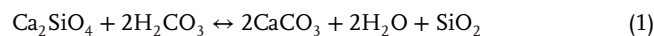# Interrogation of Ecotoxic Elements Distribution in Slag and Precipitated Calcite through a Machine Learning-Based Approach Aided by Mass Spectrometry

Faisal W. K. Khudhur,* Matthew Divers, Mark Wildman, John M. MacDonald, and Joshua Franz Einsle

CO_2 mineralization in slag has been widely investigated as a potential solution for offsetting steelmaking industry emissions. However, it can be associated with ecotoxic elements release (e.g., V and Cr). The presence of such elements in heterogenous slag at the micro-scale remains difficult for analysis since microstructural features can be missed during microscopy data inspection, thereby presenting a challenge in understanding how ecotoxic elements exist in slag. Here, an unsupervised machine learning-based technique is used to analyze slag's microstructural features. Energy Dispersive Spectroscopy (EDS) data are analyzed through Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) method. Results show that passive CO_2 mineralization has occurred in situ in the studied samples, on the surface, and within their pores. Additionally, V and Cr regions with equivalent diameters < 42 µm can exist within slag, potentially making such elements prone to mobilization due to slag pulverization. Interrogation of the samples with Laser Ablation Inductively Coupled Plasma Mass Spectroscopy (LA-ICP-MS) confirms the distribution of the elements obtained from the clustering algorithm and further demonstrates that up to 84 and 9 ppm of V and Cr are incorporated in the precipitated calcite, respectively. This implies that ecotoxic elements may be immobilized through calcite precipitation.

1. Introduction

Ironmaking and steelmaking industry are vital to infrastructure and economic development as steel is an essential component in various industries, with crude steel production reaching 160 Mt in 2021.[1] However, this industry generates ≈7% of global anthropogenic CO_2 emissions, at a rate of 1.8 kg of CO_2 per kg of steel.[2,3] The increased awareness of sustainable development goals resulted in several proposals to offset the CO_2 emitted from this industry, as documented in[4] and the references therein. One of these proposals is to sequester CO_2 through passive mineralization. This process follows a waste-to-resource pathway by utilizing slag – a byproduct from this industry that is composed of metals and silicon oxides – as a reactant. Passive CO_2 mineralization has been documented at different locations worldwide.[5] For example, slag heaps near Consett, UK have captured up to $7.8{\times}10^4$ kg CO_2 y^{-1} [6] However, as this process occurs with minimal human intervention, the rate of CO_2 uptake can be slow due to different reasons, including low CO_2 concentration in atmospheric air and low dissolution rates of several minerals that constitute these wastes. Consequently, attempts have been made to enhance CO_2 uptake by enhancing watering frequency and reducing the size of the slag lumps to allow for a larger reactive surface area.[7,8] Passive CO_2 mineralization occurs during slag weathering as various minerals in slag dissolve to provide Ca^{2+} cations that react with dissolved carbonates to precipitate calcite, as exemplified in the cases of larnite (Ca_2SiO_4) and gehlenite $(Ca_2Al_2SiO_7)$, in Equations (1) and (2), respectively:[9–11]

$$Ca_2SiO_4 + 2H_2CO_3 \leftrightarrow 2CaCO_3 + 2H_2O + SiO_2 \tag{1}$$

$$2Ca_2Al_2SiO_7 + 2H_2CO_3 \leftrightarrow 2CaCO_3 + 2H_2O + Al_2O_3 + SiO_2 \tag{2}$$

However, the dissolution of slag is associated with releasing ecotoxic elements that originate from the raw materials that were used in the ironmaking and steelmaking process. Upon interaction with water, slag can produce leachates that contain oxyanions like As, Cr, and V.[12] The elevated content of these elements (e.g., 120 µg l^{-1} of V) in the leachates, as well as their alkaline pH, can exceed the environmentally

F. W. K. Khudhur, M. Divers, M. Wildman, J. M. MacDonald, J. F. Einsle
School of Geographical and Earth Sciences
University of Glasgow
Glasgow G12 8QQ, UK
E-mail: f.khudhur.1@research.gla.ac.uk
M. Divers
Department of Civil and Environmental Engineering
University of Strathclyde
Glasgow G1 1XJ, UK

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

acceptable discharge limits, thereby making them a potential hazard to nearby water bodies.[13,14] However, CO_2 mineralization can immobilize the ecotoxic elements that are released during weathering of tailings.[15] For example, at Woodsreef in Australia, the hydromagnesite ($Mg_5(CO_3)_4(OH)_2 \cdot 4H_2O$) and pyroaurite ($Mg_6Fe^{3+}_2(OH)_{16}CO_3 \cdot 4H_2O$) formed as a result of CO_2 mineralization contained up to 100's ppm of transition elements (Cr, Co, Ni, V).[15] Previous work observed that elements leaching from slag are largely affected by the CO_2 mineralization process.[16,17] In the absence of CO_2, the leachates from the slag had a pH value of 11.9 and a V concentration of 493 µg L^{-1}. However, as CO_2 is introduced, it dissolves in water and consumes the OH^- to produce carbonates which in turn react with leached Ca^{2+} from slag to produce $CaCO_3$, leading to a reduced pH value of 8.0 and a V concentration of 859 µg L^{-1}.[17]

Slag mineralogical composition depends on the properties of the ore used, the processing method, and its management practice, resulting in compositionally heterogeneous slag at macroscopic and microscopic levels. Previous work investigated the bulk properties of slag such as their mineralogy and major and minor elements contents.[14,18] However, there is a lack of understanding as to how ecotoxic elements distribution can affect the applicability of slag in passive mineralization applications since these elements are prone to mobility during slag weathering. For example, it is not clear if such elements exist within well-defined grains or if they preferentially substitute cations in different slag minerals. Understanding the nature of ecotoxic elements presence in slag will help in designing waste management practices that promote CO_2 mineralization and hinder the release of such elements to the environment. Scanning Electron Microscopy (SEM) coupled with Energy Dispersive Spectroscopy (EDS) analysis has been used to visualize the elements distribution in slag and reaction fronts at which minerals dissolve and calcite precipitates.[19,20] Increased automation makes the collection of large area (i.e., 10's of mm^2) hyperspectral maps a routine task. The collected maps consist of large datasets that can contain millions of spectra and can be exported as data arrays. Consequently, machine learning algorithms can be leveraged to determine the present phases and to extract valuable information related to the studied materials. To illustrate, while it is difficult to visualize minor elements across a thin section, since such elements may exist within fine structures that are easily missed, dimensionality reduction and clustering algorithms enhance the signal-to-noise ratio and provide a method to identify such structures and extract their representative spectra.[21] In this work, we use dimensionality reduction and clustering techniques to study EDS data collected from slag samples, to segment the different regions within these samples as per their chemical compositions, and to visualize the distribution of low-concentration (<5%) elements. The purpose of this technique is to identify the composition of different regions in slag, thereby assisting in minerals identification. The clustering algorithm is further used to identify regions of interest where ecotoxic elements (e.g., V, Cr) reside, and to identify regions on which line-scans LA-ICP-MS measurements were performed. EDS and LA-ICP-MS quantification methods were compared to discuss the similarity of measurements obtained from both techniques. We further investigated ecotoxic elements immobilization by measuring their concentrations in the precipitated calcite through LA-ICP-MS.

2. Results

The studied samples were collected from a legacy slag site near the former Ravenscraig steelworks Scotland, where ironmaking and steelmaking operations produced Blast Furnace (BF) and Basic Oxygen Furnace (BOF) slag.[22] The polished thin sections of these samples were investigated by Backscatter Electrons (BSE) imaging, EDS, and Raman spectroscopy to identify the present elements and mineralogy of the samples. **Figure 1** shows the BSE images (with a magnification of x400, having a pixel size 1.4 µm) of the studied slag samples. These images show that the samples possess different phases, as evident by the difference of the grey-level values that reflect different chemistry within the samples, that is, areas appear brighter as their content of elements with higher atomic numbers increases. The BSE images also show that the void space within the slag can host material with a uniform grey level. This material can also form on the external surface of the slag. Examples of EDS spectra taken from circular regions demonstrate that this material is composed of Ca, O, and C (spectra A1, B1, and C1) while the slag is composed of Mg, Al, Si, Ca, Fe, and O (spectra A2, B2 and C2). Note that the EDS technique has been reported to have a detection limit of 0.1%−0.5% by weight, making it unsuitable for trace elements detection, with the latter being present at 0.01% scale.[23–26] Figure 1b shows that for sample B, a distinctive intermediate layer forms on the external surface of the slag. This layer contains several localities with varying grey level values, reflecting the heterogeneity in this layer. Based on the measured Raman shifts (**Figure 2**), the observed material on the slag surface and the pore linings is indexed as calcite and the slag minerals in samples A and C resemble åkermanite-gehlenite. However, on different locations in sample B the Raman spectra are inconclusive, showing a large peak broadening between 700 and 900 cm^{-1}, as exemplified by Raman spectra B2 and B3.

Due to the apparent heterogeneity of the samples and the inconclusiveness of the acquired Raman spectra, the acquired EDS maps of the samples were further used to segment each sample to areas based on their chemical composition. This segmentation is achieved through dimensionality reduction and clustering. The dimensionality reduction was performed by Principal Component Analysis (PCA) through Singular Value Decomposition (SVD), and it demonstrates that the studied data can be represented by 5 (sample A), 8 (sample B), and 5 (sample C) principal components, with these components being able to explain up to 42% (sample A), 32% (sample B) and 18% (sample C) of the total variance (**Figure 3a**). These principal components can be thought of as orthogonal compositional vectors that provide a reasonable representation of the complete dataset while being significantly smaller in number when compared to the number of actual measurements (1024 measurements at each pixel in the current study), as explained elsewhere.[27] Note that as shown in **Table 1** (in the Experimental section) and Figure 3a, higher cumulative explained variance at a given number of components is associated with higher counts per pixel, thereby resulting in a higher signal-to-noise ratio. Beyond this point, adding further components does not increase the cumulative explained variance by more than 0.5% per component, meaning that additional components do not help in explaining the data further and can be considered as noise. Figure 3b–d) shows the acquired EDS data
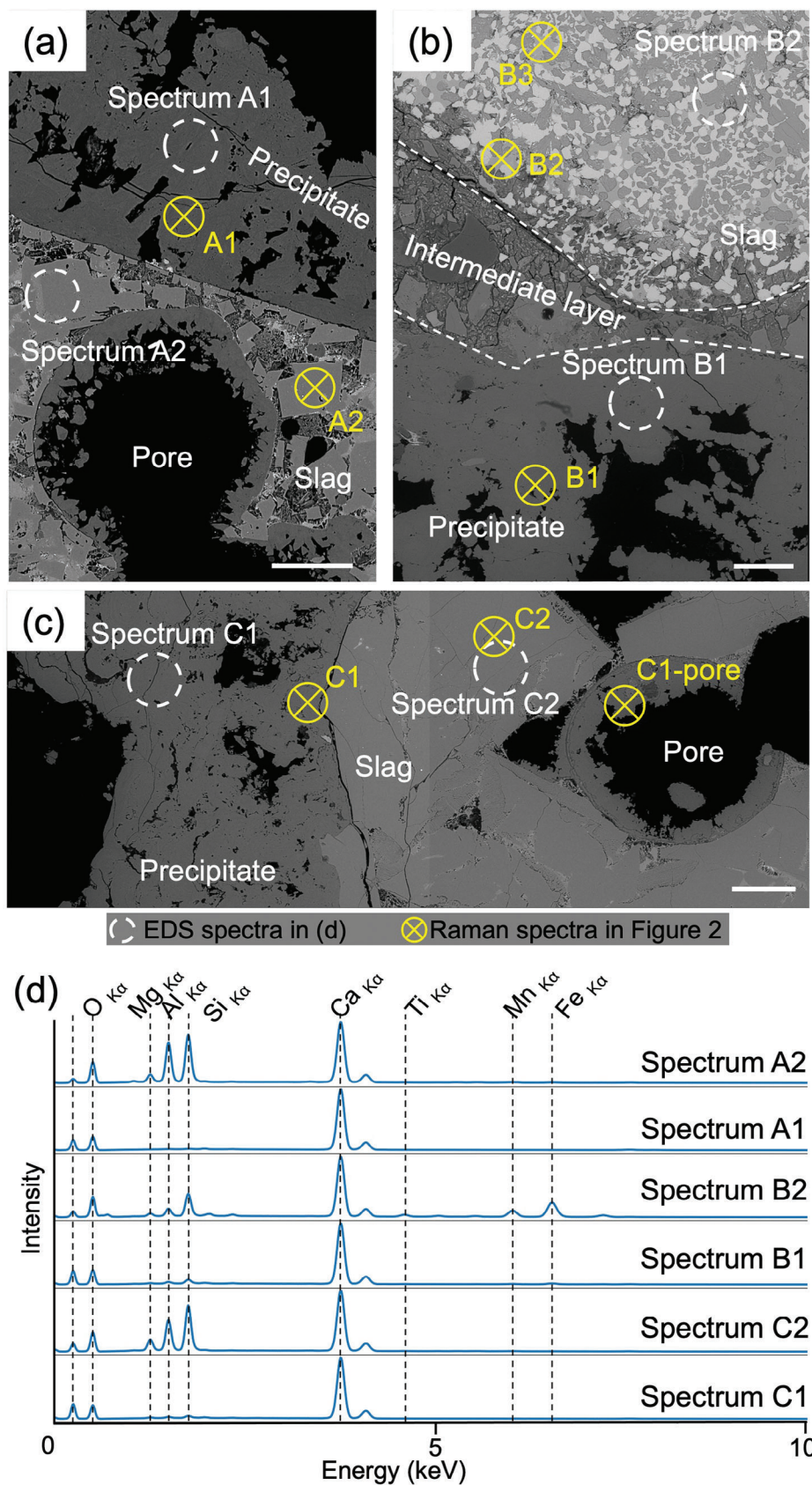
2300559 (2 of 14)

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

Figure 1. Backscatter electron image for samples A(a), B(b), and C(c). Scalebars are 100 μm.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
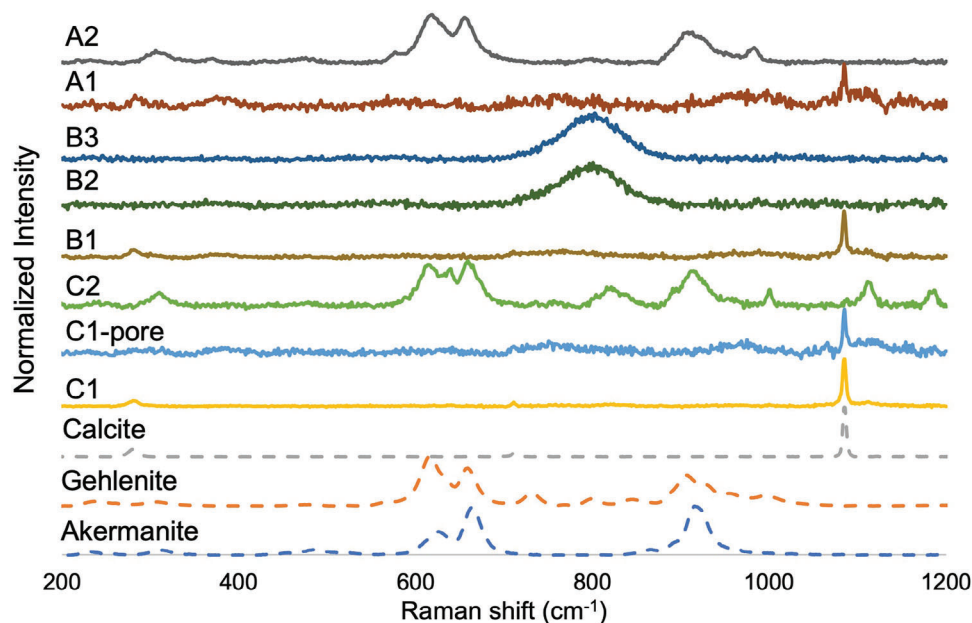SYSTEMS**

www.advsustainsys.com

Figure 2. Raman spectra were collected from locations annotated in Figure 1. Dashed lines represent Raman spectra of the standards as collected from the RRuff database.
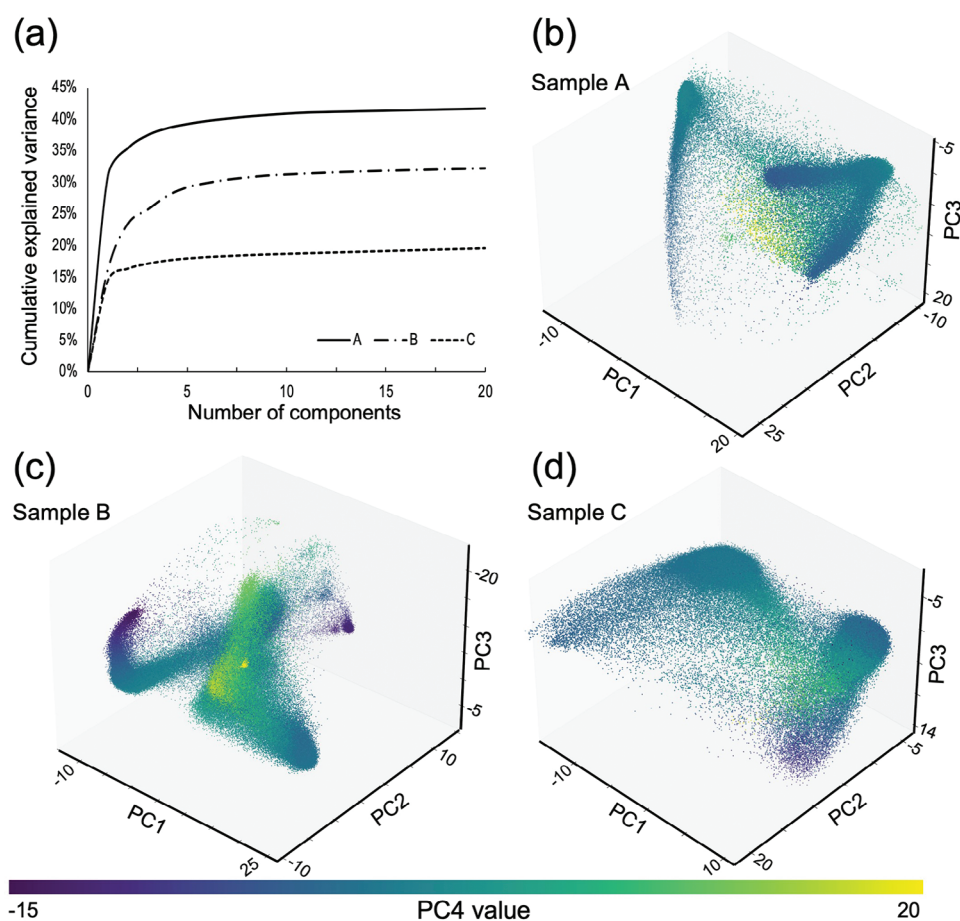

Figure 3. Results of dimensionality reduction, show the cumulative explained variance as a function of the number of principal components for the studied samples a). Images b), c), and d) shows the EDS data plotted as a function of their PC1, PC2, and PC3, and color-coded as per their PC4 values.

2300559 (4 of 14)

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

Table 1. SEM and EDS settings were used for data collection.

SEM settings	Sample A	Sample B	Sample C
Accelerating voltage [kV]		20	
Aperture size [µm]		60	
Dwell time [µs]		250	
Frame count		200	
Magnification		400	
Maps Properties			
Physical size [µm x µm]	728×546	734×1040	1380×545
BSE image pixel size [µm]	0.71	0.71	0.71
EDS map dimensions [pixels]	512×384	516×728	972×383
EDS map pixel size [µm]	1.42	1.42	1.42
Counts per pixel	2725	1792	1665
Live time [hour]	2.7	5.2	5.2
Data size [GB]	1.6	3.0	3.0

when plotted as per their principal components (PCs) values, depicting how these data group together and form assemblies. These assemblies can be spotted in Figure 3 although the number of assemblies differs across the studied samples due to varied chemistry. For example, Figure 3c depicts several areas of high point density, reflecting the larger heterogeneity in that sample. Some smaller assemblies in Figure 3c are evident, reflecting the presence of several clusters with each having a distinct chemical composition. Figure 3d only shows that data points group around two main clusters and the rest of the points are dispersed in the data space. **Figure 4** depicts the chemical signature of different PCs of sample A (Figures S2 and S3 for samples B and C, respectively, Supporting Information). The output components in Figure 4b do not always exist in the actual measurement,[28] as exemplified in the negative peaks of some elements. In these spectra, a positive peak value reflects a positive correlation with a given energy channel, while a negative peak value reflects a negative correlation.[29] Figure 4c,g depicts the values of each principal component at each pixel across the studied samples. For example, brighter regions in the loading map of PC1 (4c) demonstrate a stronger contribution of the spectrum described by PC1. In those bright regions, one would expect higher abundances of all the positive elemental peaks (Mg, Al, Si, Ti, Mn) and an absence of negative elements (C, Ca). Conversely, in the darker intensity regions of the loading map, one would expect to find higher concentrations of the negative elemental peaks, with an absence to low contribution from positive peaks. The dimensionally reduced data were then clustered using the HDBSCAN algorithm, and the clustering results reveal the existence of 7 clusters in sample A (with outliers representing ≈10%, **Figure 5a**), 14 clusters in sample B (with outliers representing ≈30%, Figure 5b) and 5 clusters in sample C clusters (with outliers representing = ≈20%, Figure 5c). The obtained cluster maps were then back-projected on the original EDS maps so that physically interpretable spectra of these clusters could be obtained. The thus obtained spectra can be used to quantify different elements in a given in a given cluster thereby allowing minerals identification. For example, the blue clusters contain a Ca to O molar ratio of ≈0.3, which corresponds to the ratios of these elements in calcite ($CaCO_3$). The cal-

cite cluster can be identified in all the samples, while åkermanite-gehlenite dominates the slag in samples A and C as evident from the clusters distribution map. In addition to the calcite cluster, Sample B contains several clusters with variable compositions, as exemplified in Figure 5b. The annotated cluster in sample B (Figure 5b) is of particular concern since it contains pronounced peaks of Cr and V which are known for their ecotoxicity. EDS quantification in **Figure 6c,d** demonstrates that this cluster is predominately composed of Ca (30.9 wt.%), O (33.6 wt.%), Fe (11.1 wt.%), and Ti (11.1 wt.%), with Cr (1.7 wt.%) and V (1.5 wt.%). This cluster appears at different locations within the studied area, and it is characterized by irregular shapes and particles having equivalent diameters of up to 42 µm. Other phases in Sample B show no detectable concentration of any ecotoxic elements, and they contain variable amounts of Fe, Ca, Mn, Al, and Ti. LA-ICP-MS results in Figure 6e,g show that elevated concentrations of V and Cr (9000 and 11 000 ppm, respectively) coincide with the locations of the Cr/V-containing cluster. We highlight that the clusters distribution map shown in Figure 6b and the elements distribution shown in Figure 6e,g differ since the spot size in the LA-ICP-MS line scans is ≈38 times larger than the pixel size of the EDS analysis (38 vs 1.4 µm). Various locations within the calcite that precipitated on the external surface of the slag show variable incorporation of different elements with V having a maximum composition of 84 ppm, while Mn concentration reaches 5420 ppm in calcite precipitated within the pore space (**Figure 7**).

3. Discussion

3.1. Samples Properties

It is important to note that slag is a heterogeneous material, therefore the present results may only be representative of the studied samples. Nevertheless, these samples were selected for analysis as they reflect different calcite precipitation patterns (e.g., in pores or on external surfaces) as well as different compositions that are commonly encountered in slag.[30] Therefore, the present results can give an idea about the nature of CO_2 mineralization in slag and the existence of ecotoxic elements. The BSE and EDS imaging of the studied samples portray typical features of slag shown in previous studies performed on slag samples collected from this site and from different sites.[31,32] These features include the coexistence of multiple phases with variable compositions as well as the existence of pore space within the samples, thereby providing potential permeability and surface area at which slag dissolution and calcite precipitation reactions can occur. Raman spectroscopy and EDS clustering analysis of samples A and C show that they contain åkermanite-gehlenite minerals, which are commonly observed in BF slag, and their presence has been documented in this site and several sites worldwide.[14,33–36] For sample B, however, Raman spectra taken from different spots within the matrix could not be indexed. We speculate that these observed Raman spectra are a consequence of alteration that different mineral phases are subjected to during the steelmaking process, or during the chemical alteration that slag may be subjected to during hydration and subsequent minerals dissolution. For sample B, quantification of several clusters shows low concentration of Si (<10 wt %) and high concentration of iron (up to 40 wt %), suggesting that this sample may be a BOF slag.
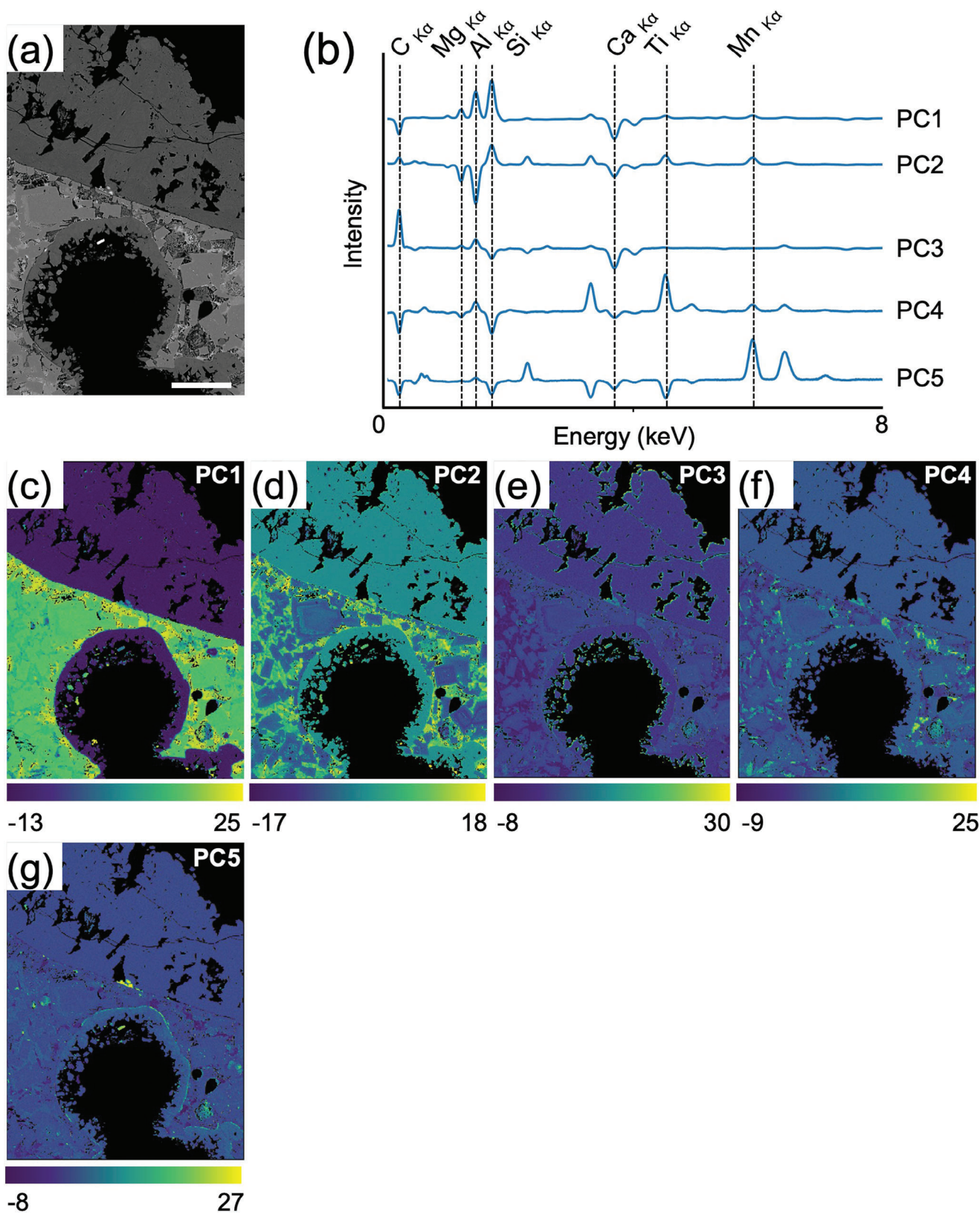
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

Figure 4. Demonstration of dimensionality reduction of Sample A through PCA. a) BSE image of sample A. The scale bar is 100 µm. B) The chemical signature of each PC. Images c) through g) show the values of PC1 through 5 (score maps) at each pixel in the image.
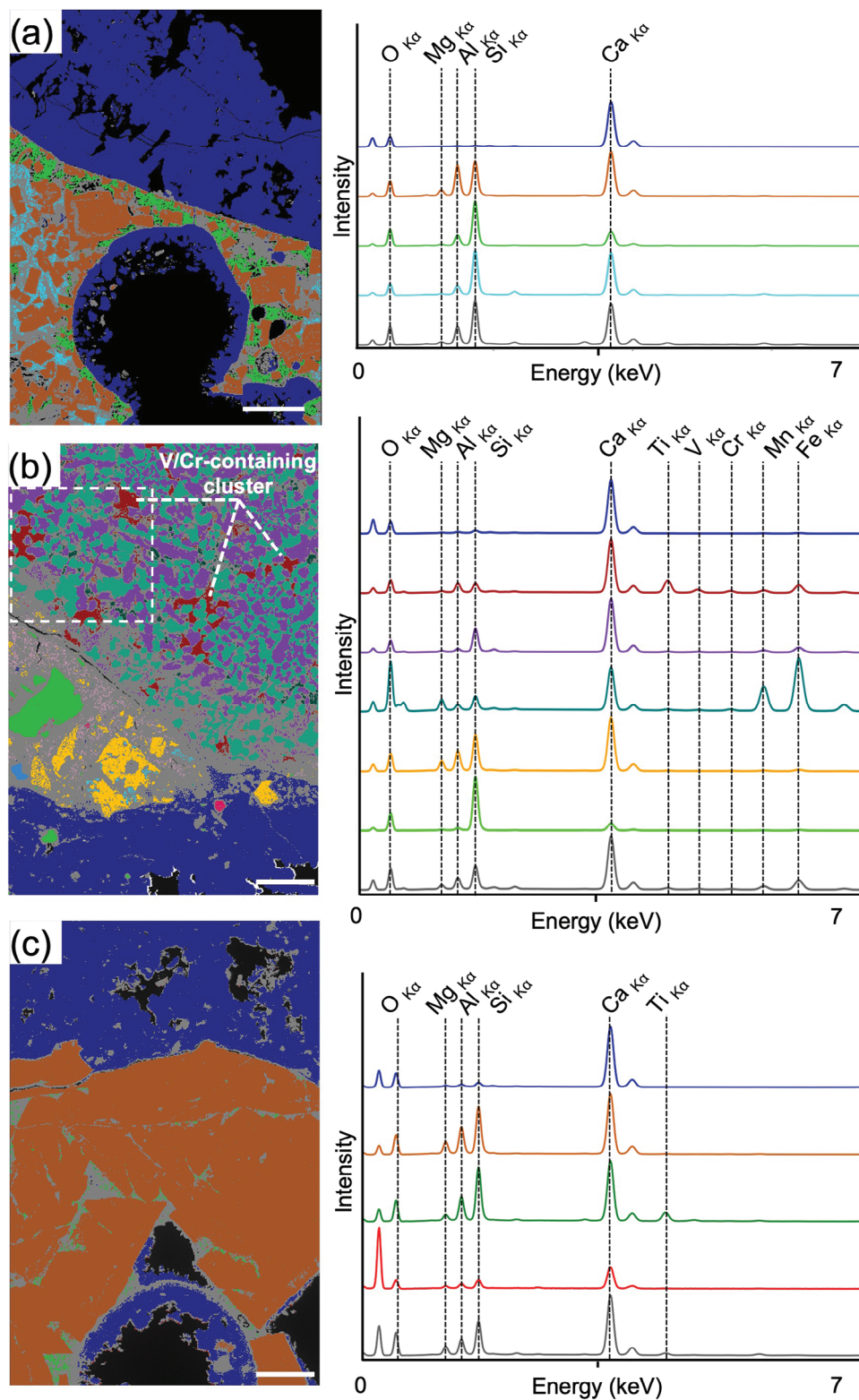
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

Figure 5. HDBSCAN results of EDS data of samples A(a), B(b), and C(c). Each color represents a distinct cluster. Scale bars are 100 µm. The color of each cluster corresponds to that shown in the spectra plot. The grey color represents outliers. The boxed area in image B is magnified in Figure 6.
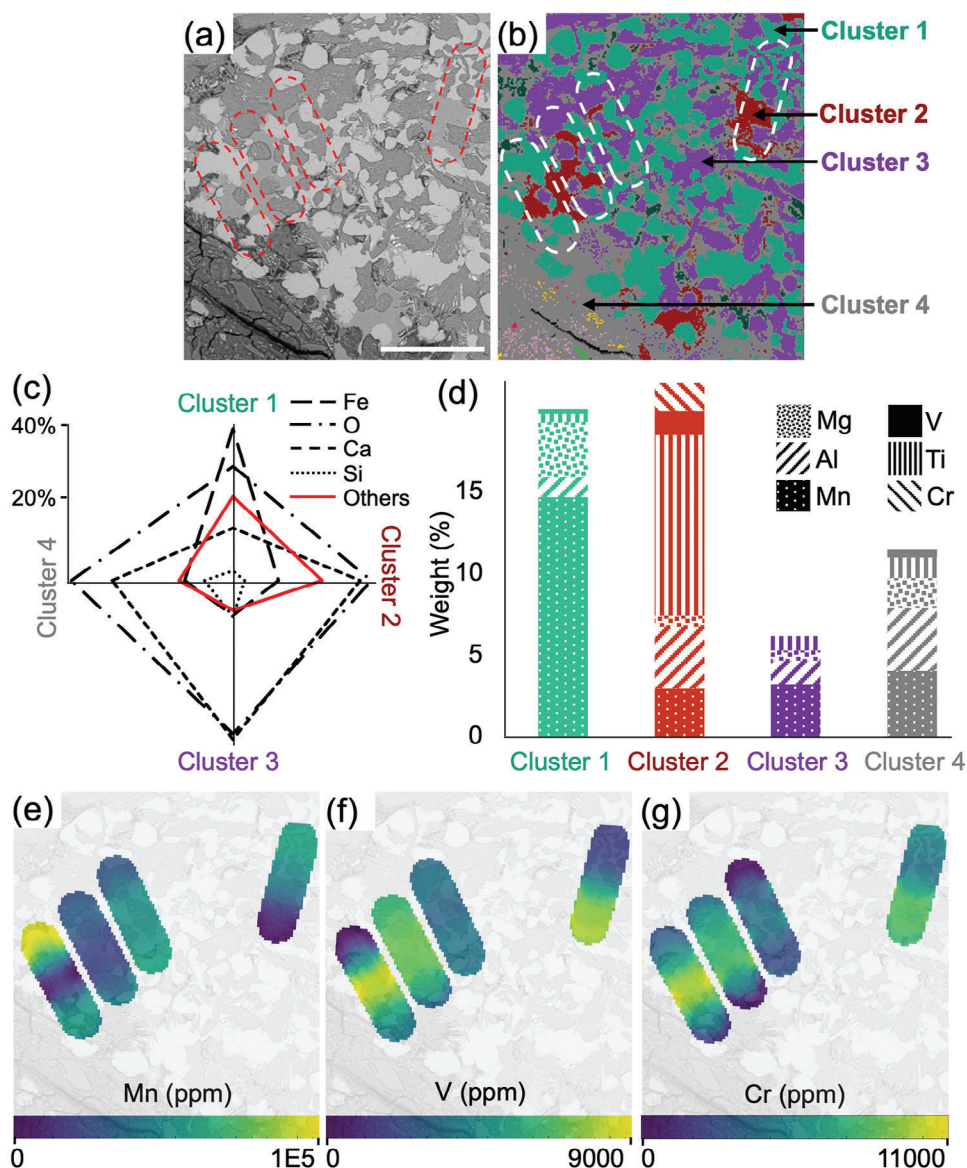
Figure 6. Concentrations of different elements across sample B. a) shows a BSE image of sample B and shows the locations of the lines scanned by the LA-ICP-MS across the sample. b) magnified results of the HDBSCAN. c) and d) show the concentrations of the elements in different clusters, as obtained by the EDS standardless, carbon-free quantification. e), f), and g) show the concentrations of Mn, V, and Cr, respectively, in the dashed regions in images (a) and (b) as obtained from the LA-ICP-MS.

While samples A and C contain åkermanite-gehlenite minerals, which slowly dissolve to produce Ca^{2+} that can be used to capture CO_2 to produce $CaCO_3$,[37] it is evident that passive CO_2 mineralization and calcite precipitation did occur across all the studied samples, on their external surface and within their pore space. Unlike our previous observations, however, the thickness of precipitated calcite reaches 300 µm versus a maximum thickness of 50 µm observed in our previous study.[36] We attribute this to the fact that the currently investigated samples were collected from a pond near the slag deposit site. Consequently, the water enhanced the calcite precipitation, in contrast to our previous study in which calcite precipitation took place within the pore space of the samples. While the samples investigated in this study have thicker calcite layers on their surfaces, we highlight that calcite precipitation can provide a negative feedback mechanism to CO_2 mineralization. This is because carbonate precipitation may form an impermeable barrier that hinders calcite dissolution and further CO_2 mineralization, analogous to experimental CO_2 mineralization in columns and in natural settings.[38,39]

3.2. Clustering Results Analysis

The dimensionality reduction of the studied samples shows that with the chosen number of PC's, the cumulative explained variance reaches ≈42%, reflecting that additional components still

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
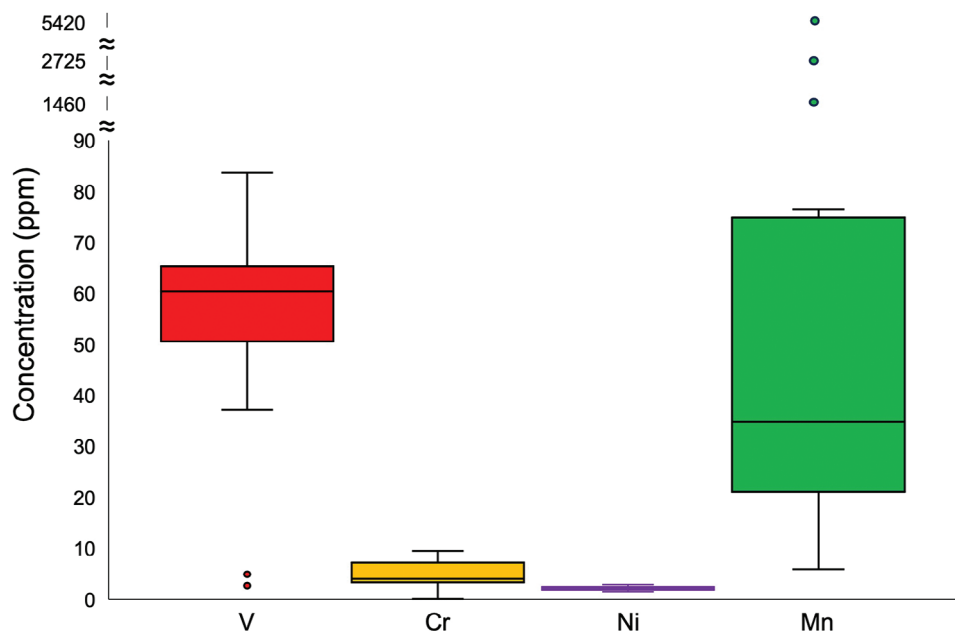SYSTEMS**

www.advsustainsys.com

Figure 7. Median-inclusive box and whisker plot of the concentration of different elements in the precipitated calcite in the three studied samples as determined by LA-ICP-MS. Outliers reflect concentrations in calcite that precipitated in the pore space. Values and errors of each data point are presented in Supplementary Table S1 (Supporting Information).

explain the data, as shown in Figures S4-S6 (Supporting Information). Nevertheless, the cutoff point used in our analysis was chosen by observing the point beyond which additional components increase the cumulative explained variance by less than 0.5%. Additional components increase the computational time, and may reduce the effectiveness of the clustering performance due to the curse of dimensionality as reported elsewhere.[28,40] The application of the HDBSCAN method on the dimensionally reduced data depicts that the studied samples exhibit a variety of complexity and that the provided workflow can analyze samples containing several phases that have characteristic compositions. Samples A and C containing 7 and 5 clusters respectively, including up to ≈20% outliers, while sample C shows 14 clusters with ≈35% outliers, as depicted in Figure 5. This can be attributed to the heterogeneous nature of slag as its chemical and mineralogical compositions strongly depend on the property of the ore used, processing technique, and waste management practice, with the latter strongly affecting its crystallinity and carbon capture potential.[41] Additionally, samples are continuously altered by different environmental factors. For example, water can percolate into the slag heap and result in the dissolution of minerals, or it can carry ions that precipitate on the surface of the slag or within its pore space. To illustrate, in sample B, it is observed that an intermediate layer is evident between the calcite cluster and the matrix phase. We previously documented the existence of similar layers at the interface between slag minerals and the precipitated calcite in pore space within other slag samples,[36] where we attributed the observed layers to the dissolution of calcium ions from complex minerals that dissolve non-stoichiometrically as reported elsewhere.[42] Our previous observations demonstrated that this layer is homogenous and predominantly composed of Si and O. In sample B, however, the clustering result shows that this layer has low Si content (< 5%) and it

contains several constituents of variable chemical compositions, as exemplified in Figure 5b. Prior to calcite precipitation, this layer represents the portion of slag that is subjected to chemical alteration, mostly due to interaction with water. This water contains carbonate ions that are consumed in the carbonation reaction to produce calcite. It can also carry other elements that can be immobilized due to calcite precipitation, as shown in the colored localities within this region. Across the studied samples, outliers can appear within grains because of surface defects, or more commonly at clusters boundaries. This is in part due to the interaction volume that generates EDS signals can be larger than the pixel resolution, therefore producing outliers where grains of different chemical composition overlap. This issue is more evident in Sample B, which contains several clusters within the slag matrix, resulting in a larger content of outliers within the matrix. Additionally, we note that the data of each cluster shown in Figure 7 can be exported and further analyzed through P/B ZAF EDS quantification. That is, the raw spectrum of a given spectra can be exported and then quantified to gain more insight into mineral identification. For example, the blue cluster in all the samples shown in Figure 7 has Ca/O atomic concentration ratios of ≈0.3, consistent with the corresponding ratio of these elements in calcite ($CaCO_3$) and with the Raman spectra acquired from within these clusters. Likewise, the brown cluster in samples A and C is shown to have a Ca/Si atomic ratio of 1.8. This value is between 1 for åkermanite ($Ca_2MgSi_2O_7$) and 2 for gehlenite ($Ca_2Al_2SiO_7$). These minerals are endmembers that form solutions that are found in slag,[43] in agreement with the provided Raman results that showed similarity of the acquired spectra with the åkermanite and gehlenite standard spectra (Figure 2). However, a larger deviation is observed when comparing the atomic ratio of Ca/(Al+Mg+Si), calculated to be 0.8 and 0.9 in samples A and C, to the corresponding value in åkermanite/gehlenite

($\approx$0.7). This could be a result of the material measured being not a pure end member and some pixels being positioned at grain boundaries. The latter can cause boundary pixels to have contributions from more than one mineral, therefore presenting a deviation from the ideal mineral compositions. Additionally, the provided EDS quantification through the P/B ZAF method is based on a standardless method, which makes it more prone to errors when compared to standard-based methods.

Figure 6b shows a magnified image of sample B, demonstrating the large heterogeneity of this sample. EDS quantification shows that the largest content of V and Cr is in cluster 2, which is present at various locations within the studied thin section, and it has an equivalent diameter of up to 42 μm. Previous work demonstrated that slag can be used in passive CO_2 mineralization, although the observed CO_2 uptake is usually low, that is, in the order of 60 kgCO_2/1000 kg slag.[36,44] Consequently, it has been suggested that slag size reduction through pulverization can be used to increase the CO_2 uptake rate. For example, the size reduction of slag average diameter from 532.1 to 43.9 μm increased the CO_2 uptake by around two orders of magnitude under the same experimental conditions, and they attributed this rise to increasing the dissolution rate of slag minerals.[45] However, based on the present analysis, we highlight that the mobility of ecotoxic elements should be further evaluated at 42 μm or less can. Previous work suggested that the size reduction of rocks and ores until particle size reaches grain size makes the elements that reside in these grains more mobile.[46] In light of the current findings, size reduction to below 42 μm could make the observed Cr and V more prone to mobilization due to the enhanced surface area that is available for elements leaching.

3.3. Sequestration of Ecotoxic Elements in the Precipitated Calcite

The LA-ICP-MS results on Sample B in Figure 6e,g exemplify how slag can contain various amounts of ecotoxic elements. Inspection of the locations of cluster 2 with respect to the laser path demonstrates that this cluster coincides with areas of high Cr and V concentrations as calculated by ICP-MS, indicating that the clustering method can segregate clusters based on differences in ecotoxic elemental compositions. These elements can be mobilized during water percolation or slag dissolution and can provide lasting environmental concerns. Through entrapment of these elements in carbonates, Hamilton and coworkers reported that up to 102, 440, and 784 ppm of Cr, Ni, and Co, respectively can be captured within hydromagnesite and pyroaurite that precipitate during weathering of ultramafic tailings.[15] These elements were incorporated within carbonates mostly due to the substitution of Mg in the crystalline structure of the carbonates or as distinct grains within carbonate cement. The concentrations of ecotoxic metals that were available within the ore minerals were below the detection limit in the pit water, suggesting that the CO_2 mineralization process can sequester ecotoxic elements.[15] In our study, however, no grains containing ecotoxic elements were detected in the precipitated calcite, and low concentrations of Cr (maximum of 9 ppm) and V (maximum of 84 ppm) were measured in the precipitated calcite as depicted in Figure 7, although LA-ICP-MS shows that the cluster in the slag that is rich in V and Cr contains

up to 9000 and 110 000 ppm of V and Cr. In the present analysis, the studied elements have ionic radii smaller than that of Ca^{2+} and should be suitable for incorporation within the precipitated calcite as they can fit into the lattice site without disrupting the calcite lattice.[47] Nevertheless, Mn is observed to exist at a high concentration within the calcite that precipitates within the pore space. We attribute this to the high incorporation of Mn that can result from the slow precipitation of calcite that occurs in porous media, as reported elsewhere.[47] While the current results show that carbonate precipitation can immobilize some ecotoxic elements, previous work and modeling studies suggested that carbonate precipitation may affect the solubility and concentrations of V and Cr in the leachates.[17,48,49] For example, it has been explained that calcium vanadate $Ca_3(VO_4)_2$ solubility controls the release of V in a slag environment.[16,17] To illustrate, in the absence of CO_2, the released Ca^{2+} during the leaching process is at a concentration sufficient to hinder the dissolution of $Ca_3(VO_4)_2$ (log K_{sp} = −17.97). However, the presence of CO_3^{2-} results in the formation of $CaCO_3$, which provides a sink for Ca^{2+} and promotes the dissolution of $Ca_3(VO_4)_2$. This indicates that larger CO_2 mineralization can result in a higher release of V. With respect to Cr, no mineral in their thermodynamic database could mimic its leaching as a function of pH, and at various carbonation levels.[16] However, it was reported that higher Cr concentration in the leachate was observed at higher carbonation and that its leaching is proportional to basicity reduction.[50]

4. Conclusion

As slag contains alkaline earth oxides in the form of silicate minerals (e.g., åkermanite), it can be used in CO_2 removal applications. However, Ecotoxic elements in slag can be a source of pollution that may present a drawback when deploying passive mineralization as a CO_2 removal solution. While the literature largely focuses on the bulk content of these elements in slag, this work focuses on interrogating slag samples at the microstructural level through a machine learning-based approach. This approach is based on using dimensionality reduction followed by the HDB-SCAN method to segment EDS data into distinct clusters, each with its own chemical signature. In addition to providing clear evidence of calcite precipitation in the studied samples, the provided workflow can segment the data into clearly identifiable clusters that are otherwise easily missed through visual inspection of BSE and EDS data. For example, slag can contain ecotoxic elements (V/Cr) in grains having equivalent diameters of less than 42 μm. While it is common to pulverize slag to increase the surface area, thereby enhancing the reaction rate, our work demonstrates that size reduction to particles smaller than 42 μm should be avoided to prevent the mobilization of these elements and their consequent delivery to nearby water resources. Additionally, the presented analysis shows the content of ecotoxic elements in the precipitated calcite (mainly V, averaging at 55 ppm), suggesting that calcite precipitation may immobilize ecotoxic elements that are released due to slag weathering. While dissolution of slag minerals is required for calcite precipitation, it also makes ecotoxic elements more prone to mobilization. This problem should be accounted for during the economic and life cycle evaluation of enhanced weathering of slag. The presented workflow shows that the clustering algorithm can segregate different portions of
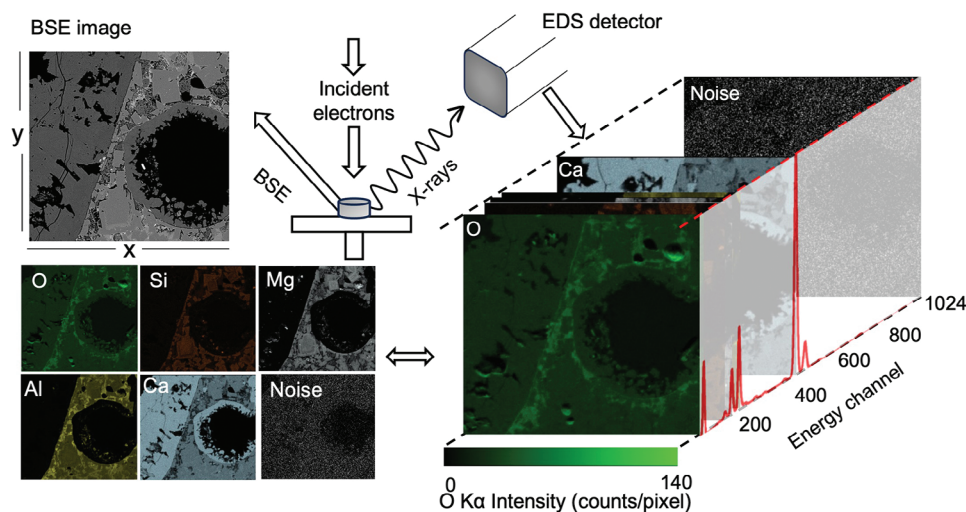
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

Figure 8. Demonstration of the data structure obtained through BSE/EDS. The count of each energy level at each pixel is recorded in the EDS data, as exemplified in the case of Oxygen (green map). The edge length of the BSE image is 500 μm.

a thin section into different clusters that can otherwise be easily missed during visual inspection of the microscopy data, and the locations of the observed clusters matched that displayed from the LA-ICP-MS. The provided workflow relies on freely available, open source thereby making it accessible to a wider community. However, given that this workflow is applied only on a small number of samples (n = 3) in the current work, future work can involve applying it to a larger number of samples to increase the representativeness of the thus obtained results. Additionally, the availability of an open-source EDS quantification algorithm can enhance the accessibility of the provided workflow and will enable studying the effect of using different machine learning parameters/algorithms on the reproducibility of clusters quantification.

5. Experimental Section

SEM-EDS Data Acquisition System: Slag samples were collected from the former Ravenscraig Steelworks site in Motherwell, Scotland. Details about the location of this site and its history were summarized in the previous study. In light of the previous observations of microstructural features of slag and previously acquired backscattered electron images of slag,[31,36,32] three samples (A, B, and C) were shortlisted for the present analysis. These samples were selected so that they reflect the variation in slag sample microstructure (e.g., presence of pores space, presence of materials with different atomic numbers, presence of pore filling materials and precipitates) and so they demonstrate different calcite precipitation patterns (i.e., in pore space and on the external surface of the slag). This enables understand how different precipitation environments may immobilize ecotoxic elements. The samples were mounted, and three thin sections (one per sample) were prepared and imaged using the Zeiss Axio Imager M2m microscope (Supplementary Figure S1, Supporting Information), polished, and carbon-coated for imaging with the SEM. Backscatter electrons (BSE) and EDS signal were collected using Zeiss Sigma VP Field-Emission Gun SEM (VP-FEG-SEM) equipped with Oxford Instruments 170 mm^2 silicon drift EDS detector based in the GEMS facility at the University of Glasgow. For each sample, the acquired data tiles were stitched and montaged with Oxford Instrument AZTec 6.0 software. Previous work demonstrated that ≈700 counts per pixel was sufficient to facilitate data

analysis.[40] However, there is no direct entry for the number of counts per pixel in SEM-EDS control software. The counts per pixel can be increased through increasing the dwell time and/or the number of acquired frames. Consequently, several attempts were performed to find suitable combinations of settings that yield greater than 1000 counts per pixel without introducing surface damage and minimal drift, in a reasonable time (≈ 5 h). The acquisition parameters are summarized in Table 1.

Dimensionality Reduction and Clustering: The obtained EDS map data were exported as .raw and .rpl files from Aztec 6.0 software and were imported and separately processed in Jupyter Notebook using HyperSpy, SciPy, pandas, Matplotlib, Seaborn, and NumPy dependencies.[51–56] The current workflow builds on the work of Tominaga and coworkers[40] for processing of the EDS data and it is described as follows. The obtained BSE images were used to visualize the samples and to identify pore space such that it could be masked from subsequent dimensionality reduction and clustering analysis. Pore space provides no information regarding the sample mineralogy and elements' distribution and can be ignored in dimensionality analysis to reduce the size of the processed data, thereby reducing the processing time. The pore space mask of a given sample was created by manually thresholding the corresponding BSE image through the Scikit image library in Python.[57] The Poissonian noise in the EDS data was normalized following the method of Keenan and Katoula[29] from the modified HyperSpy source code. **Figure 8** depicts the structure of the collected data. Each EDS map is divided into pixels in the x and y direction, and the pixel size is dictated by the required resolution and magnification. The EDS detector measures the energy emitted from each pixel as a result of the interaction with the incident electrons. The energy ranges across 2048 levels that span over 20 keV. As energy levels are characteristic of elements, more than one energy level can be associated with each pixel, depending on the chemical composition at that pixel, as exemplified in O, Si, Mg, Al, and Ca maps shown in Figure 8. Therefore, the obtained data are structured as follows: x and y values that identify the location of the pixel in the sample, and values for each of the 2048 energy levels. As the map data contain over 400000 pixels and 2048 energy channels, they are characterized by high dimensionality. Inspection of the acquired EDS data demonstrates that they show no peak beyond 1024 energy channels, so the energy axis was cropped beyond this point. The dimensionality reduction is then required to eliminate noise by identifying linear relationships in the data that describe most of the relevant chemical variation, later used to cluster the data. Dimensionality reduction reduces the computational cost and improves the performance of clustering algorithms that utilize distance-based metrics.[28,57–59] Consequently, Principal Component Analysis (PCA) was used to project the EDS data to lower dimen-

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
SUSTAINABLE
SYSTEMS**

www.advsustainsys.com

sional space using PCA singular value decomposition (SVD)-based implementation in Scikit-learn.[60] Reduced-dimensionality data were then clustered using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm.[61,62] This probabilistic method does not require prior knowledge of the number or geometry of the clusters within a dataset, and it does not require all points within a dataset to be clustered as it can identify noise as outliers. Here HDBSCAN implementation was used in Python to cluster the obtained EDS data into different phases, each with a distinctive chemical signature, to visualize the distribution of different elements within the studied samples.[63] The HDBSCAN hard cluster assignments were used to identify regions of interest for further analysis and quantification using the standardless, peak-to-background method (P/B-ZAF) implemented within Bruker ESPRIT software. This EDS quantification technique yields the mass and atomic ratios of different elements in a given spectrum, thereby facilitating mineral identification and sample characterization. User inputs to the provided workflow are required for peaks identification and spectra calibration, choosing a number of the dimensions for the dimensionally reduced data, and choosing values for the HDBSCAN parameters (e.g., minimum cluster size and minimum sample numbers). The user can also use a mask to exclude a given area (like pores) from the analysis. The analysis was performed on a Dell Precision 5820 workstation with Intel Xeon W-2175 CPU and 32 GB RAM. Running the provided workflow on the largest dataset (Sample B) required less than 10 min, as measured using the time.time() function. To calculate the morphological properties of a given cluster, a binary image of that cluster was noise-reduced and then analyzed using the morphology and label measurement operation in the Scikit image library in Python.[57] Details of used packages and Python scripts are available in the supplementary information.

This open-source unsupervised approach provides a vendor-independent and agnostic method for the identification and quantitation of the mineral phases present in a specimen. Commercial solutions for automated mineralogy such as the QEMSCAN, Zeiss's Mineralogic or Tescan's TIMA all rely on the coupling of vendor-specific instrument automation combined with a proprietary database of reference EDS spectra (more information about these systems are available elsewhere.[63–66] Ignoring vendor-specific variations, the crux of the technique rests on matching the collected spectra to a library of reference spectra – in some ways very similar to simple forms of supervised regression. Importantly the result of this fitting process results in a set of proprietary data formats, which are difficult to work with and require vendor licenses. In contrast, the workflow leverages the power of the underlying patterns in the data collected. This exploratory data analysis approach allows for the discovery of minor phases which would often get overlooked due to the weak signals being not apparent due to low abundance. In this way, user bias was avoided in the assignment of phases, while leveraging experimental expertise in the processing and classification of the phases identified. Further by leveraging open-source software tools, these are not dependent on any one vendor's system design and instead offer a method for a broad range of users to take advantage of the statistically dense data that they routinely collect with their existing instrumentation. Additionally, the workflow can be adapted to address system-specific peculiarities. Finally, in the present work, the output of the workflow was assessed by using Raman spectroscopy to ensure that different regions within a given cluster do represent the same material and that the compositions calculated through the EDS quantification technique in a given cluster match the mineral identification obtained through Raman spectroscopy.

Raman Spectroscopy and LA-ICP-MS: For mineral identification in the three studied samples, Raman spectroscopy was performed using an In-Via Raman microscope (Renishaw) within the GEMS facility at, University of Glasgow. 514 nm monochromatic light and 1000 grating were used to identify the intensity of Raman shifts to be used for mineral identification. Data were processed by truncation and background subtraction using WiRE 4.1 software (Renishaw) and they were indexed with reference to the RRuff database.[67] ASI RESOlution Laser ablation system (Australian Scientific Instruments) with ArF 193 nm laser (ATL Lasertechnik) and Thermo iCAP RQ mass spectrometer (Thermo Fisher Scientific) was used for quantification of elements within areas of interest in slag in Sample B that were

identified through clustering and within the precipitated carbonates of all the three samples. Line scans were chosen over stationary spots since the chosen spot size of 38 μm is larger than the features of interest that were identified from the clustering analysis. Analysis of line scans enables calculations of elements concentrations along a given line to good accuracy and precision of <5%, as shown elsewhere.[68] At the condition of the LA-ICP-MS, the depth of the ablated material is ≈5 μm. Consequently, the thus produced concentration profiles can be compared with the quantifications of EDS. Areas that resemble 38 μm-wide lines were ablated 5 Hz laser having energy of 3.8–3.9 mJ. The following signals were collected: ^{29}Si, ^{44}Ca, ^{48}Ti, ^{51}V, ^{52}Cr, ^{55}Mn, ^{59}Co and ^{60}Ni. ^{44}Ca was collected as it was used as the internal standard for quantification of trace elements in calcite, while ^{51}V, ^{52}Cr, ^{55}Mn, ^{59}Co, and ^{60}Ni were collected as they are anticipated to be present in calcite due to their association with leachates produced during slag weathering.[69] For the analysis of the slag, the NIST SRM 610 standard was used in accordance with the literature since no better reference material that resembles slag properties has been identified.[70] Due to the heterogeneous nature of the studied sample, only semi-quantitative data are reported as no internal standard can be found at a fixed composition throughout the scanned area. For analysis of calcite, the MACS 3 standard was used, with calcium being identified as the internal standard. Quality assessment of the collected data and the uncertainty of the measurement are demonstrated in Tables S1 and S2 (Supporting Information) respectively. The acquired LA-ICP-MS data were analyzed with Iolite4 software with reference to the GeoReM database.[71]

Supporting Information

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

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

F.W.K.K: conceptualization, methodology, data collection/analysis, and Wrote-original draft. M.D: performed analysis, methodology, wrote review, and edited. M.W: performed data collection and analysis. J.M: Supervised, Funding acquisition, methodology, wrote-review, and edited. J.F.E.: performed supervision, analysis, methodology, wrote-review, and edited.

Data Availability Statement

The data that support the findings of this study are openly available in [Zenodo] at [https://doi.org/10.5281/zenodo.8389122], reference number [72] and on the University of Glasgow repository [https://doi.org/10.5525/gla.researchdata.1584]

[1] World Steel Association, "World Steel in Figures," can be found under https://worldsteel.org/steel-topics/statistics/world-steel-in-figures-2022/, **2023**.

[2] R. An, B. Yu, R. Li, Y.-M. Wei, *Appl. Energy* **2018**, *226*, 862.

[3] World Steel Association, "World Steel in Figures 2019," can be found under https://worldsteel.org/wp-content/uploads/2019-World-Steel-in-Figures.pdf, **2019**.

[4] L. Li, T.-C. Ling, S.-Y. Pan, *Sci. Total Environ.* **2022**, *806*, 150280.

[5] F. W. K. Khudhur, J. M. MacDonald, A. Macente, L. Daly, *Sci. Total Environ.* **2022**, *823*, 153553.

[6] W. M. Mayes, A. L. Riley, H. I. Gomes, P. Brabham, J. Hamlyn, H. Pullin, P. Renforth, *Environ. Sci. Technol.* **2018**, *52*, 7892.

[7] G. P. Assima, F. Larachi, G. Beaudoin, J. Molson, *Ind. Eng. Chem. Res.* **2012**, *51*, 8726.

[8] A. Polettini, R. Pomi, A. Stramazzo, *J. Environ. Manag.* **2016**, *167*, 185.

[9] K. S. Lackner, C. H. Wendt, D. P. Butt, E. L. Joyce, D. H. Sharp, *Energy* **1995**, *20*, 1153.

[10] H. Béarat, M. J. McKelvy, A. V. G. Chizmeshya, D. Gormley, R. Nunez, R. W. Carpenter, K. Squires, G. H. Wolf, *Environ. Sci. Technol.* **2006**, *40*, 4802.

[11] H. Pullin, A. W. Bray, I. T. Burke, D. D. Muir, D. J. Sapsford, W. M. Mayes, P. Renforth, *Environ. Sci. Technol.* **2019**, *53*, 9502.

[12] H. I. Gomes, W. M. Mayes, M. Rogerson, D. I. Stewart, I. T. Burked, *J. Clean. Prod.* **2016**, *112*, 3571.

[13] W. M. Mayes, P. L. Younger, J. Aumônier, *Water. Air. Soil Pollut.* **2008**, *195*, 35.

[14] N. M. Piatak, M. B. Parsons, R. R. Seal, *Appl. Geochemistry* **2015**, *57*, 236.

[15] J. L. Hamilton, S. A. Wilson, B. Morgan, C. C. Turvey, D. J. Paterson, S. M. Jowitt, J. McCutcheon, G. Southam, *Int. J. Greenh. Gas Control* **2018**, *71*, 155.

[16] W. J. J. Huijgen, R. N. J. Comans, *Environ. Sci. Technol.* **2006**, *40*, 2790.

[17] A. J. Hobson, D. I. Stewart, A. W. Bray, R. J. G. Mortimer, W. M. Mayes, M. Rogerson, I. T. Burke, *Environ. Sci. Technol.* **2017**, *51*, 7823.

[18] I. Z. Yildirim, M. Prezzi, *Adv. Civ. Eng.* **2011**, *2011*, 463638.

[19] N. Jacquemet, J. Pironon, V. Lagneau, J. Saint-Marc, *Appl. Geochemistry* **2012**, *27*, 782.

[20] E. E. Chang, S. Y. Pan, Y. H. Chen, C. S. Tan, P. C. Chiang, *J. Hazard. Mater.* **2012**, *227–228*, 97.

[21] J. Ryu, H. Kim, R. M. Kim, S. Kim, J. Jo, S. Lee, K. T. Nam, Y. C. Joo, G. C. Yi, J. Lee, M. Kim, *Ultramicroscopy* **2021**, *231*, 113314.

[22] J. Love, J. Stevens, *The Scottish Steel Industry*, University of Edinburgh, Newington, Edinburgh **1989**.

[23] S. Nasrazadani, S. Hassani, *Modern analytical techniques in failure analysis of aerospace, chemical and oil and gas industries. In: Handbook of materials failure analysis with case studies from the oil and gas industry* (Eds: A. S. H. Makhlouf, M. Aliofkhazraei), Butterworth-Heinemann, **2016**, 39.

[24] W. J. Wolfgong, *Chemical analysis techniques for failure analysis: part 1 common instrumental methods. In: Handbook of materials failure analysis with case studies from the aerospace and automotive industry* (Eds: A. S. H. Makhlouf, M. Aliofkhazraei), Butterworth-Heinemann, **2016**, 279.

[25] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. M. Ritchie, J. H. J. Scott, D. C. Joy, *Scanning Electron Microscopy and X-Ray Microanalysis*, Springer, New York, **2017**.

[26] S. Burgess, "Determining accurate detection limits for EDS analysis using simulated spectra – Oxford Instruments," can be found under https://www.oxinst.com/blogs/determining-accurate-detection-limits-for-eds-analysis-using-simulated-spectra, **2022**.

[27] J. Shlens, *Arxiv* **2014**, https://arxiv.org/abs/1404.1100v1

[28] E. C. Duran, Z. Kho, J. F. Einsle, I. Azaceta, S. A. Cavill, A. Kerrigan, V. K. Lazarov, A. S. Eggeman, *Comput. Mater. Sci.* **2023**, *228*, 112336.

[29] M. R. Keenan, P. G. Kotula, *Surf. Interface Anal.* **2004**, *36*, 203.

[30] A. L. Riley, J. M. MacDonald, I. T. Burke, P. Renforth, A. P. Jarvis, K. A. Hudson-Edwards, J. McKie, W. M. Mayes, *J. Geochemical Explor.* **2020**, *219*, 106630.

[31] F. W. K. Khudhur, A. Macente, J. M. MacDonald, L. Daly, *Microsc. Microanal.* **2022**, *28*, 1514.

[32] J. M. MacDonald, F. W. K. Khudhur, R. Carter, B. Plomer, C. Wilson, C. Slaymark, *Appl. Geochemistry* **2023**, *152*, 105649.

[33] D. G. Grubb, D. R. V. Berggren, *J. Hazardous, Toxic, Radioact. Waste* **2018**, *22*, 04018030.

[34] R. I. Iacobescu, Y. Pontikes, D. Koumpouri, G. N. Angelopoulos, *Cem. Concr. Compos.* **2013**, *44*, 1.

[35] J. S. Chukwuma, H. Pullin, P. Renforth, *Miner. Eng.* **2021**, *173*, 107232.

[36] F. W. K. Khudhur, J. M. MacDonald, L. Daly, A. Macente, L. Spruženiece, S. Griffin, C. Wilson, *Micron* **2023**, *174*, 103532.

[37] F. Engström, D. Adolfsson, C. Samuelsson, Å. Sandström, B. Björkman, *Miner. Eng.* **2013**, *41*, 46.

[38] I. M. Power, A. L. Harrison, G. M. Dipple, S. A. Wilson, P. B. Kelemen, M. Hitch, G. Southam, *Rev. Mineral. Geochemistry* **2013**, *77*, 305.

[39] C. Noiriel, C. I. Steefel, L. Yang, J. Ajo-Franklin, *Chem. Geol.* **2012**, *318–319*, 60.

[40] M. Tominaga, E. Ortiz, J. F. Einsle, N. F. Ryoichi Vento, M. O. Schrenk, I. Buisman, I. S. Ezad, D. Cardace, *Geophys. Res. Lett.* **2021**, *48*, e2020GL088472.

[41] L. Kriskova, Y. Pontikes, L. Pandelaers, Ö. Cizer, P. T. Jones, K. Van Balen, B. Blanpain, *Metall. Mater. Trans. B* **2013**, *44*, 1173.

[42] S. N. Lekakh, C. H. Rawlins, D. G. C. Robertson, V. L. Richards, K. D. Peaslee, *Metall. Mater. Trans. B Process Metall. Mater. Process. Sci.* **2008**, *39*, 125.

[43] A. A. Reddy, D. U. Tulyaganov, A. Goel, S. Kapoor, M. J. Pascual, J. M. F. Ferreira, *J. Mater. Sci.* **2013**, *48*, 4128.

[44] C. Crouzet, F. Brunet, G. Montes-Hernandez, N. Recham, N. Findling, J. H. Ferrasse, B. Goffé, *Front. Energy Res.* **2017**, *5*, 2017.00029.

[45] A. Polettini, R. Pomi, A. Stramazzo, *Chem. Eng. J.* **2016**, *298*, 26.

[46] M. Kelvin, E. Whiteman, J. Petrus, M. Leybourne, V. Nkuna, *Miner. Eng.* **2022**, *176*, 107322.

[47] J. D. Rimstidt, A. Balog, J. Webb, *Geochim. Cosmochim. Acta* **1998**, *62*, 1851.

[48] L. De Windt, P. Chaurand, J. Rose, *Waste Manag.* **2011**, *31*, 225.

[49] G. Cornelis, C. A. Johnson, T. Van Gerven, C. Vandecasteele, *Appl. Geochemistry* **2008**, *23*, 955.

[50] R. M. Santos, D. Ling, A. Sarvaramini, M. Guo, J. Elsen, F. Larachi, G. Beaudoin, B. Blanpain, T. Van Gerven, *Chem. Eng. J.* **2012**, *203*, 239.

[51] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, et al., *Nat. Methods* **2020**, *17*, 261.

[52] F. de la Peña, E. Prestat, V. T. Fauske, P. Burdet, J. Lähnemann, P. Jokubauskas, T. Furnival, M. Nord, T. Ostasevicius, K. E. MacArthur, D. N. Johnstone, M. Sarahan, J. Taillon, T. Aarholt, pquinn-dls, V. Migunov, A. Eljarrat, J. Caron, C. Francis, T. Nemoto, T. Poon, S. Mazzucco, actions-user, N. Tappy, N. Cautaerts, S. Somnath, T. Slater, M. Walls, F. Winkler, H. W. Ånes, *Zenodo* **2022**, https://doi.org/10.5281/ZENODO.7263263.

[53] The pandas development team, *Zenodo* **2023**, https://doi.org/10.5281/zenodo.8364959.

[54] J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*, 90.

[55] M. L. Waskom, *J. Open Source Softw.* **2021**, *6*, 3021.

[56] S. Van Der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, *Peer J.* **2014**, *2014*, e453.

[57] T. Landers, R. L. Rosenberg, *Encycl. Database Syst.* **2009**, *6*, 846.

[58] J. Blanco-Portals, F. Peiró, S. Estradé, *Microsc. Microanal.* **2022**, *28*, 109.

[59] C. C. Aggarwal, A. Hinneburg, D. A. Keim, *On the surprising behaviour of distance metrics in high dimensional space. In Database theory-ICDT 2001. ICDT 2001* (Eds: J. Van den Bussche, V. Vianu). Lecture Notes in Computer Science vol 1973. Springer, Berlin, Heidelberg. **2001**, 420.

[60] F. Pedregosa, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot andÉdouarddand, A. Duchesnay, F. D. Edouarddduchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.

[61] R. J. G. B. Campello, D. Moulavi, J. Sander, in *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Springer, Berlin, Heidelberg **2013**, pp. 160.

[62] L. McInnes, J. Healy, S. Astels, *J. Open Source Softw.* **2017**, *2*, 205.

[63] R. Fandrich, Y. Gu, D. Burrows, K. Moeller, *Int. J. Miner. Process.* **2007**, *84*, 310.

[64] P. Gottlieb, G. Wilkie, D. Sutherland, E. Ho-Tun, S. Suthers, K. Perera, B. Jenkins, S. Spencer, A. Butcher, J. Rayner, *JOM* **2000**, *52*, 24.

[65] T. Hrstka, P. Gottlieb, R. Skála, K. Breiter, D. Motl, *J. Geosci.* **2018**, *63*, 47.

[66] R. Juránek, J. Výravský, M. Kolář, D. Motl, P. Zemčík, *Comput. Geosci.* **2022**, *165*, 105109.

[67] B. Lafuente, R. T. Downs, H. Yang, N. Stone, in *Highlights Mineral. Crystallogr,* (Eds.: T. Armbruster, R. M. Danisi, W. De Gruyter), De Gruyter, Berlin, Germany, **2015**, pp. 1–30.

[68] R. Hennekam, T. Jilbert, P. R. D. Mason, G. J. de Lange, G. J. Reichart, *Chem. Geol.* **2015**, *403*, 42.

[69] H. I. Gomes, W. M. Mayes, H. A. Baxter, A. P. Jarvis, I. T. Burke, D. I. Stewart, M. Rogerson, *J. Clean. Prod.* **2018**, *202*, 401.

[70] E. Bérard, C. Pécheyran, P. Dillmann, S. Leroy, E. Vega, A. Williams, C. Verna, V. Toureille, *J. Anal. At. Spectrom.* **2020**, *35*, 2582.

[71] K. P. Jochum, U. Nohl, K. Herwig, E. Lammel, B. Stoll, A. W. Hofmann, *Geostand. Geoanal. Res.* **2005**, *29*, 333.

[72] F. W. K. Khudhur, M. Divers, J. M. MacDonald, J. F. Einsle, M. Wildman, *Zenodo* **2023**, https://doi.org/10.5281/zenodo.8389122.