterfall plot showing the normalized left polarization PEEM spectra of selected pixels. d) Waterfall plot showing average normalized spectra of average pixel intensity for the left polarization of selected neurons. The shaded regions show the standard deviation of the intensity for all pixels assigned to each neuron. A line is plotted at 709.4 eV as guide.

regions (ferrimagnetic, yellow in Figure 1e) as olive (neuron C8). Figure 2b reveals features that were not readily observable in Figure 1. These include the phase boundary of the oxidation products, magnetic and non-magnetic regions, and distinct domains (primarily beige, green, and light blue) comprising the unoxidized region. In the following sections, we will explore these features in greater detail. To facilitate comparison of conventional method and SOM-RPM, a particle representing iron oxidation products (yellow in Figure 1c,d; dark purple in Figure 2b) and a region containing magnetic and non-magnetic FeS (yellow and blue in Figure 1e; olive and pink in Figure 2b) are highlighted as region A and B respectively.

2.3. Chemical State Analysis (Fe_{1-x}S Vs Oxidation Products)

To begin, we focus on Region A highlighted in Figure 2b, which represents a grain of iron oxidation products. Magnified images of Region A are reproduced and shown in Figure 3a,b, depict-

ing the PEEM map (reproduced from Figure 1c) and similarity map (reproduced from Figure 2b), respectively. For analysis, we selected nine different pixels located at varying distances from the center of the grain. The numbers in brackets show the location of neuron in the color SOM (Figure 2a). We then plot the PEEM spectra of the selected pixels in Figure 3c and compare the mean pixel intensity of the selected neurons using their respective assigned colors in Figure 3d.

The color-differentiated regions in the SOM-RPM map highlight subtle differences in the spectral characteristics of the sample that are not easily visible in the original map. In both figures, the particle appears to have a circular shape, with a central region that is brighter than the surrounding area, corresponding to the iron oxidation products. However, there are some differences in the shape of the particle between the two maps. In the original PEEM map, the particle appears to have a slightly irregular boundary, with some protrusions and indentations along the edge. In contrast, in the SOM-RPM processed map, the particle appears to have a more uniform circular shape, with a smooth

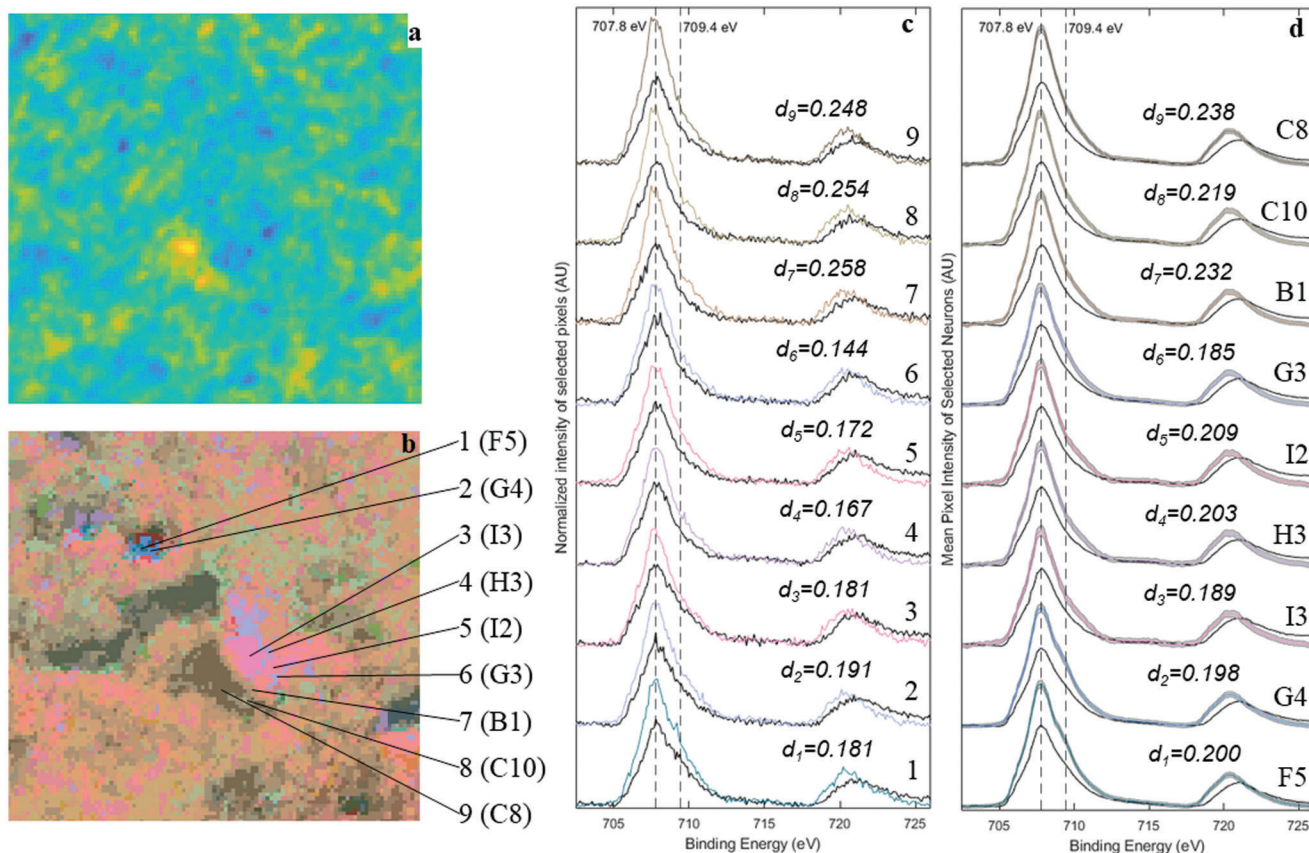


Figure 4. Comparison between conventional and SOM-RPM approach in the analysis of the magnetic property of pyrrhotite on Region B highlighted in Figure 2b. a) Magnified XMCD map on Region B in Figure 1e. The image is reproduced here to allow better comparison. b) Magnified similarity map on Region B in Figure 2b. Nine ROIs are selected for discussion, numbers in brackets show the assigned neurons of selected pixels. c) Waterfall plot showing the normalized PEEM spectra of selected pixels. The asymmetry between LHS (black) and RHS (color) polarized Fe^{2+} peaks ($d_{1,g}$) of each pixel is also displayed. d) Waterfall plot showing the mean pixel spectra of selected neurons. The shaded regions show the standard deviation of the intensity for all pixels assigned to each neuron. The asymmetry between LHS (black) and RHS (color) polarized Fe^{2+} peaks ($d_{1,g}$) of each neuron are displayed. Two lines at 707.8 eV and 709.4 eV are plotted as guides.

boundary. This difference suggests that the unbiased consideration of the full XMCD-PEEM spectra which the SOM-RPM approach utilizes has increased the ability to map oxidation at and across this boundary. The smoother appearance also reflects the decreased sensitivity of SOM-RPM to random noise and fluctuations in the dataset.

The waterfall plot depicting the original PEEM spectra of selected pixels is shown in Figure 3c. Figure 3d, on the other hand, shows the respective mean \pm standard deviation of the intensities of all pixels assigned to the same neurons as the selected pixels, with plot coloring matching the SOM-RPM coloring of the neurons. It can be observed in Figure 3d that as we move toward the center of the oxidized particle, a shoulder begins to emerge and progressively intensifies. This observation suggests that the oxidation process occurring in the region is not a sharp transition, but rather occurs in a gradient manner from the outer edge toward the center of the particle. The emergence of the shoulder feature in the spectra of the selected regions may be attributed to the presence of different oxidation states or phases within the particle; while the typical PEEM processing does not clearly show any structured variation around this re-

gion, the SOM-RPM map allocates these points to different neurons, allowing for a clear image of the emergence of these different phases. This demonstrates the potential for this approach to provide valuable insights into the understanding of the oxidation behavior and mechanisms of iron-based materials, and more generally enable the assessment of phase formation in materials.

2.4. Magnetic State Analysis (FeS Vs Fe_7S_8)

To study the application of SOM-RPM in identifying regions with different magnetic property, a region containing high and low magnetic contrast is selected. Figure 4a,b provide a magnified view of Region B on the XMCD (reproduced from Figure 1b) and similarity maps (reproduced from Figure 2b), respectively. To analyze the region further, we randomly selected nine different pixels and identified their assigned neurons in brackets. In Figure 4a, ROIs 1 to 6 exhibit low XMCD contrast. Given that the degree of perceived color alteration represents the Euclidean distance between neurons, we further explored the notable

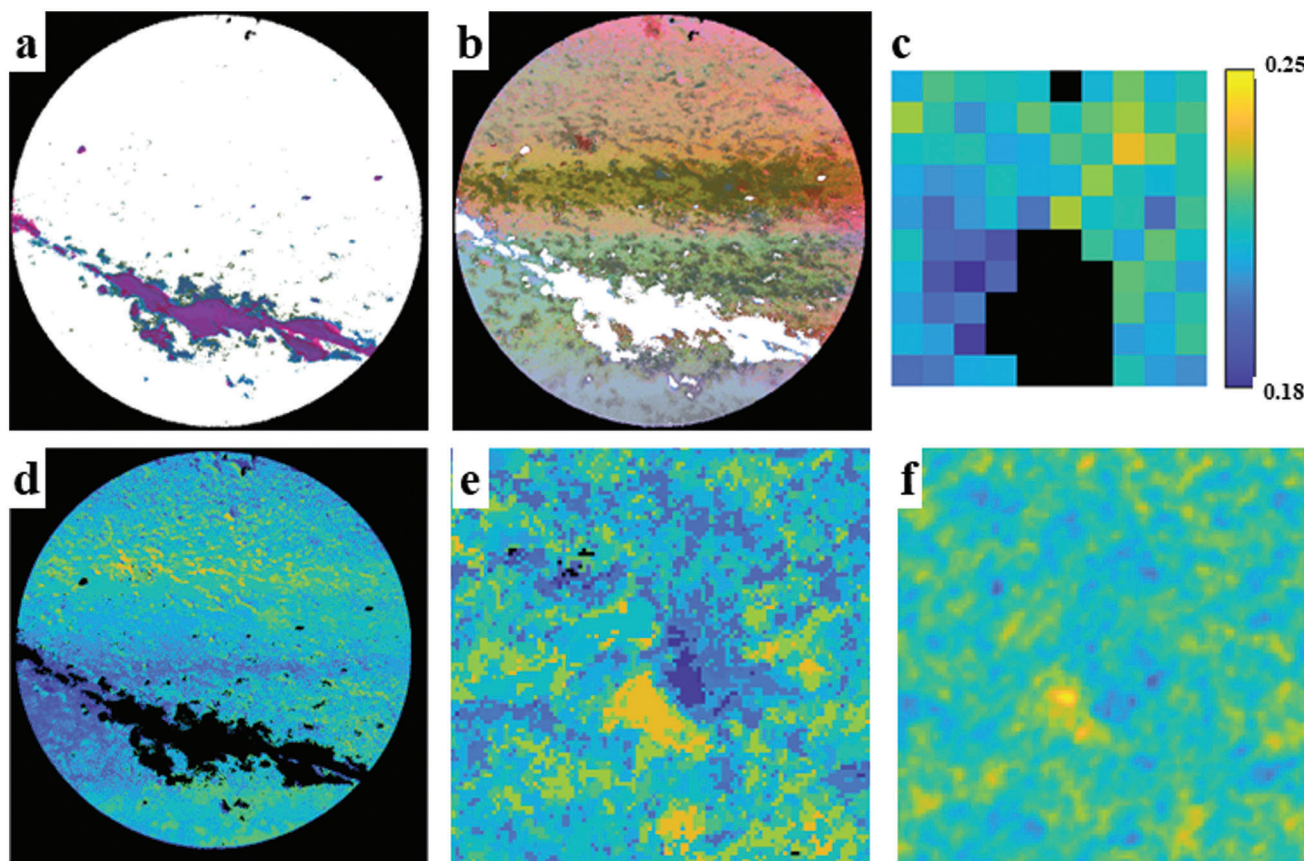


Figure 5. Analysis of the magnetic property of pyrrhotite using SOM-RPM approach. a) Similarity map of iron oxidation product. b) Similarity map after removing pixels (shown in white for contrast) from the oxidized region. c) SOM topographic map recolored by the difference of the Fe^{2+} peak between left and right polarization for the remaining neurons. The color scale shows the asymmetry values represented by each neuron. d) Similarity map colored by degree of magnetism. e) Magnified similarity map colored by degree of magnetism on Region B. f) Magnified XMCD map on Region B in Figure 1e. The image is reproduced here to allow better comparison.

divergence in color assignment between ROIs 1 and 2, as compared to ROIs 3 to 6.

Pixel intensities and neuron intensities are plotted in Figure 4c,d, respectively, to study this observation. LHS polarization is shown in black and the RHS polarization in assigned pixel/neuron assigned color. The grey shaded region in Figure 4d indicates the standard deviation of pixels assigned to each neuron. The insets in Figure 4c,d show the asymmetry values at 707.8 eV of selected pixels ($d_i = \frac{I_{p_i, \text{RHS}} - I_{p_i, \text{LHS}}}{I_{p_i, \text{RHS}} + I_{p_i, \text{LHS}}}$) and mean normalized pixel intensity ($d_i = \frac{\text{Norm } I_{N_i, \text{RHS}} - \text{Norm } I_{N_i, \text{LHS}}}{\text{Norm } I_{N_i, \text{RHS}} + \text{Norm } I_{N_i, \text{LHS}}}$) of selected neurons respectively. Normalized intensity is used for neurons as the pixel intensity is normalized for SOM training. The dotted vertical lines at 707.8 eV and 709.4 eV were added as visual guides. To emphasize, these spectra are the mean spectra of the clustered pixels. They are not an approximation generated by the model.

As stated previously, the low standard deviation calculated (grey shaded region) for these neurons serves to validate the accurate derivation of the spectral features illustrated in Figure 4d from the allocated pixels. ROIs 1 and 2, colored green and blue in Figure 4b, were selected from a slightly oxidized region, as indicated by the emergence of the shoulder of their SOM output

spectra (Figure 4d) at 710.9 eV. Using the conventional data visualization technique, the increase in intensity is challenging to attribute solely to the magnetic property or iron oxidation since magnetic property would affect the overall intensity of the Fe^{2+} peak, which overlaps the iron oxidation peak. ROIs 3 to 9 were selected at different distances around the magnetic domain of unoxidized region, and the spectra of the neurons were consistent with the XMCD maps.

In order to assess the magnetization, and hence the iron deficiency of the pyrrhotite crystal, we implemented a methodology to compare these properties across the surface. Initially, we excluded 17 neurons that represented oxidized regions (as depicted in Figure 5a) from the analysis. This step was necessary to avoid potential interference caused by the alteration of the correlation between iron vacancy and magnetism induced by the oxidation process. Figure 5b displays the remaining pixels after the removal of the oxidized regions. Subsequently, we proceeded to calculate the contrast in the intensity of the Fe^{2+} peak between the left and right polarized beams using the remaining neuron weights. Figure 5c shows the contrast values of the neurons in the SOM, while Figure 5d shows these values mapped back to the pixels' spatial locations. The black neurons in Figure 5c indicate

that the neurons representing oxidation products are clustered together on the SOM.

The discussion presented in Section 3.1 highlights the structural differences between the iron-rich, non-magnetic region comprised of hexagonal FeS crystal and the iron-deficient, highly magnetic region comprised of monoclinic Fe₇S₈ crystal. Figure 5d provides insight into the spatial distribution of magnetic regions in relation to iron oxidation products. Notably, the oxidized regions appear to be surrounded by non-magnetic or low magnetic regions (blue), while the moderately to highly magnetic regions (yellow) are located further away from these products. These findings are consistent with the activation energies of hexagonal and monoclinic pyrrhotite as discussed in Section 3.1. Figure 5e,f present the recolored similarity map and original XMCD map magnified in Region B for direct comparison. The fact that these two maps closely resemble each other indicates a reliable connection between the asymmetry calculated for single pixels and neurons.

2.5. Use of SOM-RPM for Analysis of PEEM Data

Having explored the ways in which a SOM-RPM map can be used to explore XMCD-PEEM data for the pyrrhotite sample, we now turn to discussing, in more general terms, the use of SOM-RPM for PEEM analysis.

First, we would like to emphasize the role of SOM-RPM in the analytical workflow. We note that, like any unsupervised learning algorithm, it should be applied in conjunction with domain expertise. Indeed, SOM-RPM does not present scientifically interesting conclusions on its own – this is the domain of the scientist who is applying the SOM-RPM algorithm. For scientific rigour, it is always important to incorporate domain expertise into the interpretation and evaluation of the model. This study presents a demonstration of this, whereby at each point in the analytical process we have considered the veracity of the SOM-RPM clustering and visualizations (e.g., Figures 3d and 4d and Figures S4–S13, Supporting Information).

Looking more specifically at the spectra shown in Figures 3d and 4d and Figures S4–S13 (Supporting Information): A key feature of SOM-RPM, in contrast with more traditional analytical workflows (such as simply binning neighboring pixels or applying a smoothing algorithm to obtain a smoother spectrum, requiring the sacrifice of spatial or energy resolution), SOM-RPM allows the creation of maps in which unbiased statistical averaging of similar spectra is possible. This allows the retention of the high lateral spatial resolution and energy resolution of the original XMCD-PEEM measurement, making SOM-RPM an ideal algorithm for this type of analysis. Furthermore, in Figure 4c, there are indications of a shoulder in the single pixel spectra of ROI 1 and 2 (the original spectra of ROI 1, 3, 5, 7, and 9 selected in Figure 4 are plotted in the Figure S14, Supporting Information). However, it is very difficult to discern due to the noise in the data, which is a common complication in PEEM based measurements. With the use of SOM-RPM, these pixels with similar spectra are grouped together, such that the resulting similarity map provides the starting point for generating these kinds of insights, which still come directly from the data. The fact that SOM-RPM assigns a different color to these two ROIs highlights its ability to differ-

entiate between pixels with high similarity, enabling the visualization of subtle differences in material phases, which would be easily overlooked using the conventional approach.

After processing the dataset with SOM-RPM, the retained spectral properties enable us to eliminate undesired (oxidized) pixels and apply an alternative color scale to the model (depicted in Figure 5c,d). This produces a direct and clear visualization of the relevant property – magnetic anisotropy. By assigning distinct colors to the SOM, we facilitate the interpretation of variations in Fe²⁺ peak intensity contrast throughout the sample. Figure 5e,f illustrates that the SOM-RPM algorithm refines the definition of magnetic and non-magnetic domains, leading to clearer differentiation compared to the original magnetic contrast image. This results from the averaging out of fluctuations by SOM-RPM.

It is crucial to note that the model's purpose is not to accurately replicate the asymmetry ratio. This information is inherent and straightforward to calculate and visualize. Instead, SOM-RPM offers a means to visualize this asymmetry within the broader context of overall spectral similarity. Conversely, it allows us to explore spectral similarity in relation to asymmetry. This reveals diverse subphases with distinct spectral attributes that align with the same (or nearly identical) asymmetry. Consequently, our approach provides a more comprehensive analysis of the data.

Instead of displaying the data in separate images, which could be cumbersome for researchers to compare and study correlations, SOM-RPM can present an objective, comprehensive representation of multiple datasets, probing different properties within a single image. This is a data reduction and analysis approach which is uniquely well suited to the experimental capabilities of a PEEM instrument. One of the reasons PEEM is considered so useful is that it is possible, with an appropriate apparatus and experimental design, to create multiple high-resolution images which are of the same spatial region of a sample, but different data products (for example, core level X-ray photoelectron spectroscopy, valence band, and XMCD at multiple absorption edges all on the same sample area).

By using SOM-RPM as a tool in PEEM analysis it is possible to build maps correlating similarities in these disparate spectral products and develop a more nuanced understanding of the composition and properties of a material surface at nanometer length-scales. Critically, we note that this unsupervised ML approach to data reduction and visualization requires no a priori assumptions about the sample, and interpretation of the information is left to the researcher, who can apply best practice in the analysis. While the demonstration shown in this work is a natural mineral sample, there is nothing in this workflow which would prevent the use of this approach for characterizing the properties of other samples, such as engineered quantum structures.

3. Conclusion

We investigated the potential of using SOM-RPM for analyzing and visualizing complex PEEM datasets generated by synchrotron-based XMCD-PEEM. Traditional visualization approaches for XMCD-PEEM data may not fully capture the complexity of heterogeneous samples. By applying SOM-RPM to the XMCD-PEEM data acquired from a pyrrhotite sample, we demonstrated the creation of a colored topographic map that represents spectral similarities and dissimilarities among the pixels.

This approach provides a more intuitive and easily interpretable representation of the data without the need for data binning or spectral smoothing. The results of the SOM-RPM analysis were compared to the conventional visualization approach, highlighting the advantages of SOM-RPM in revealing features that were not readily observable using a more conventional method to the data.

The SOM-RPM analysis of the XMCD-PEEM data enabled a valuable visualization of the pyrrhotite sample. By comparing the results to the conventional approach, the SOM-RPM-generated similarity map provided insights into the distribution of iron oxidation products, magnetic and non-magnetic regions, and distinct domains within the unoxidized region. The SOM-RPM visualization approach allowed for the identification of phase boundaries and the characterization of different regions based on their spectral properties. This analysis enhances the visualization of phases with subtle differences embedded in pyrrhotite. These observations demonstrate the utility of our approach in discerning subphases within samples, a critical facet within the micro-mineralogy, paleomagnetic, and geomagnetic domains. Such precision becomes paramount in preventing misinterpretation of minerals that bear high similarities.

The successful application of SOM-RPM to XMCD-PEEM data suggests its potential as a complementary tool for other PEEM-based measurements, making SOM-RPM a useful data reduction tool for understanding correlations in the rich high resolution hyperspectral data that PEEM is capable of generating.

4. Experimental Section

Pyrrhotite Sample Preparation: The sample of mixed pyrrhotite was sourced from Brukung mine, South Australia. The sample was confirmed to be a mixture of monoclinic and hexagonal pyrrhotite with a 45:55 ratio, by quantitative XRD. Powder X-ray Diffraction (PXRD) analysis was carried out at Flinders Microscopy and Microanalysis (Flinders University) using a Bruker D8 Advance Eco Powder X-Ray Diffractometer with a $\text{Co K}\alpha$ radiation source ($\lambda = 1.7902 \text{ \AA}$). The PXRD analysis is presented in Figure S1 (Supporting Information). Samples were ground to a particle size of $<50 \mu\text{m}$ using a quartz mortar and pestle. The sample was analyzed on a zero-background silicon substrate disc. XRD data were all collected across the 2θ range of 10 to 90° with a step size of 0.02° and 0.5 s per step. Qualitative analysis of samples was completed using Diffrac.EVA and matched with reference spectra from the Crystallography Open Database (COD). Sections $\approx 1 \text{ cm}^2$ and 1 mm thick were cut from the block for analysis. The sample surface was polished using wet and dry sandpaper of increasingly fine grain size from 400, 800 to 1200 grit, then polished with $1 \mu\text{m}$ diamond paste. The sample was then sonicated for 3 min in ultrapure water to remove the diamond paste and particles. The sample was then leached in Milli-Q water adjusted to pH 1 with concentrated sulfuric acid, for 1 h just prior to analysis, to clean and remove iron hydroxide.^[25]

Synchrotron XMCD-PEEM Setup: The XPEEM data was collected on beamline BL05B2 at the National Synchrotron Radiation Research Centre (NSRRC) in Hsinchu, Taiwan. The BL05B2 beamline uses an elliptically polarized undulator (EPU5) capable of left, right, and linear polarization, and an Omicron FOCUS-IS PEEM as the microscope. The beamline has a spherical-grating monochromator, yielding very high photon flux (2×10^{12} photons S^{-1} at 800 eV in a $0.4 \text{ mm} \times 0.2 \text{ mm}$) with spatial resolution better than 50 nm.^[26,27] The emitted photoelectrons were accelerated toward an electron lens column and projected onto an aluminum coated yttrium aluminium garnet crystal screen. The real-time, sample surface im-

ages were acquired by a charge coupled device detector mounted behind the screen in total electron yield mode. The analysis chamber vacuum was held at ultrahigh vacuum (10^{-10} Torr) during analysis. An Fe L edge stack of images was acquired while scanning the desired photon energy range across the Fe L-absorption edge, from 700 to 730 eV, with a 0.1 eV step, so that each pixel in the stacked image contained a complete near edge X-ray absorption spectrum. The data was energy calibrated using a Fe wire standard.

Data Export and Preprocessing: The 1024×1024 pixels hyperspectral data was exported from the XSM reader 0.99 software in TIFF format for each energy level in both polarity from 702.0 to 726.0 eV in 0.1 eV step, thus giving 482 features in total for each pixel. The TIFF image stack was then imported into MATLAB R2020b (v9.7) for preprocessing and analysis. To remove low-intensity pixels outside the field of view, 308 400 pixels were filtered out with a total intensity less than 10% of the maximum pixel value, giving 740 176 remaining pixels. To correct for the varying distribution of the light source, the dataset underwent initial processing using a baseline removal function called `msbackadj` implemented in the MATLAB Bioinformatics Toolbox, with a window size of 100 and step size of 50 to remove the background electron intensity. The baseline removal function performs baseline correction on a raw signal by estimating and regressing the baseline within windows, and then adjusts the intensity values based on the estimated baseline. This step aimed to normalize the data and mitigate the effects of non-uniform illumination, ensuring more accurate and reliable results for subsequent analysis.^[28]

SOM-RPM: The detailed algorithm used for the SOM-RPM model is discussed in Gardner et al. 2019,^[11] 2020,^[15] and 2021.^[13] To provide some context, the principle of SOM is explained in terms of the current study. SOM is an unsupervised ML algorithm that can be used to visualize and cluster high-dimensional data in a low-dimensional space. A SOM is consisted of network units known as neurons and each neuron is built up of weight layers. In the study, each weight layer represents the intensity of photoelectrons detected at each energy level, giving 482 weight layers (241 layers for each polarity). The total electron count for each pixel was normalized for SOM training.

As part of the training, each of the 740 176 pixels in the PEEM image was associated with a winning neuron, defined as neuron with closest weight vector (using shortest Euclidean distance from the sample vector). The weight vectors of winning neuron and its neighboring neurons were then iteratively updated to minimize the difference between pixel intensities and the weights of neurons in the high-dimensional space. A principal component-based approach was used to initialize the SOM weights. More details about this initialization method, and the SOM algorithm more generally, are provided in Ballabio et al.^[29–31] The resultant topographic map exhibits a toroidal structure and serves as a low-dimensional model of the data topology, whereby the topological distances in the high-dimensional data are modeled as topographical distances between neurons in the map. In the case, each neuron in the topographic map contains a full weights spectrum from both polarities of the Fe L-edge XANES spectra.

The SOM was then color coded using the RPM technique, which is a multidimensional scaling technique, to provide a model for the SOM itself in which distances between neurons are accurately represented.^[15] Here, International Commission on Illumination (CIE) $L^*a^*b^*$ color space was used, which was designed such that the perceived color change is approximately proportional to the Euclidean distance of the neurons to further improve the accuracy of the resulting map.^[13,15] Finally, the color tagged pixels were plotted in their original position forming an image known as similarity map.

The SOM-RPM workflow, as previously described, was performed using MATLAB R2020b (v9.7).^[11,13] Square SOMs with the size of 6×6 , 10×10 , and 14×14 neurons with a toroidal topology were utilized for pixel clustering. SOMs were trained for 10 000 epochs, which took ≈ 7 h. This number of epochs was selected to ensure convergence. RPM was performed using in-house MATLAB scripts. The SOM training was done on a 16 CPU cluster with 128 GB total RAM. CIE lab color space was used for the RPM algorithm for a more accurate distance-oriented coloring of the SOM neurons.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by an Office of National Intelligence – National Intelligence and Security Discovery Research Grant (NI210100127) and an Australian Research Council Future Fellowship (Harmer) (FT110100099), funded by the Australian Government. This work was performed in part at Microscopy Australia, a company established under the National Collaborative Research Infrastructure Strategy, through the South Australian Regional Facility, Flinders Microscopy and Microanalysis (FMMA), Flinders University. This work was performed in part at the Australian National Fabrication Facility (ANFF), a company established under the National Collaborative Research Infrastructure Strategy, through the La Trobe University Centre for Materials and Surface Science. The PEEM data was collected at the National Synchrotron Radiation Research Centre (NSRRC), Taiwan under proposal 2014-2-077-1 (S.L.H.). The travel funding was provided by the International Synchrotron Access Program (ISAP) managed by the Australian Synchrotron AS/IA133/7716. The authors acknowledge Zoe Pettifer for sample polishing assistance. The authors acknowledge the Milano Chemometrics and QSAR Research Group for the development of the Kohonen and CP-ANN Toolbox for MATLAB.

Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

hyperspectral imaging, machine learning, pyrrhotite, relation perspective imaging, SOM-RPM, XMCD-PEEM

Received: July 6, 2023
Revised: September 13, 2023
Published online:

[1] S. Günther, *Prog. Surf. Sci.* **2002**, *70*, 187.

[2] N. Waterfield Price, R. D. Johnson, W. Saenrang, F. Maccherozzi, S. S. Dhesi, A. Bombardi, F. P. Chmiel, C. B. Eom, P. G. Radaelli, *Phys. Rev. Lett.* **2016**, *117*, 177601.

- [3] R. Juge, N. Sisodia, J. U. Larrañaga, Q. Zhang, V. T. Pham, K. G. Rana, B. Sarpi, N. Mille, S. Stanescu, R. Belkhou, M.-A. Mawass, N. Novakovic-Marinkovic, F. Kronast, M. Weigand, J. Gräfe, S. Wintz, S. Finizio, J. Raabe, L. Aballe, M. Foerster, M. Belmeguenai, L. D. Buda-Prejbeanu, J. Pelloux-Prayer, J. M. Shaw, H. T. Nembach, L. Ranno, G. Gaudin, O. Boule, *Nat. Commun.* **2022**, *13*, 4807.
- [4] M. Foerster, F. Macià, N. Statuto, S. Finizio, A. Hernández-Mínguez, S. Lendínez, P. V. Santos, J. Fontcuberta, J. M. Hernández, M. Kläui, L. Aballe, *Nat. Commun.* **2017**, *8*, 407.
- [5] P. Roese, C. Keutner, U. Berges, P. Espeter, C. Westphal, *Mater. Mater. Trans. A* **2017**, *48*, 1484.
- [6] R. G. Acres, S. L. Harmer, D. A. Beattie, *J. Synchrotron Radiat.* **2010**, *17*, 606.
- [7] V. Lopez-Flores, M. Mawass, J. Herrero-Albillos, A. A. Ünal, S. Valencia, F. Kronast, C. Boeglin, *J. Phys.: Condens. Matter* **2020**, *32*, 465801.
- [8] C. Baldasseroni, G. K. Pálsson, C. Bordel, S. Valencia, A. A. Unal, F. Kronast, S. Nemsak, C. S. Fadley, J. A. Borchers, B. B. Maranville, F. Hellman, *J. Appl. Phys.* **2014**, *115*, 043919.
- [9] T. Ueno, H. Hino, A. Hashimoto, Y. Takeichi, M. Sawada, K. Ono, *Comput. Mater.* **2018**, *4*, 4.
- [10] H. Iwasawa, T. Ueno, T. Masui, S. Tajima, *Quantum Mater.* **2022**, *7*, 24.
- [11] W. Gardner, S. M. Cutts, B. W. Muir, R. T. Jones, P. J. Pigram, *Anal. Chem.* **2019**, *91*, 13855.
- [12] W. Gardner, A. L. Hook, M. R. Alexander, D. Ballabio, S. M. Cutts, B. W. Muir, P. J. Pigram, *Anal. Chem.* **2020**, *92*, 6587.
- [13] W. Gardner, S. M. Cutts, D. R. Phillips, P. J. Pigram, *Biopolymers* **2021**, *112*, 23400.
- [14] S. E. Bamford, D. Yalcin, W. Gardner, D. A. Winkler, T. M. Kohl, B. W. Muir, S. Howard, E. A. Bruton, P. J. Pigram, *Anal. Chem.* **2023**, *95*, 7968.
- [15] W. Gardner, R. Maliki, S. M. Cutts, B. W. Muir, D. Ballabio, D. A. Winkler, P. J. Pigram, *Anal. Chem.* **2020**, *92*, 10450.
- [16] C. Mang, A. Kontny, D. Harries, F. Langenhorst, L. Hecht, *Meteorit. Planet. Sci.* **2012**, *47*, 277.
- [17] N. Belzile, Y.-W. Chen, M.-F. Cai, Y. Li, *J. Geochem. Explor.* **2004**, *84*, 65.
- [18] C. I. Pearce, *Rev. Mineral. Geochem.* **2006**, *61*, 127.
- [19] C.-S. Horng, A. P. Roberts, *Earth Planet. Sci. Lett.* **2006**, *241*, 750.
- [20] Y. Boyjoo, H. Shi, E. Olsson, Q. Cai, Z.-S. Wu, J. Liu, G. Q. (M.) Lu, *Adv. Energy Mater.* **2020**, *10*, 2000651.
- [21] Yu. Mikhlin, Ye. Tomashevich, *Phys. Chem. Miner.* **2005**, *32*, 19.
- [22] H. Miyauchi, T. Koide, T. Shidara, N. Nakajima, H. Kawabe, H. Fukutani, K. Shimada, A. Fujimori, K. Iio, T. Kamimura, *J. Electron Spectrosc. Relat. Phenom.* **1996**, *78*, 259.
- [23] C. Baldasseroni, C. Bordel, A. X. Gray, A. M. Kaiser, F. Kronast, J. Herrero-Albillos, C. M. Schneider, C. S. Fadley, F. Hellman, *Appl. Phys. Lett.* **2012**, *100*, 159924.
- [24] H. Wang, I. Salvason, *Phase Transitions* **2005**, *78*, 547.
- [25] X. Tang, Y. Chen, *Int. J. Min. Sci. Technol.* **2022**, *32*, 847.
- [26] A. D. Smith, G. Cressey, P. F. Schofield, B. A. Cressey, *J. Synchrotron Radiat.* **1998**, *5*, 1108.
- [27] D. H. Wei, Y.-L. Chan, Y.-J. Hsu, *J. Electron Spectrosc. Relat. Phenom.* **2012**, *185*, 429.
- [28] M. S. Claus, S. Gerd, *Rep. Prog. Phys.* **2002**, *65*, 1785.
- [29] D. Ballabio, M. Vasighi, *Chemom. Intell. Lab. Syst.* **2012**, *118*, 24.
- [30] D. Ballabio, M. Vasighi, V. Consonni, M. Kompany-Zareh, *Chemom. Intell. Lab. Syst.* **2011**, *105*, 56.
- [31] D. Ballabio, V. Consonni, R. Todeschini, *Chemom. Intell. Lab. Syst.* **2009**, *98*, 115.