TECHNICAL ARTICLE

Multivariate statistical analysis of micro-X-ray fluorescence spectral images

Mark A. Rodriguez,[a] Paul G. Kotula, James J. M. Griego, Jason E. Heath, Stephen J. Bauer, and Daniel E. Wesolowski
Sandia National Laboratories, Albuquerque, New Mexico 87185-1411

Multivariate statistical analysis (MSA) is applied to the extraction of chemically relevant signals acquired with a micro-X-ray fluorescence (μ-XRF) mapping (full-spectral imaging) system. The separation of components into individual histograms enables separation of overlapping peaks, which is useful in qualitatively determining the presence of chemical species that have overlapping emission lines, and holds potential for quantitative analysis of constituent phases via these same histograms. The usefulness of MSA for μ-XRF analysis is demonstrated by application to a geological rock core obtained from a subsurface compressed air energy storage (CAES) site. Coupling of the μ-XRF results to those of quantitative powder X-ray diffraction analysis enables improved detection of trace phases present in the geological specimen. The MSA indicates that the spatial distribution of pyrite, a potentially reactive phase by oxidation, has low concentration and thus minimal impact on CAES operations. © *International Centre for Diffraction Data*
[doi:10.1017/S0885715612000243]

I. INTRODUCTION

Compressed air energy storage (CAES) is an important technology for renewable energy systems such as wind and solar power, because of the inherently intermittent nature of power-generation of these systems. To cope with the intermittent nature, it is ideal to store power that is generated "off-peak" in a temporary energy storage facility, until peak demand requires its release. CAES systems are designed to store energy as pressurized air, pumped into large underground caverns or porous rock formations such as saline aquifers (Succar and Williams, 2008). When stored energy is required, the pressure is released to power electrical generators that restore electricity to the power grid. The rock formations for CAES require detailed characterization regarding their geological constituents to predict and optimize performance. A concern is the possible reaction of gases present in the compressed air, such as oxygen, with phases present in the formation. Pyrite oxidation could alter the pH, salinity, and mineralogy. Oxidation could potentially produce colloidal products, such as ferric hydroxide and melanterite, and dissolution of carbonates with associated precipitation of gypsum (Succar and Williams, 2008). Thus, oxidation can substantially impact CAES through alteration of porosity and permeability, which in turn govern storage capacity and flow rates of air into and out of the formation. The use of X-ray diffraction (XRD) and micro-X-ray fluorescence (μ-XRF) in the analysis of one such core-drilled specimen taken from a borehole at a CAES facility is discussed. The use of principal component analysis (PCA), a common form of multivariate statistical analysis (MSA), to decompose the convoluted μ-XRF dataset containing elemental information regarding at

least eight distinct phases is also discussed. These XRF results are then used to improve trace phase identification of the quantitative XRD results.

II. EXPERIMENTAL

The rock core was obtained from a borehole to support site characterization of a potential CAES facility in Iowa. A small piece ($\sim$24 × $\sim$40 × $\sim$15 mm^3) was cut from the original core block. From that piece, a small portion was removed and ground to a powder ($\sim$5 g) for routine powder XRD analysis. The remaining portion of sample was impregnated with a red fluorescent, low-viscosity resin under vacuum. The sample was then mounted on a glass slide and polished to a thickness of $\sim$30 μm. This thin-section specimen was analyzed using optical petrography (using a Leitz Wetzlar Orthoplan-Pol polarizing microscope) as well as subsequent μ-XRF analysis.

A Bruker M4 Tornado μ-XRF mapping system was used for μ-XRF analysis. The instrument was equipped with a micro-focused Rh source (50 kV, 600 μA) with a polycapillary optic ($\sim$30 μm spot size). The detector system used two silicon-drift detectors to collect fluorescence spectra from the specimen. The specimen was secured to the x–y translation stage within the M4 chamber, and XRF spectra were collected under vacuum conditions ($\sim$10^{-3} Torr). The XRF-mapping dataset for the thin-section specimen was collected as a large datacube with full X-ray spectra (4096 channels, 0–40 keV) collected at each pixel in a two-dimensional array. The step size used for the μ-XRF spatial map was 50 μm. This resulted in a 502 × 860 matrix for the map which covered an area of $\sim$25 × 43 mm^2, thus incorporating the entire cross-sectioned specimen. Total data collection time was $\sim$4 h. The dimensionally of the datacube (502 × 860 × 4096) was greater than 1.7 billion elements, and encompassed a file size of 747 MB. Elemental maps for individual atomic

---

species were generated within the M4 software package. For PCA, the datacube was read into an in-house modified PCA software package written in MATLAB (The Mathworks, 2008). MSA, which seeks to maximize the mutual simplicity of the spatial components (Keenan, 2009), took 90 s on a PC equipped with four-core 3.8 GHz Intel Xeon processors and 8 Gb of memory. Full details on the use of PCA are beyond the scope of this paper and the reader may refer to literature on this topic (e.g. Jollife, 2002; Keenan and Kotula, 2004; Rodriguez, *et al.*, 2007, 2010).

For XRD analysis, the ground powder specimen (<100 μm particle size) was loaded into a side-drifted specimen holder against filter paper (to reduce preferred orientation effects). Standard powder XRD analysis was performed using a Siemens model D500 $\theta - \theta$ diffractometer equipped with a sealed-tube (Cu$K\alpha$) source, fixed (1°) slits, a diffracted-beam graphite monochromator, and a scintillation detector. Generator settings were 40 kV and 30 mA. Scan parameters were a 2θ angular range of 10–65°, a step size of 0.04°, and a count time of 4 s. Phase identification was performed within the program Jade (ver. 9.3, Materials Data, Inc., Livermore, CA). The resulting XRD dataset was quantitatively analyzed using Rietveld refinement within the program GSAS (Larson and Von Dreele, 2000; Toby, 2001).

III. RESULTS AND DISCUSSION

Figure 1 reveals that the specimen is almost entirely composed of the mineral quartz (PDF entry 00-046-1045; also see Table I). An initial single-phase refinement of the powder XRD data within GSAS revealed a reasonable fit, as shown in the inset labeled "Quartz only" (upper right corner of Figure 1). However, careful evaluation of the low-intensity peaks near the background (see zoomed region, inset in Figure 1) reveals small peaks from additional phases. Detailed evaluation of the low-intensity "trace" phases was desirable, as initial optical petrography analysis (not shown) suggested additional phases besides the dominant quartz. In particular, the presence of pyrite was confirmed by optical microscopy. Note that petrographic microscopy techniques such as point-counting (Dickinson, 1970; Gazzi, 1966) can require hours of detailed measurements. PCA-augmented μ-XRF analysis generates, in minutes, the same spatially distributed chemical information as obtained through these time-consuming microscopy methods. Interestingly, powder XRD analysis, used specifically to identify these trace phases, could not confirm the presence of pyrite within the sample on the basis of the obtained pattern, as shown in Figure 1. Hence μ-XRF mapping was undertaken, that is, to test whether the chemical signature for the suspected pyrite could be isolated spatially using μ-XRF.

Typical PCA uses algorithms that search large datasets for commonalities within the dimensionally of the data matrix. The resulting PCA outputs are displayed in generic terms as "loadings" and their corresponding "scores" (see Rodriguez, *et al.*, 2007). For the case of this datacube, the loadings (also called "components") are extracted as isolated histograms, along with the corresponding scores that can be extracted as *x*–*y* spatial maps. The typical PCA map output for a component is a (rainbow) color contour image (red = strong, blue = weak) to indicate how that component is distributed spatially. The histograms derived from PCA look like typical XRF spectra and can be treated as such. The combination of component histograms (abbreviated as "Cmpt-#" throughout the text) and their corresponding contour maps can adequately be recombined to fully describe the measured dataset, while removing background noise. Analysis of the μ-XRF datacube using PCA yielded as many as 13 possible
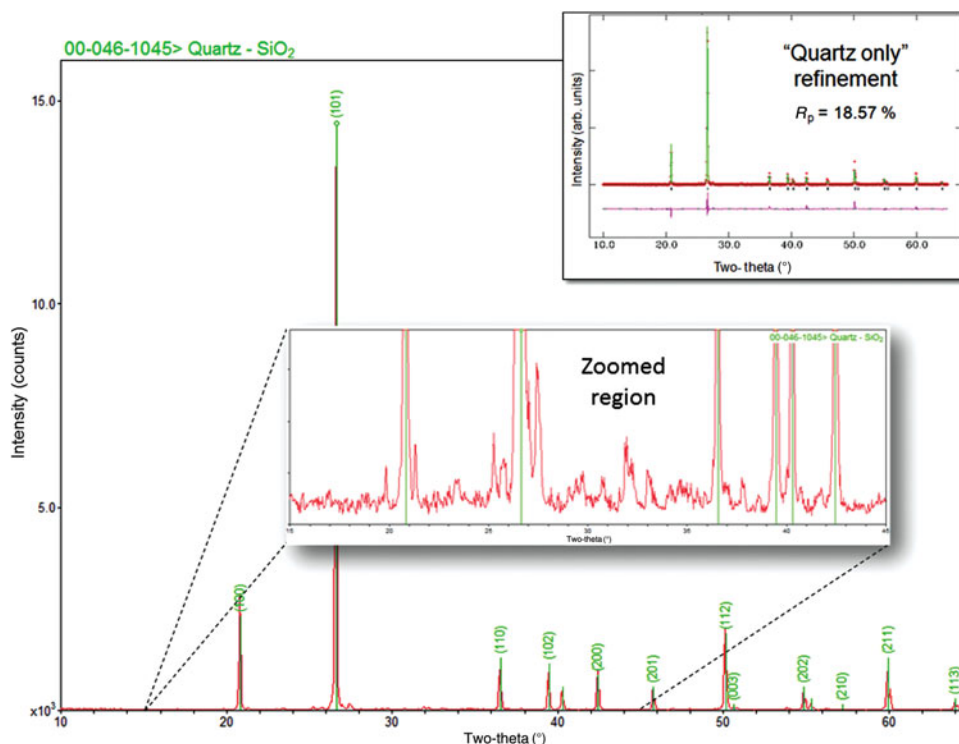

Figure 1. Initial XRD pattern showing domination of quartz content, with only trace peaks of additional phases present in the background.

components necessary to describe the dataset. Some had to do with specific chemical signatures that could easily be assigned to phases within the sample. Others were more subtle effects. Table I outlines all the assigned components from PCA as well as the quantified XRD results from powder XRD analysis. The comparative results of XRD and μ-XRF analyses will be discussed in detail later.

Spectra and PCA maps for all 13 components are not presented. Instead, a few important components are selected to illustrate the functionality of the PCA technique. To orient the reader to PCA output, Figure 2 shows the first derived component (Cmpt-1), along with other images for comparison. The histogram in Figure 2, shown as Intensity vs. Energy, is a straightforward XRF spectrum with the Si$K\alpha$ peak as the main observed emission line, as well as some small peaks from other elements near the background level. This histogram can easily be assigned to the quartz phase. As the μ-XRF system is unable to detect elements below sodium, the only peak expected to be observed from the quartz phase is the Si$K\alpha$ line. The contour map for the first component, shown in the upper right corner of Figure 2 and labeled as "Cmpt-1 PCA map", shows this Si signal coming from nearly all the locations across the specimen. Only small regions indicate any absence of this signal. For comparison, an optical image (labeled "Video image") is included in Figure 2 so one can observe the appearance of the specimen under typical optical conditions. Pink-staining is seen within some areas of the specimen and around the edges; this is from the added epoxy resin. Regions of darker coloring suggest the existence of different phases besides the dominant quartz. In addition to the video image and the PCA map, the Si map generated by the M4 software package is also included, which simply isolates signal from the Si$K\alpha$ emission line and plots its relative magnitude as a function of sample position. This is dubbed the "keV-slice" as it is an energy cross-section through the datacube. The Si$K\alpha$ keV-slice map is similar to the PCA map, indicating the strong presence of Si across the specimen, with small, isolated regions indicating absence of Si-containing minerals. Figures 1 and 2 clearly indicate that any additional phases present in this sample are minor or trace constituents, making this sample particularly challenging with regard to phase identification.

Figure 3 illustrates the fourth derived component of the analysis. The solid line in the Intensity vs. Energy plot (labeled Cmpt-4) shows the spectrum that is very similar to that obtained from an independent pyrite mineral sample, shown as the offset (dashed line) in the plot. The intensity ratios for the S$K\alpha$, Fe$K\alpha$, and Fe$K\beta$ lines are all nearly identical between the two spectra. In essence, the derived component histogram Cmpt-4 is the chemical spectrum for pyrite. The signature is *quantitative*, and can be used to characterize phase presence on the basis of species concentration. Some additional elemental lines are present in the component spectrum, namely the Si$K\alpha$ and the Ti$K\alpha$ peaks. These peaks are not likely part of pyrite, but their origin is likely "bleed-in" of the first component (silicon from quartz) and a second identified component (titanium from anatase, TiO_2; see Table I). The presence of the extra Si$K\alpha$ and the Ti$K\alpha$ peaks shows that the PCA technique is not always perfect for the generation of pure components, but that it sometimes suffers from "component mixing." This is especially true when a datacube contains one or two components that dominate (e.g. quartz), whereas other components are indicative of more subtle effects (i.e. pyrite). In fact, nearly all the 13 derived components had some bleed-in from the quartz component as indicated by a small Si peak in each spectrum. This can make quantification of compounds containing silicates more difficult. However, this issue had little impact on the relative intensity ratios of the S$K\alpha$ and Fe$K\alpha$, $K\beta$ peaks observed in Cmpt-4.

The keV-slice maps shown in Figure 3 for S and Fe are very interesting. They actually do not show a very high degree of correlation. A long, vertical vein is seen on the right side of the S keV-slice map, suggesting a sulfur-containing mineral. This vein is not present in the Fe keV-slice map. Careful analysis of other keV-slice maps revealed that the Ca map (not shown) also displayed this vein. The material was ultimately identified as being from a calcium sulfate mineral gypsum. This was confirmed on the basis of weak peaks in the XRD pattern and the presence of other similar spatial mapping features from other borehole specimens obtained from nearby depths. Note that PCA *did* find a component for gypsum (Cmpt-5) with proper chemical ratios of S and Ca (see Table I). In truth, many of the features in the S and Fe keV-slice maps are not actually signals from pyrite. This

Table I. Summary of X-ray diffraction (XRD) and micro-X-ray fluorescence (μ-XRF) analyses.

PCA rank no. from μ-XRF data	Key elements of component from μ-XRF	Identified material	Wt% from Rietveld refinement[a]	PDF[b] entry number
1	Si	Quartz	94.3(4)	00-046-1045
2	Ti	Anatase	0.5(1)	00-021-1272
3	Ca, Cl	Epoxy resin	–	
4	Fe, S	Pyrite	0.2	00-042-1340
5	Ca, S	Gypsum	0.3	00-033-0311
6	Ca, P	Hydroxylapatite	1.6(2)	04-011-6221
7	Cl	Epoxy in voids	–	
8	K, Al, Si, Fe	Orthoclase	1.9(2)	04-009-3700
8		Microcline	1.0(2)	04-008-1783
9	Si	Si Laue	–	
10	Ca, Si, As	Glass slide	–	
11	Zr, Hf	Zircon	0.2	00-006-0266
12	Si	Si Laue	–	
13	Si	Si Laue	–	

[a]Bracketed value refers to 1σ error on last significant digit.
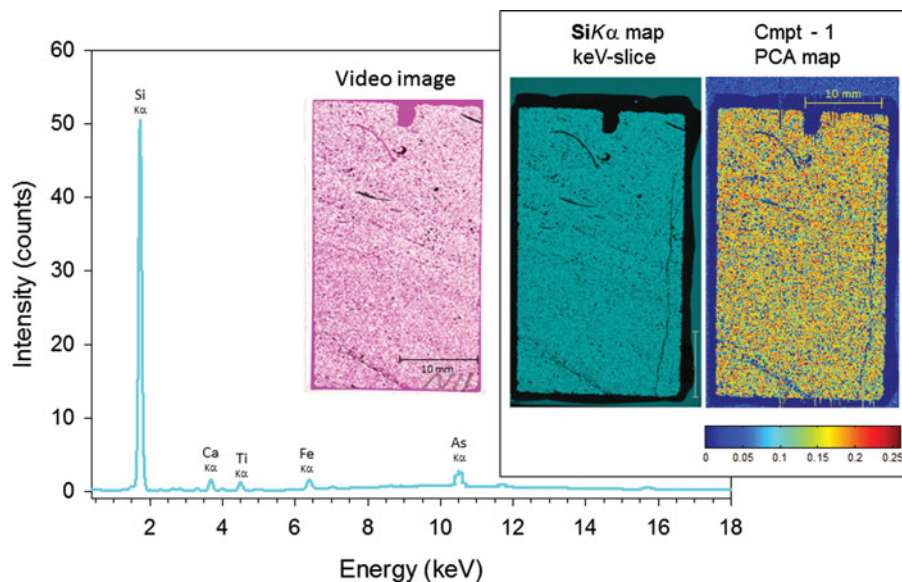[b]Powder Diffraction File (PDF-4 +, 2010): ICDD, Newtown Square, PA.

Figure 2. The Intensity vs. Energy spectrum for the first isolated PCA component (Cmpt-1), along with the corresponding PCA map for Cmpt-1 (shown as a color contour map; upper right). A chemical map for the Si$K\alpha$ keV-slice is shown for comparison, along with a video image from optical microscopy.

fact underscores the difficulty of isolating chemical signatures of trace phases when the atomic species of that desired constituent phase are also present in other phases in the specimen. What one really needs is to spatially identify the commonality of location for the Fe and S fluorescence signals that possess the proper emission-line intensity-ratio to be assigned to the FeS_2 chemical composition. This, in fact, is exactly what the Cmpt-4 PCA map reveals, as seen in Figure 3 (upper right). This color contour map is nearly featureless, with a blue background level across most of the map area. However, dispersed across the map are small regions, shown as red dots, which indicate strong "scores" for the Cmpt-4 histogram. These

small dots indicate the presence of tiny pyrite grains (100–500 μm) within the specimen matrix. One such dot is highlighted with an arrow, shown in the upper right portion of the Cmpt-4 PCA map. Arrows also highlight this same position on the Fe and S keV-slice maps; this demonstrates that S and Fe signals are observed in this location, with approximately the same spatial dimensions as that of the PCA map. Such fine-tuned analysis would be very difficult without the use of PCA techniques. The obvious low quantity of pyrite in this specimen explains the difficulty in detecting this trace phase using conventional powder XRD analysis. In addition, the information derived from the Cmpt-4 PCA map regarding
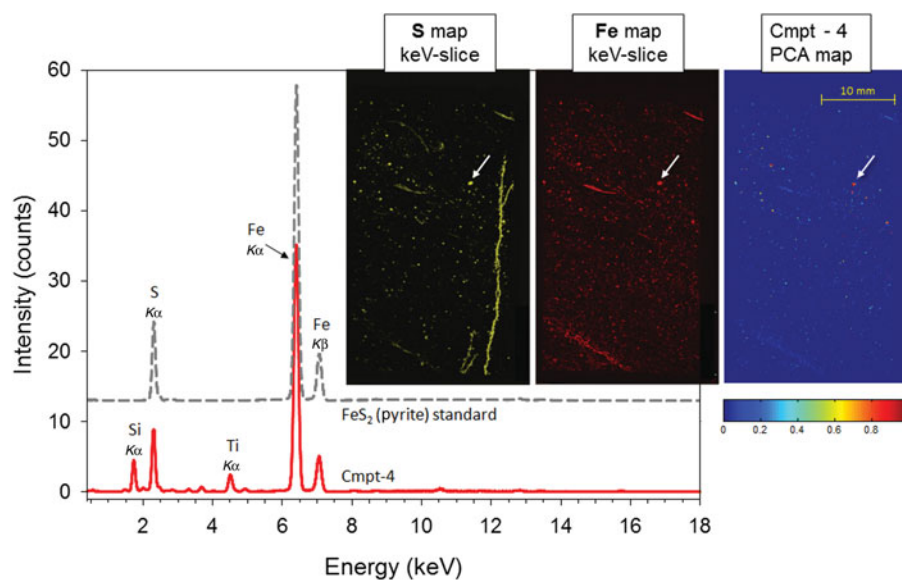

Figure 3. PCA-derived Cmpt-4. The Intensity vs. Energy plot shows the obtained Cmpt-4 spectrum (solid line) along with a spectrum from a pyrite standard (offset dashed line). keV-slice maps for S and Fe are shown along with the PCA-derived contour map for Cmpt-4. A small arrow on map images indicates the location of a ~500-μm pyrite grain.
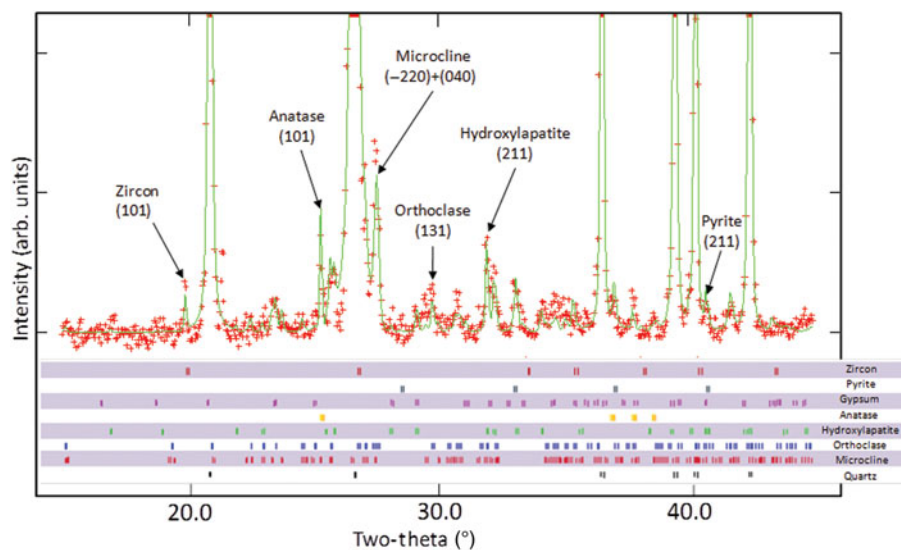
Figure 4. Zoomed range of final Rietveld refinement showing fitting of trace phases. Quartz peaks go off-scale. Reflection tick-marks underneath the fit pattern indicate locations of possible peaks from individual phases. Selected phases with assigned (*hkl*) values have been labeled for the trace phases (R_p = 17.16%).

the small size, and spatially isolated nature of the pyrite grains, increased confidence that the pyrite phase would not pose a major concern regarding acidic damage within the bore-hole. The spatial mapping by μ-XRF was critical in the diagnosis of pyrite as isolated grains, and this information was useful in assessing a minimal impact of pyrite reaction during CAES operation.

Figure 4 shows the final fit of the Rietveld refinement for the powder XRD data. The plot contains the same zoomed-in range as shown in the inset of Figure 1. The small impurity peaks are now accounted for in the refinement shown in Figure 4. The μ-XRF analysis coupled with PCA enabled straightforward clarification of the trace phases. Table I summarizes the synergistic results obtained by the merger of XRD and μ-XRF analyses. Along with quartz, additional phases were confirmed to be present: orthoclase, hydroxylapatite, microcline, anatase, gypsum, zircon, and pyrite. Note that the order in terms of PCA ranking of the derived components listed in Table I does not always follow the quantity of a given phase (i.e. wt%). This is because PCA rank depends on dataset variance and not on concentration. Figure 4 shows specific indexed peaks for these trace phases. These labeled peaks were selected because they had significant relative intensities for the given phase (>40% I_{rel} as documented in PDF entries for these phases listed in Table I), and they did not overlap major peaks from the other phases. The residual error (R_p) obtained from Rietveld analysis reduced from 18.57% for the "Quartz only" refinement (Figure 1) to 17.16% after fitting the trace phases (Figure 4). Although this is not a large change, the reduction of the residual error by ∼1.5%, along with the generally good fit of the observed to calculated patterns in Figure 4, indicates an accurate assessment of the trace phases in this specimen.

The PCA components were of great assistance in the phase identification. Specifically, isolation of Cmpt-6 as containing both phosphorous and calcium enabled the quick identification of hydroxylapatite likely from fossilized remains within the specimen. Microcline and orthoclase have the same chemical signature, and therefore were extracted as one

component (Cmpt-8) in the PCA technique. Phase identification searches of the XRD pattern with K, Al, and Si resulted in the identification of both these phases. The same process worked for the trace phases of zircon and pyrite. In fact, the quantitative results for zircon and pyrite trace phases using Rietveld refinement indicate their presence at ∼0.2 wt%. This is typically at the limit of quantification for powder diffraction measurements, as the error associated with the phase fraction is of the same magnitude as the quantity refined to be present. Even with this limitation in quantification, qualitatively, the phases can be isolated spatially in terms of chemical signature, and confirmed present on the basis of the location of diffraction peaks in the bulk powder analysis. This was critical in the confirmation of pyrite within this specimen.

Notably, a defining (*hkl*) for the gypsum phase is absent from Figure 4. Significant peak overlap occurred for all the major reflections in this phase. The pattern showed consistency regarding possible intensity, which could account for this phase, but the intensity always overlapped other major reflections from other phases. Gypsum was confirmed on the basis of several factors. Firstly, the μ-XRF clearly showed a component with Ca and S in a 1:1 atomic ratio (see Cmpt-5 in Table I), which suggested a sulfate. Secondly, other core-drill specimens taken from similar depths showed gypsum, but with higher concentrations, which made the identification routine. Thirdly, other calcium sulfate-containing phases common in geological formations, such as anhydrite and bassanite, were inconsistent with the XRD data. Lastly, gypsum was accommodated within the Rietveld refinement without difficulty and refined to a realistic concentration.

In Table I, note again that not all PCA components had origins in specific phases present in the core-drill cross-section. Actually, 6 of the 13 components could be considered artifacts of the analysis. This includes components from the epoxy resin (i.e. Cmpt-3 and Cmpt-7), a component from the underlying glass slide (Cmpt-10), and components from Si Laue peaks (Cmpt-9, Cmpt-12, and Cmpt-13). These artifacts are very interesting because they demonstrate the ability
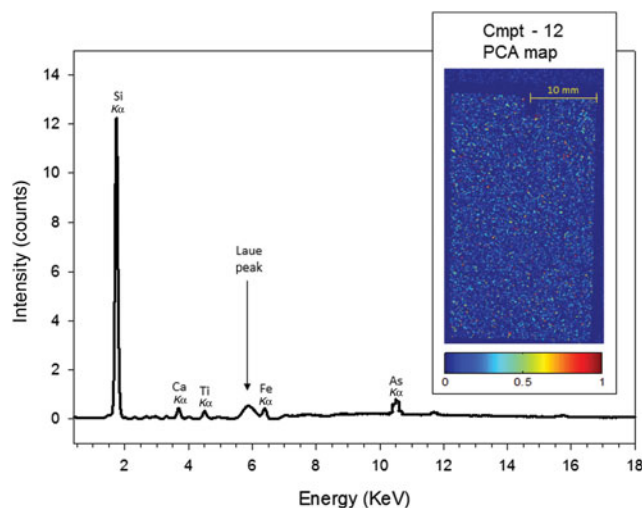
Figure 5. Derived PCA spectrum and corresponding PCA map for Cmpt-12. The Intensity vs. Energy plot is identical to that of Figure 1, but with the addition of a broad Laue peak at ~5.8 keV. The PCA map shows that the distribution of this component is correlated with large, individual grains distributed throughout the specimen.

of PCA to remove systematic errors that might otherwise contaminate the quantitative analysis of given phases. This is particularly true of Laue peaks. For example, the Cmpt-12 histogram, shown in Figure 5, is nearly identical to that of Cmpt-1 (quartz). However, there is one major difference: a broadened peak at ~5.8 keV. This peak would not easily identify with any expected elemental species from the specimen, but could possibly be mistaken for $MnK\alpha$ (5.90 keV). The fact that this peak showed a broadened peak profile led to immediate suspicion regarding its origin. Evaluation of the PCA map for Cmpt-12, as shown in Figure 5, supports the identification of this component as being from Laue diffraction effects. As seen in the Cmpt-12 PCA map, the appearance of individual grains, evenly distributed throughout the specimen in a similar fashion to that of the quartz phase, is highly suggestive of Laue diffraction. In addition, the spectrum for Cmpt-12 tags the Laue peak with the $SiK\alpha$ peak, assuring that this effect is from the quartz phase. Geometrical considerations regarding the beam and detector angular orientation tentatively assign the Laue energy as being diffraction from the quartz (102) and/or (110) planes. Here again, the effectiveness of the PCA technique is observed. The method not only helps remove potentially misleading artifacts that may bias quantitative results, but also separates these artifacts for specific assignment, which often leads to additional information gleaned from the technique.

IV. CONCLUSION

The analysis of μ-XRF datasets using PCA revealed a highly detailed, spatially constrained set of components. Detection of trace phases in the core-drilled cross-section was greatly enhanced by PCA when coupled with quantitative powder XRD results. Derived PCA components are quantitative for determining chemical composition. PCA-augmented μ-XRF analysis generates, within seconds, the same spatially distributed chemical information as obtained from many hours of petrographic microscopy. Results support geological characterization of the renewable energy site for CAES. Pyrite oxidation is not a major concern during CAES operation because the mineral was detected at low quantities and it was well dispersed throughout the mineral matrix from the core-drilled specimen.

Dickinson, W. R. (**1970**). "Interpreting Detrital Modes of Graywacke and Arkose," J. Sediment. Petrol. **40**, 695–707.

Gazzi, P. (**1966**). "Le Arenarie del Flysch Sopracretaceo dell'Appennino Modenese: Correlazioni con il Flysch di Monghidoro," Miner. Petrograf. Acta **12**, 69–97.

Jollife, I. T. (**2002**). *Principal Component Analysis* (Springer-Verlag, New York), 2nd ed.

Keenan, M. R. and Kotula, P. G. (**2004**). "Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images," Surf. Interface Anal. **36**, 203–212.

Keenan, M. R. (**2009**). "Exploiting spatial-domain simplicity in spectral image analysis," Surf. Interface Anal. **41**, 79–87.

Larson, A. C. and Von Dreele, R. B. (**2000**). *General Structure Analysis System (GSAS)* (Report LAUR No. 86-748). Los Alamos, New Mexico: Los Alamos National Laboratory.

Rodriguez, M. A., Keenan, M. R., and Nagasubramanian, G. (**2007**). "In situ X-ray diffraction analysis of $(CF_x)_n$ batteries: signal extraction by multivariate analysis," J. Appl. Cryst. **40**, 1097–1104.

Rodriguez, M. A., Van Benthem, M. H., Ingersoll, D., Vogel, S. C., and Reiche, H. M. (**2010**). "*In situ* analysis of $LiFePO_4$ batteries: Signal extraction by multivariate analysis," Powder Diffr. **25**, 143–148.

Succar, S. and Williams, R. H. (**2008**). *Compressed Air Energy Storage: Theory, Resources, and Applications for Wind Power* (Energy Analysis Group, Princeton Environmental Institute, Princeton, NJ, USA), 81 pp.

The Mathworks (**2008**). *MATLAB*. Version 7.7.0.471 (The Mathworks Inc., Natick, MA, USA).

Toby, B. H. (**2001**). "*EXPGUI*, a graphical user interface for *GSAS*," J. Appl. Cryst. **34**, 210–213.