Multivariate factor analysis of heavy minerals concentrate from Athabasca oil sands tailings by X-ray photoelectron spectroscopy

Marshall, Gregory M.; Kingston, David M.; Moran, Kevin; Mercier, Patrick H. J.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1002/sia.5201
Surface and Interface Analysis, 46, 4, pp. 197-209, 2013-02-03

Multivariate Factor Analysis of Heavy Minerals Concentrate from Athabasca Oil Sands Tailings by X-ray Photoelectron Spectroscopy

Journal:	*Surface and Interface Analysis*
Manuscript ID:	SIA-12-0117.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Marshall, Gregory; National Research Council Canada, NRC Energy, Mining and Environment Portfolio Kingston, David; National Research Council Canada, NRC Energy, Mining and Environment Portfolio Moran, Kevin; Titanium Corporation Inc., Process Development Mercier, Patrick; National Research Council Canada, NRC Energy, Mining and Environment Portfolio
Keywords:	multivariate analysis, XPS chemical imaging, mineralogy, oil sands, varimax rotation

Multivariate Factor Analysis of Heavy Minerals Concentrate from Athabasca Oil Sands Tailings by X-ray Photoelectron Spectroscopy

Gregory M. Marshall [1*], David M. Kingston [1], Kevin Moran [2], Patrick H.J. Mercier [1]

[1] NRC Energy, Mining and Environment Portfolio, National Research Council Canada, Ottawa, ON, K1A 0R6, Canada

[2] Titanium Corporation, Edmonton, Alberta, T5J 1G4, Canada

* gregory.marshall@nrc-cnrc.gc.ca

Multivariate factor analysis of X-ray Photoelectron Spectroscopy data acquired from Athabasca oil sands heavy minerals concentrate was used to identify the primary mineral components and their physical associations. Using large-area spectroscopy, a Principal Components Analysis (PCA) and Varimax rotation of the PCA spectral loadings matrix demonstrated that the most significant factors index the mineral chemistry by virtue of the within-factor spectral correlations. Analysis of the Varimax rotated factor scores indicated the physical character of several mineral associations. Emphasis is placed on the high-value materials, namely, zircon and the titanium-bearing minerals. In spectral imaging mode, Varimax rotation in the spatial domain applied to a PCA noise-reduced data reconstruction was used to render component images illustrating the spatial distribution of selected mineral chemistries. The component images also revealed evidence of surface species consistent with pyrite weathering. Data is supported with optical microscopy and energy-dispersive X-ray spectroscopy. Our work demonstrates the utility of multivariate spectroscopic techniques in the analysis of complex mineral chemistry.

Keywords: multivariate analysis, XPS chemical imaging, oil sands, mineralogy, varimax rotation

1. Introduction

The commercial extraction of bitumen from oil sands feedstock employs separation processes based on flotation chemistry. Mineral waste solids are a by-product of these extraction operations, commonly referred to as tailings. Bitumen froth treatment tailings are known to concentrate high-value heavy minerals, such as zircon ($ZrSiO_4$), titanium dioxide (TiO_2), and ilmenite ($FeTiO_3$) in quantities sufficient to motivate recovery efforts.[1,2] Details of the engineering process have been reviewed,[3] and mineralogical characterizations have been reported.[4,5] Heavy mineral recovery is largely based on magnetic and electrostatic separation techniques,[3,6] which rely on the surface condition of the constituent mineral solids. X-ray Photoelectron Spectroscopy (XPS) is a fundamental surface analysis tool that extracts chemical state information; it has the potential to identify factors critical to the mineral surface condition, and the heavy mineral recovery yield consequently.

The surface complexity of Heavy Minerals Concentrate (HMC) demands the use of multivariate approaches in order to interpret XPS spectra in physically meaningful terms. Within this framework, factor analysis is an approach that has the primary goal of simplicity, i.e., a reduction of the number of variables that contribute to a factor representation the data.[7] Two methods of factor analysis that are commonly employed are the Principal Component Analysis (PCA), and Varimax factor rotation,[8] which is a variant of the class of orthogonal transformations known as orthomax rotations. Whereas the former results in abstract factors that are normally limited to a subjective analysis, the latter has been shown in the context of imaging spectra to be effective at spatially resolving the chemical species present.[9–12]

In this paper, both large-area spectroscopy and spectral imaging modes are employed in the investigation of complex HMCs. In the former, each spectrum is comprised of photoemission

from multiple particles, and in the latter, a spectrum is associated with each pixel in the image plane. Pixels may comprise one, or possibly more, mineral species. In both modes, Varimax rotation serves as the primary means of multivariate analysis. PCA is used to obtain an initial set of vectors for Varimax rotation and to provide an effective means of noise reduction.

The objective of the analysis is to render complex spectra in terms of their constituent mineral components and spatial distributions. Secondly, by interpretation of the factor scores, the physical character of selected mineral associations may be observed. By association, we mean two or more mineral types that are physically coincident on a particle basis, e.g., by particle inclusion, intergrowth, intermixing etc. The term association is used in contrast to particle liberation, which refers to the independence of one mineral species from another. In addition to rendering the spectra and images of the primary minerals, Varimax score imaging reveals evidence of surface components throughout the HMC consistent with the chemical products of pyrite weathering. This feature is potentially relevant to heavy mineral separation processes that are affected by the surface condition of the HMC constituents.

While XPS imaging studies of rock and mineral samples using PCA methods have been previously reported with the objective of supporting noise reduction computational methodologies,[13,14] and quantification of elemental species,[15] our results are novel in the application of factor rotations to investigate mineral speciation. Data is supported with optical microscopy and Energy-Dispersive X-ray spectroscopy (EDX).

2. Factor Analysis Methods

All numerical processing was conducted in the MATLAB (v. R2011b) development environment, except where noted below. XPS spectra recorded by either large-area spectroscopy or spectral imaging modes were cast into $m \times n$ matrices $\mathbf{X}$, according to the number of large-area

spots or imaging pixels (m), and the number of spectrum channels (n). Data pre-processing was applied on a case-by-case basis as treated later. A Singular Value Decomposition (SVD) was then employed to perform a factorization of the pre-processed dataset. In general, the SVD is,

$$\underset{m\times n}{\mathbf{X}} = \underset{m\times q}{\mathbf{U}}\ \underset{q\times q}{\mathbf{\Sigma}}\ \underset{q\times n}{\mathbf{V}^{\mathrm{T}}} \tag{1}$$

where **U** and **V** are matrices of the left and right (column-wise) singular vectors, respectively, T denotes the matrix transpose, and **Σ** is a diagonal matrix of the non-zero singular values sorted in decreasing order. The index q specifies the number of non-zeros, and is the smaller of m or n, unless $m = n$, in which case $q = n,m$. Eqn (1) is referred to as the economy SVD in MATLAB, and is contrasted with the full SVD that returns the null space in addition.[16] A PCA of the dataset is obtained directly from Eqn (1) by,

$$\mathbf{X} = \mathbf{S}\mathbf{V}^{\mathrm{T}} \tag{2}$$

where $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}$. In Eqn (2), matrix **S** is a $m\times q$ matrix of mutually orthogonal column vectors. PCA transforms **X** to a new set of uncorrelated variables, called principal components, ordered such that each successive component accounts for a decreasing fraction of the data variance. The j^{th} column vector of **S** lists the contributions of the j^{th} principal component to the m spectra. These are called scores. The j^{th} column vector of **V** contains the projections of a unit vector along the j^{th} principal direction onto axes defining the coordinate system of the n original variables. These are called loadings. As the principal components are ordered by decreasing proportion of variance, truncation at an appropriate level of significance provides an effective means of noise reduction. In Fig. 1, a schematic of the decomposition in Eqn (1) illustrates the truncation of noise vectors to n_p factors of significance.

Varimax rotation is a matrix operation ($f_{\mathbf{R}}$: $\mathbf{W} \rightarrow \mathbf{WR}$) that maximizes the sum of the column-wise variances of the squared elements of the input matrix (**W**), viz.

$$f_{\mathbf{R}}(\mathbf{W}) = \max \sum_{j} \sigma_j^2(\mathbf{W}_{i,j}^2) \quad (3)$$

This transformation can be used to produce a new set of variables for **X**, called factors. Note that factors are synonymous with principal components where PCA has been applied. The intent of Varimax rotation is to improve the physical interpretability of factors by maximizing their simplicity, i.e., by minimizing the number of within-factor correlations. Ideally, a factor representation of **X** can be made with vectors representing the pure chemical components, or the elemental spectra. In the present context, simplicity implies a minimum number of mineral species per spot/pixel, or elements per spectrum.

SVD is used to obtain an initial set of vectors for Varimax rotation. Matrix **U** forms the basis of the component space, and **V** forms the basis of the spectral space. Varimax rotation matrix **R** is orthogonal ($\mathbf{R}^T\mathbf{R} = \mathbf{R}\mathbf{R}^T = \mathbf{I}_q$), where $\mathbf{I}_q$ is the $q \times q$ identity matrix. Under rotation **R**, the orthonormality of the column vectors of **U** and **V** is preserved. Consequently, **R** effects an orthogonal rotation of the basis to which it is applied. Specifically, **UR** denotes a rotation of the component basis and **VR** denotes a rotation of the spectral basis. Consider these rotations, in Eqns (4) and (5) respectively, in the following factor representations of **X**,

$$\mathbf{X} = \mathbf{U}\mathbf{R}(\mathbf{V}\mathbf{\Sigma}\mathbf{R})^T \quad (4)$$

where the input matrix to the Varimax operation may be either $\mathbf{W} = \mathbf{U}$, or $\mathbf{W} = \mathbf{V}\mathbf{\Sigma}$, and

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{R}(\mathbf{V}\mathbf{R})^T \quad (5)$$

where $\mathbf{W} = \mathbf{V}$, or $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}$. Implicit in Eqns (4) and (5) are different rotations **R** depending on the choice of **W**. Note that in Eqn (4), $(\mathbf{V}\mathbf{\Sigma}\mathbf{R})^T\mathbf{V}\mathbf{\Sigma}\mathbf{R}$ does not form a diagonal matrix in general, i.e., $\mathbf{V}\mathbf{\Sigma}\mathbf{R}$ is not composed of mutually orthogonal column vectors. Similar arguments apply to $\mathbf{U}\mathbf{\Sigma}\mathbf{R}$ in Eqn (5). This has important consequences in that, simply stated, Varimax trades simplicity in

one space for relaxed contrast in the other.[12] Where the initial matrix, either **V** or **U,** is first column-scaled (right-multiplied) by **Σ** to form input matrix **W** in Eqn (4) or (5), the transformation $\mathbf{W} \rightarrow \mathbf{WR}$ preserves the column orthonormality of vectors in the complementary space, i.e., in the matrix **U** or **V**, respectively. Such **Σ**-scaling is common in factor modeling if the goal is to produce a matrix of mutually orthonormal factors. Application of Eqns (4) and (5) will be considered in view of the sample and experimental conditions as the situation dictates.

3. Experimental Details

As received HMC samples provided by Titanium Corporation were cleaned in toluene by Accelerated Solvent Extraction (ASE) using a Dionex model ASE 350 to remove residual bitumen. The cleaned mineral particles were distributed uniformly on double-sided pressure-sensitive tape backed on a brass coupon for XPS measurements.

XPS spectra and images were recorded using a Kratos Analytical model Axis-Ultra DLD (delay-line detector) with a monochromated Al Kα X-ray source (1486.7 eV). In large-area spectroscopy mode, high-resolution spectra were acquired on a 700 μm × 300 μm spot basis, 37 spots were recorded, each spot separated by 200 μm in a grid pattern spanning a sample area of approximately 20 mm^2, or about 5-10 particles per spot. The analyzer pass energy was fixed at 40 eV, and a charge neutralizing thermionic current was applied. Data were collected in 0.1 eV steps over the following binding energy regions: Fe 2p, O 1s, Ti 2p, Ca 2p, Zr 3d, S 2p, Si 2p, and Al 2p, for a total of 1308 concatenated channels. In imaging mode, spectra were collected as a binding energy series of images dimensioned 800 μm × 800 μm in size, spaced in 0.2 eV steps, using a pass energy of 80 eV, charge neutralization, and a per image acquisition time of 4 minutes. The imaging stack consisted of 256×256 pixel images, concatenated over the same binding energy regions, for a total of 567 spectral channels. For both spectroscopy and imaging

modes, binding energy calibration was made to the Ti 2p (3/2) spin-orbit branch at 458.6 eV for approximate charge correction. The C 1s line was not recorded in either mode, owing to the multiplexed contribution from several sources, including a background from the adhesive tape.

Optical microscopy, Leica Microsystems: model DMRXE, and a Scanning Electron Microscope (SEM), JEOL: model JSM-840A, equipped with an EDX spectrometer, Oxford Instruments: model 6560 INCAx-sight / light element, provided supporting data. SEM-EDX spectra were acquired from Au sputter-coated samples on conductive carbon tape at a spectral resolution of 129 eV, and 20 eV data spacing, using a 20 kV accelerating voltage and a 2 nA beam current. Mineral spectra were generated from 512×384 pixel SEM-EDX images at 75× magnification by identifying pixel groupings in elemental ternary diagrams using the INCA microanalysis software.

The minerals reported in this work have been independently observed as crystalline phases by X-ray Diffraction (XRD) in unpublished work. Data is available on request.

4. Results and Discussion

4.1 Optical Imaging and SEM-EDX

Fig. 2 displays optical images of the most frequently observed particles: gypsum ($CaSO_4$-$2H_2O$), pyrite (FeS_2), quartz (SiO_2), ilmenite ($FeTiO_3$), leucoxene, and zircon ($ZrSiO_4$). Pyrite oxidizes in the presence of water to produce several iron and sulphur bearing species including hydro(oxides) and sulphates.[17,18] Leucoxene is an alteration (weathering) product of ilmenite resulting from oxidation and the leaching of ferric (Fe^{3+}) iron.[19] The terminal product of alteration includes polymorphs of TiO_2, primarily rutile, but may also include anatase.[20] A loosely defined mixture, leucoxene is not a recognized mineral species.[21] In Fig. 2e), elbow twin formations characteristic of rutile are evident.[20] In Fig. 2f), the typical morphology of

zircon crystal is easily observed and its birefringence is made clear by extinction under 90° rotation of a polarization analyzer.

Fig. 3 shows the SEM-EDX spectra corresponding to the above materials. Counts are additive over pixels selected from elemental ternary diagrams. Qualitatively, the main elements reflect the underlying mineral chemistry. However, due to a variety of morphologically based effects, accurate stoichiometric analyses are best achieved on polished sections not given here. That being said, an estimation of the elemental fractions is possible for the purpose of mineral classification. Consider the alteration of ilmenite. As a weathering process, it is natural to expect a distribution of the fraction Fe/(Fe+Ti) between stoichiometric ilmenite (50 %), and TiO_2 (~ 0%), which has been documented for similarly sourced HMCs by Kaminsky et al.[4] By choosing a boundary criterion of 10%, the Ti-bearing minerals' spectra in Fig. 3 can be labelled as leucoxene and *altered* ilmenite. The resulting spectra represent approximately equal proportion of these terminal and transition species, with group average Fe/(Fe+Ti) ratios of 4% and 29% respectively, indicating a strong degree of alteration.

4.2 Large-Area Spectroscopy

XPS spectra obtained in large-area spectroscopy mode were first Shirley background subtracted using CasaXPS software. This removed channel correlations to lower binding energies resulting from inelastic scattering. Fig. 4 shows an overlay of all large-area spectra in the Ti 2p and Zr 3d binding energy regions after background subtraction. It is clear that both elements are represented in each spot specific spectrum and, likewise, the remaining elements were represented in each spot. The data matrix $\mathbf{X}$ was then autoscaled ($\mathbf{X}_{sc}$), i.e., mean-centered and normalized such that its column vectors had unit variance. In matrix form this is written as,

$$\mathbf{X}_{sc} = \left[\mathbf{I}_m - \frac{1}{m}\mathbf{i}_m^{T}\mathbf{i}_m\right]\mathbf{X}\boldsymbol{\Omega}^{-1} \quad (6)$$

where $\mathbf{i}_m$ is a m-element row vector of ones, and $\mathbf{\Omega}$ is a $n \times n$ diagonal matrix formed from a column-wise evaluation of the standard deviations of $\mathbf{X}$. In large-area spectroscopy mode, there is a dominant average spectrum that mean-centering removes. This average spectrum is indicated in Fig. 4 by the solid line. Its removal brings the spot-spot variation into greater significance. Subsequently, normalization confers unit variance to each channel. The variance captured in the principal components is then defined about the normalized mean spectrum. Without normalization, a large spread in variance was observed among the principal components, most likely due to mineral abundance differences, which made identification of the level of factor significance difficult. One disadvantage of column-wise normalization is that on-peak channel variance is down-scaled to the level of off-peak channel noise. To mitigate deleterious effects, off-peak channel contributions were truncated by concatenation of binding energy regions, as indicated by Fig. 4.

The level of factor significance can be determined by the breakpoint in a scree plot, which is, in general, a graph of the eigenvalues ($\mathbf{\Lambda}$) of the data matrix $\mathbf{X}^T\mathbf{X}$. The eigenvalues can be obtained directly from the matrix of singular values since $\mathbf{\Sigma}^2 = \mathbf{\Lambda}$. Following SVD of the matrix $\mathbf{X}_{sc}$, the eigenvalues of $\mathbf{X}_{sc}{}^T\mathbf{X}_{sc}$ refer to the variances of the principal components by a scale factor of $1/(m\text{-}1)$. Fig. 5 plots the variance (left), and cumulative variance (right) for all principal components. Having removed the mean spectrum in pre-processing, there are only m-1 non-zero singular values, therefore, the m^{th} principal component is null. The first six principal components, which account for about 40% of the total variance, were considered most significant for the purpose of Varimax factor representation. Truncation at six factors was based *a priori* on the number (6) of primary mineral chemistries in Fig. 3, and by consideration of the breakpoint in Fig. 5 that occurs between factors 5 and 6.

A score plot is a graph of the scores on any two factors and is used to illustrate relationships between the *m* objects in a given factor representation. Such plots are commonly used to illustrate group classifications. In large-area spectroscopy, the principal components were sufficiently multiplexed such that no classifications were discernible. For autoscaled data, this manifested as an unstructured distribution of scores about the origin of the PCA score plots, examples are which are provided in the Supporting Information. This indicates a distinct lack of preferred direction in the component space, therefore, rotation of the component basis was rejected in favour of spectral basis rotation in Eqn (5). For similar reasons, input matrix $\mathbf{W} = \mathbf{U\Sigma}$ was rejected in favour of $\mathbf{W} = \mathbf{V}$ for use in the algorithm defined by Eqn (3).

The MATLAB code for the Varimax routine follows the discussion by Harman,[22] and is provided in the Supporting Information. In the routine, the rows of the input matrix **W** are divided by their lengths as a user option. This is the Kaiser normalization,[8] which gives each variable equal weight prior to rotation. After rotation, the lengths on **WR** are restored with the original lengths of **W**, as they are rotation invariant. A variable's length squared is its communality, which is the relative proportion of a variable's variance accounted for in the retained factors. Note that Kaiser normalization down-scales on-peak channels to the level of off-peak channels. However, as discussed in the context of autoscaling, concatenation of the XPS regions about their respective on-peak channels minimizes the number of off-peak channels contributing to the dataset.

4.2.1 Varimax Factor Loadings

In order to compare the PCA loadings matrix **V** and Varimax loadings matrix $\mathbf{G} = \mathbf{VR}$ with restored spectral units, **V** and **G** were row-scaled by the matrix of standard deviations, i.e., $\mathbf{\Omega V}$ and $\mathbf{\Omega G}$ such that reconstitution of the original data with Eqns (1) and (6) is,

$$\mathbf{X} = \mathbf{M} + \mathbf{X}_{sc}\boldsymbol{\Omega} \tag{7}$$

$$\mathbf{X}_{sc} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{T} \tag{8}$$

$$\mathbf{X} = \mathbf{M} + \mathbf{U}\boldsymbol{\Sigma}\left(\boldsymbol{\Omega}\mathbf{V}\right)^{T} \tag{9}$$

and similarly, following a Varimax rotation of the spectral basis,

$$\mathbf{X} = \mathbf{M} + \mathbf{U}\boldsymbol{\Sigma}\mathbf{R}\left(\boldsymbol{\Omega}\mathbf{G}\right)^{T} \tag{10}$$

In the above equations, **M** is a $m \times n$ matrix of m replicates of the mean spectrum of **X**. Furthermore, on account of row-scaling by **Ω,** note that Eqns (9) and (10) do not represent linear combinations of **X** in terms of mutually orthogonal spectral loading vectors, since $\mathbf{V}^{T}\boldsymbol{\Omega}^{2}\mathbf{V}$ and $\mathbf{G}^{T}\boldsymbol{\Omega}^{2}\mathbf{G}$ are not diagonal matrices in general.

Fig. 6 graphs the spectral loadings of the first five factors following PCA (**ΩV**, solid line) and Varimax rotation (**ΩG**, filled area). Within each Varimax factor, inspection of the photoelectron lines yield an index to the elements contributing to the minerals' speciation. The Varimax factors conform to the mineral chemistries observed in SEM-EDX (Fig. 3) with the exception of factor 3, which is the index of Ti-bearing minerals. This case is discussed below. The loading on Varimax factor 6 indicated a positive correlation of Al 2p, Si 2p and O 1s photoemissions suggesting alumino-silicate of unspecified type, not germane to further discussion and, therefore, was not included in Fig. 6. Similarly, other small-quantity mineral varieties exist in the HMC at lower levels of significance.

Before continuing with the analysis, some discussion on the effect of the various scaling options is warranted. To reiterate, both autoscaling and Kaiser normalization were used in generating the results in Fig. 6. Without autoscaling, i.e., mean-centering only, cross-mineral elemental correlations were observed in the Varimax loadings, which provided a less accurate

representation of the mineral chemistry. Consequently, mean-centering was rejected on the grounds that it did not yield a physically interpretable result. With autoscaling applied, we observed that with or without Kaiser normalization Varimax returned nearly identical results. Inspection of the communalities of matrix **V** revealed that similar on-peak maxima among the XPS regions were evident in the autoscaled data. The consequence of similar initial communalities is that the re-weighting effect of Kaiser normalization is less pronounced, which manifested in nearly identical Varimax results. At a minimum, this equivalence reinforces the statement that Varimax is stable with respect to Kaiser down-scaling when concatenation of binding energy regions is employed.

Returning to the analysis of Fig. 6, observe that Varimax factor 1 corresponds to pyrite. Its loading in the S 2p region indicates a sulphate (SO_4^{2-}) chemical state evident by the intensity of the S 2p (3/2) photoemission near 169.4 eV. The presence of sulphate means the pyrite surface has been oxidized. Note further that Varimax factor 1 and PCA factor 1 are similar. In spectral analysis, the first principal component can be interpreted as the average morphological variation,[13,23] which in this case represents the topography and particle density variation from one analysis spot to the next. The high correspondence of PCA factor 1 with a Varimax loading for pyrite suggests that pyrite accounts for most of the morphological variation in the large-area spectra. This point will be considered further in section 4.3.3.

In the set of Varimax factors, there is a lack of speciation between the Ti-bearing minerals leucoxene and altered ilmenite. This is to be expected given the extent of ilmenite alteration and the resulting elemental overlap with TiO_2 polymorphs. Furthermore, quantitative phase analysis by XRD has indicated that the mass ratio of crystalline $FeTiO_3$ vs. TiO_2 polymorphs is about 1:4 in the HMC. In terms of atom percent, this computes to a Fe/(Fe+Ti)

ratio of 7%, which supports the low Fe 2p intensity observed in factor 3. Consequently, the chemical state in the Ti 2p region is largely attributed to TiO_2. However, more than one peak is evident in the O 1s region. The peak at 530.0 eV is attributed to TiO_2. The origin of the peak near 531.3 eV is less clear. Its binding energy approximates that of ilmenite,[24] but its ratio with the TiO_2 peak is too large to be accounted for by ilmenite alone, given that the expected atomic oxygen ratio of these components is only 12%. The negative peak near 533 eV appears to be a residual of the Si-O correlation (silicate) observed in the corresponding PCA factor.

Most evident in terms of the effect of Varimax rotation is factor 5, which represents quartz. The improved factor simplicity is highlighted by the strengthening of the Si and O positive correlation, and the minimization of Fe, Ti, Ca, and Zr correlations. Recall that the intent of a Varimax rotation is to reduce the set of correlated variables in order to better reflect chemical speciation, which is clearly apparent in this case.

In factor 2 representing gypsum, the Varimax loading indicates a Ca 2p (3/2) line at 348.3 eV and a sulphate chemical state of S 2p (3/2) near 169.7 eV. The O 1s line is also positively correlated with Ca and S, and consists of a chemical state near 532.4 eV, indicating sulphate oxygen, and a peak broadened series of states above 534.3 eV likely due to a dihydrate phase. The O 1s peak of gypsum appears underweighted by a negative peak, which is a residual of the Si-O correlation in the PCA factor, as was similarly observed in factor 3. In previous studies characterizing similarly sourced HMCs no gypsum was reported,[4,5] however, we have observed it periodically, and in the present HMC samples in particular.

Factor 4, which is the index of zircon, will be discussed in section 4.3.3 in view of the results from spectral imaging and the results from an analysis of the factor score plots from

large-area spectroscopy. In general, note that the simplicity in factor 4 has improved upon Varimax rotation. However, the correlation of elements Zr and O to Si is weak.

As a final comment on the loadings, we note that accurate stoichiometric ratios were not always obtained for the mineral species. For example, after relative sensitivity factor (RSF) and transmission function corrections, the Ca/S ratio in factor 2 is approximately correct for gypsum at 0.84, whereas the Si/O ratio in factor 5 at 1.1 is less accurate with respect to quartz. As mentioned, residual elemental correlations compete with the elemental correlations attributed to mineral speciation. These residual correlations may be artefacts of the column-orthonormality constraint imposed by the SVD, and the inherent lack of mineral discrimination in the large-area measurement. The latter violates the assumption of simplicity implicit in the Varimax approach. As a result, some factors carry residual evidence of component mixture. Residuals may get amplified by the rescaling of **G** by **Ω**, especially in the case of silicate residuals, since the spot-spot variation of highly abundant minerals, like quartz, tends to be large. In addition, the extent of natural and in-process surface conditioning is unknown and may limit stoichiometric comparison to bulk compositions. Furthermore, removal of the mean spectrum **M** limits the chemical information in the spectral loadings. Since the mean spectrum is dominant in large-area mode, adding back the mean spectrum, as in Eqns (9) and (10), would simply mask the spot-to-spot variation, which has now been recast in terms of mineral speciation. Consequently, the value of the Varimax spectral loadings is limited to their provision of a mineral index to the variation about the mean spectrum.

4.2.2 Varimax Factor Scores

Complementing the spectral loadings is an analysis of the factor scores following Varimax rotation. Fig. 7 shows a series of four score plots resulting from rotation of the spectral

basis. The factor scores are obtained from the columns of **UΣR** in Eqn (10). They indicate the relative contribution of the component spectra (**ΩG**) to the mean-centered data. In large-area mode, larger or smaller contributions occur as a result of variations in mineral abundance within each spot measured. Therefore, positive/negative scores indicate more/less abundance of a particular component in a given measurement. Abundance increases arise from mass increases per particle (by particle size increase or fractional mass increase per particle), or an increase in the number of particles per spot of that component. Consequently, interpretation of the score plots is made by relating abundance variations to component mass or number. In the limit of large measurement area and small particle size, one would expect a random distribution of component scores about the origin of the score plots commensurate with the random nature of particle number variation. As the measurement area approaches the order of the particle size fewer particles are measured and, therefore, one would expect the scores to fall closer to the component axes. Consequently, the likelihood of capturing one component and not another increases with a decrease in the measurement area to particle size ratio. This likelihood also increases with the spatial dispersion of the components, which increases with the number of components in the mix. Finally, a lack of scores in any quadrant means that the observed abundance variation is not significantly due to particle number. In this case, patterns among the scores can be interpreted in physical terms that relate to variation of component mass on a per particle basis.

Since a rotation of the spectral basis was made, the complementary matrix **UΣR** of the rotated component space does not consist of mutually orthogonal column vectors. This limits the independence of the component axes, which arise from the columns of **UΣR**. In Table 1, the

angles (θ) between the respective column vectors of **UΣR** are provided so that the reader may assess the degree of component orthogonality.

In Fig. 7a), the scores on the factor representing the Ti-bearing minerals are plotted against the zircon factor scores (θ = 77.1°). In the 1st quadrant, the factor scores for zircon and the Ti-bearing minerals fall near their respective component axes (circled). This indicates a mutual independence of the two mineral species in the measurements. The relative absence of scores in the 2nd and 4th quadrant indicates that the abundance variation is not significantly due to particle number differences. Therefore, the scores in the 1st quadrant can be interpreted as representing the mutual liberation of zircon and the Ti-bearing minerals for the largest component masses. The liberation of these minerals is a well-known property of HMC;[2,4,5] it is fundamental to separation technologies that exploit differences in their respective conductivities.[3,6] The stated relationship is exclusive of the 3rd quadrant where the components appear together. Here the component mass is smaller for both species. Three cases arises: i) the two components are associated at smaller particle sizes, or more likely; ii) intermixing of small amounts of these components occurs in other particles, such as quartz; and iii) small liberated particles have been captured in the large-area spectroscopy field-of-view.

In Fig. 7b), the factor scores of zircon are plotted against the quartz scores (θ = 82.8°). Exclusive of a few outliers, the circled region captures an anti-correlation between these two components. In addition, the relative absence of scores in quadrants 1 and 3 suggests an independence from particle number variation. Consequently, the apparent anti-correlation is interpreted as resulting from differences in mass association, i.e., a continuum of intermixing such that: i) the largest zircon masses are relatively free of quartz association, ii) the largest

quartz masses are relatively free of zircon association, and iii) at component masses near the average (at the score plot origin), an equally blended composition of the two is indicated.

In support of this interpretation, consider that quartz inclusion in zircon particles has been documented in similarly sourced HMC and,[4,5] additionally, we have observed the presence of zircon association in quartz particles. Evidence of this will be presented in section 4.3.1 by imaging. In liberation analyses to date, zircon has been reported as relatively pure and highly liberated, with the exception of some quartz inclusion as cited above. Our results do not contest this for the largest and, therefore, most desirable particles. Here, the analysis of zircon-quartz association extends into particles that may be rejected as gangue material owing to their high quartz composition.

In Fig. 7c), the factor scores of pyrite are plotted against the quartz scores ($\theta = 67.2°$). While these materials are considered as gangue, the association evident in this score plot is of value in that it further supports use of the factor scores in the investigation of mineral association. Pyrite-quartz association in similarly sourced HMC has been documented.[4,5] The circled scores in the first quadrant highlight a positive correlation between the two minerals. This relationship differs from that of mutual exclusion, i.e., is one such that the quartz variation moves with pyrite variation in fixed proportion. The association is most apparent for the largest quartz masses as represented in quadrant 1. While not directly important to the extraction of high-value minerals, this association is relevant to pyrite flotation processes that use hydrophobic collector agents, such as xanthate salts; the hydrophilicity of the associated quartz would critically hinder flotation of the pyrite + collector agent complex.

In Fig. 7d), the scores on the factor of Ti-bearing minerals are plotted against the scores on the quartz factor ($\theta = 79.1°$). The scores appear distributed in all quadrants and, therefore, no

association is apparent. However, evidence of silicate intergrowths have been observed in both leucoxene and ilmenite.[4,5] Furthermore, imaging evidence of the mutual intergrowth of Ti-bearing and Si-rich minerals is presented in section 4.3.1 Analogous to the case of inclusion, mutually intergrown minerals are expected to exhibit anti-correlation in the score plot. Unfortunately, a lack of discrimination with respect to leucoxene and ilmenite may limit an interpretation in this regard. Moreover, not all Ti-rich particles are intergrown with silicate. Some suggestion of anti-correlation is captured by the circled region near the origin in Fig. 7d). Overall, however, the distribution of scores does not make it clear. Further evidence of the mineral associations described in this section is provided by SEM in the Supporting Information.

4.3 Imaging Mode Analysis

In this section, XPS spectral imaging data is first treated with PCA noise reduction and synthetic peak fitting algorithms, which is then used to render a Red-Green-Blue (RGB) colour mapping of selected mineral elements. The noise-reduced and peak-fitted spectra are also used to render Varimax rotated spatial components and spectra of selected mineral chemistries.

4.3.1 PCA-Based Data Reconstruction

The procedure for PCA-based data reconstruction of XPS imaging datasets was adapted from Walton and Fairley,[25] and is summarized as follows: i) dataset autoscaling and SVD according to Eqns (6) and (8), ii) PCA noise truncation retaining the most significant factors according to Fig. 1, iii) reconstitution of the dataset in terms of the channel variance and mean-spectrum according to Eqn (9), iv) background subtraction and fitting of the noise-reduced data, in each XPS region and at each pixel, with a synthetic envelope function representing the regional photoemission properties.

Following autoscaling and SVD, eight factors were retained upon inspection of the scree plot, shown in Fig. 8, which is provided for the first 30 principal components (of n = 567 total). The level of factor significance was determined at the point beyond which the eigenvalues of the imaging data matrix $\mathbf{X}_{sc}{}^{T}\mathbf{X}_{sc}$ varied in an exponentially decreasing manner, i.e., linearly decreasing in the semi-log plot. This criterion has been applied to the truncation of uniform noise, i.e., noise of magnitude independent of the signal magnitude. Walton and Fairley,[14] and others,[12,26,27] used square-root transformations of $\mathbf{X}$ to suppress non-uniform (Poisson) noise prior to imaging PCA. With autoscaling, the column-wise variance (σ_j^2) of $\mathbf{X}_j$ is a good approximation of the Poisson variance (σ_P^2) in the image of the j^{th} channel when there is low feature contrast.[28] Otherwise, σ_j^2 tends to overestimate the measurement uncertainty resulting in down-scaling of feature variation relative to image noise. This may suppress some significant factors in the PCA. Nevertheless, we can expect at least 5 significant factors based on the number of chemical components whereas 8 can be identified in the scree plot in Fig. 8. Moreover, the scree plot exhibits the requisite linearity in the semi-log scale beyond the 8^{th} factor indicative of uniform noise. Inspection of the loadings on these higher order factors confirmed that no apparent chemical information was lost.

As an example of the effectiveness of PCA data reconstruction, Fig. 9 illustrates: i) the pixel data (circles), ii) the PCA noise-truncated data (dashed line), and iii) the synthetic peak fit (solid line) in the Zr 3d and Ti 2p regions at two pixel locations in the image plane, respectively. Note that the data have been linear background subtracted according to the PCA spectrum.

Using the dataset of synthetic peak fitted spectra, Red-Green-Blue (RGB) coloured maps of selected elemental distributions can be made. RGB mapping is performed by: i) intensity correction using the RSFs (transmission function unknown for imaging datasets), ii) integration

of the synthetic spectra with zero-thresholding based on a 10% noise floor, iii) intensity normalization, i.e., integrated/corrected intensities are unit normalized to yield the percent contribution of each channel to a RGB imaging vector, and iv) image rendering where the RGB vector points to a unique color coordinate.

As an example of RGB mapping of HMC, Fig. 10 shows the elemental distributions of Ti (red), Zr (green) and Si (blue). To conform to the 3-channel limit, the input data comprised a subset of the full spectral imaging stack by concatenation of the Ti, Zr, and Si XPS regions only. The chemical features of the various particles are made clear despite the morphological roughness and X-ray shadowing inherent in HMC imaging. Inspection of the map reveals evidence of some of the physical associations described in the context of score plot interpretation in section 4.2.2. For example, Ti and Si intergrowths are observed (points A & B) in particles that are Si-rich and Ti-rich, respectively. Similarly, Si inclusions (C) are observed in Zr-rich particles, and Zr inclusions (D) may be observed in Si-rich particles. Elements Ti and Zr are largely independent as expected, and particles where there is significant intermixing of Zr and Si can also be seen (E). It should be mentioned that the Si component of zircon has been suppressed in this map in order to enhance colour contrast between zircon and quartz.

4.3.2 Rotation of the Spatial Basis

In imaging mode, the component basis referred to with respect to Eqn (4) may be considered a spatial basis, since each mutually orthonormal column vector of **U** forms an image plane. Rotation of the spatial basis was selected for Varimax factor analysis of the XPS images following the work of Keenan,[9] and Smentowski et al.[12] They reported that simplicity in the spatial domain could be leveraged to obtain physically acceptable representations of component spectra by means of orthogonal rotation of the spatial basis.

The Varimax rotated spatial components and spectra were calculated based on the PCA noise-reduced and synthetic peak-fitted spectral reconstruction (**X***). Basing the Varimax rotation on **X*** was done for the aesthetics of viewing, as the component spectra are more clearly rendered by fitted peaks. Prior to SVD, **X*** was mean-centered for reasons to be discussed below. Data was not autoscaled because of the presence of zeros in the synthetic spectra. Following SVD of the mean-centered dataset (X_μ*), the number of non-zero singular values totalled 8, which was the number of principal components retained from the PCA noise reduction step. This confirmed that synthetic peak fitting did not add or subtract any significant sources of variation.

Subsequent to SVD, rotation of the spatial basis was completed with $\mathbf{W} = \mathbf{V\Sigma}$ selected as the input matrix to the Varimax algorithm. The decomposition of **X*** is then written as,

$$\mathbf{X}^* = \mathbf{M}^* + \mathbf{X}_\mu^* \tag{11}$$

$$\mathbf{X}^* = \mathbf{M}^* + \mathbf{UR}\left(\mathbf{V\Sigma R}\right)^T \tag{12}$$

We thank a reviewer for pointing to the fact that rotation of the **Σ**-scaled spectral basis preserves the orthogonality of the spatial basis. This was confirmed by validating that indeed $(\mathbf{UR})^T\mathbf{UR} = \mathbf{I}$. Eqn (12) represents a linear combination of mutually orthonormal image planes (**UR**), herein referred to as the spatial components. The component spectra are the column vectors of **VΣR,** which are not mutually orthogonal in general.

The choice to employ $\mathbf{W} = \mathbf{V\Sigma}$ rather than $\mathbf{W} = \mathbf{U}$ in the rotation of the spatial basis was guided by inspection of the resulting component spectra in a trial and error fashion. We observed that the former rotation performed better at finding a set of components with loadings having a mineralogical interpretation. In contrast, the latter rotation yielded certain non-physical results. We highlight that this is case specific, as Keenan,[9] and Smentowski et al.[12] used rotation of **U** with good results. The difference can be attributed to surface species present in the HMC, as

detailed in the next section. In imaging mode, the presence of surface species violates the criterion of spatial simplicity implicit in rotations using $\mathbf{W} = \mathbf{U}$. Rotation using $\mathbf{W} = \mathbf{V\Sigma}$ optimizes on spectral simplicity and, therefore, is not subject to the spatial criterion, but is subject to the constraint that it yields an orthogonal change of spatial basis. Recall that rotation using $\mathbf{W} = \mathbf{V}$, as specified in Eqn (5), optimizes on spectral simplicity as well, but is constrained to yield an orthogonal change of spectral basis. The component spectra in this case presented a high degree of elemental simplicity, which is contrary to the goal of resolving the spatial detail of elemental correlations attributable to the mineral chemistries.

The choice to mean-center the data was similarly guided by inspection of the resulting component spectra and is again attributed to the presence of surface species. Since XPS is highly surface sensitive, surface species get strongly weighted in the mixing of component spectra. In this case, subtraction of the mean spectrum reduces the elemental correlations between the underlying mineral spectra and those of any surface components, but only to the extent those components are uniformly distributed. Consequently, centering has the potential to improve the simplicity of the input data and, therefore, have a positive impact on the interpretability of the Varimax result. In general, however, column-centering $\mathbf{X}$ will introduce undesirable complexity if the rows (spectra) are already simple; the spectra at pixels will become a mix of pure components and all components implicit in the mean. Consequently, centering is not recommended for rotations employing $\mathbf{W} = \mathbf{U}$. It is interesting to note that rotation $\mathbf{W} = \mathbf{V\Sigma}$ without first mean-centering $\mathbf{X}^*$ yielded a set of Varimax component spectra whose mean conformed to the replicates in $\mathbf{M}^*$, apart from scaling. Agreement in this manner provides a validation of the mean-centering approach adopted in Eqn (12).

Last in the discussion on methodology, we observed that use of Kaiser normalization was also necessary to achieving physically acceptable component spectra. Normalization evens out spectral differences among the factors in terms of abundance, RSF, and transmission function. This was especially relevant with respect to down-scaling of the O 1s region, which was highly weighted owing to the ubiquity and abundance of oxygen. The need to use Kaiser normalization is attributed to not having autoscaled the input data. Recall the Varimax result was nearly insensitive to Kaiser normalization in large-area spectroscopy mode when autoscaling was used.

4.3.3 Varimax Component Images and Spectra

The results of Varimax rotation of the spatial basis are presented in Fig. 11. The spatial components have been contrast adjusted to fully exploit the colour map, which is equally scaled in all frames. Also, a background cut-off of zero score was applied to the component images, prior to contrast adjustment, in order to aid demarcation of the particles. The residual images were inspected to verify that no meaningful negative components were removed. In contrast with RGB mapping, the associated component spectra were not RSF intensity corrected in order to avoid suppression of the Fe 2p representation; Fe 2p has the highest RSF value among the XPS regions considered. The selected factors (1 to 4) represent the following mineral species: 1) quartz, 2) iron hydro(oxide), 3) iron-titanium oxides, and 4) sulphate. Factor 5, which is attributed to zircon, is shown in Fig. 13 and will be discussed below. Gypsum was not abundant within this particular imaging frame, therefore, it was not a significant source of sulphate in Fig. 11. A positive Ca, S and O correlation did appear in factor 7, overall, however, factors 6, 7 and 8 did not have elemental correlations consistent with identifiable minerals, or appeared non-physical in the component spectra. Truncation of PCA factors 6-8 made prior to Varimax

rotation resulted in negligible difference to the results in factors 1-5 consistent with their being at a lower level of significance.

The quartz representation in factor 1 is self-evident. Clearly, the spatial distribution closely matches the distribution of elemental Si in Fig. 10. Similarly, the iron-titanium oxides match the distribution of elemental Ti. Note that the integrated area of Fe 2p in the component spectrum of factor 3 is overweight by about 2 times relative to that of Ti 2p by virtue of the larger Fe 2p RSF. An estimate of the actual Fe/(Fe+Ti) atomic ratio cannot be made without the imaging transmission function. Besides, the particle sampling is small in the imaging field-of-view, so atomic ratios should not be used to generalize the stoichiometry in HMC.

The evidence of iron hydro(oxide) and sulphate presents an interesting case, and highlights the method of leveraging simplicity in the scaled loadings (**VΣ**) in order to make a rotation of the spatial basis. This approach has resulted in resolving the spatial detail of surface components consistent with pyrite weathering. Since pyrite is unstable in the aqueous conditions present during oil sands processing and, particularly, during methods of HMC conditioning and fractionation that may involve contact with water and heat, or simply through contact with a humid ambient, it is expected that pyrite will be highly oxidized. This fact is evidenced by the sulphate chemical state of the S 2p line (169.4 eV) observed in Varimax factor 1 of Fig. 6. The pyrite weathering reaction products are known to include iron hydroxides, iron oxides, sulphates, and others, including elemental sulphur and iron sulphate.[17,18] More specifically, Cai et al. reported that the pyrite weathering layer is characterized by a S-rich outer surface layer and a Fe-rich inner transitional layer.[17] In agreement with this grouping, Varimax factors 2 and 4 were differentiated by their relative abundance of Fe and S and were classified, accordingly, by the generalized chemical terms given.

The broad distribution of these species and their spatial coincidence with the other minerals in Fig. 11 suggest that they are adventitious surface components, though the hottest regions correspond to solid pyrite formations. Possible explanations for these features are: i) break-up and distribution throughout the HMC, as weathered pyrite is very brittle, and ii) pyrite dissolution and precipitation of weathering products out of aqueous solution in process. Moreover, in large-area mode (Fig. 6), the correspondence between the loadings of PCA factor 1 and one of the Varimax factors (in this case, factor 1) is consistent with a surface component interpretation as well. Given the surface sensitivity of XPS, it stands to reason that PCA factor 1, which is an index of morphological variation, would be significantly accounted for by the elements of a uniformly distributed surface component. Since a Varimax factor with loadings nearly equivalent to that of PCA factor 1 was obtained, it follows that PCA factor 1 can be attributed to the variation of a surface component having a mineral chemistry specified by that Varimax factor. Furthermore, the surface distribution of Fe, S, and O evident in imaging mode refers back to the assumptions specific to Eqn (12) and the reasoning behind the selection of rotation and data pre-processing methods; they were necessary to achieve physically interpretable results. As was the case in large-area mode, the lost of chemical information in **M*** was a necessary consequence of yielding interpretable results in the variation about the mean.

In Fig. 12, a particle-wide distribution of elemental sulphur was confirmed by SEM-EDX consistent with a surface species that comprises sulphate. Note that sulphur was not identified as a bulk contaminant in similarly sourced HMCs by SEM-EDX on polished sections,[4] and that the sulphur observed in Fig. 12 was recorded at counts rates nearly equivalent to that of the deposited Au coating. These points assert that the sulphur, as recorded by SEM-EDX, is present as a surficial element and not as an element of the bulk mineral sample. RGB mapping of the

elemental constituents Fe and S also supports that Varimax factors 2 and 4 arise from adventitious surface species. See the Supporting Information for further details.

Returning to the analysis of factors, Figs. 13 a) and b) illustrate the component image and spectrum from Varimax factor 5 representing zircon. Parts c) and d) graph the PCA noise-reduced and synthetic peak-fitted spectra from two separate pixels originating from the particle identified by the arrow in part a). The two line colours (red and blue) in the Zr 3d and Si 2p regions correspond to the two pixels respectively. The spectra in Fig. 13 have not been RSF corrected in accordance with Fig. 11. Parts c) and d) demonstrate that not all pixels with intensity in the Zr 3d region have intensity in the Si 2p region at the chemically appropriate binding energy for zircon. The Si 2p photoemission of zircon should be located at a binding energy of about 102 eV, which is 1.4 eV lower relative to that of quartz, [29] as labelled in part d). As clear from the blue coloured spectrum, a Si 2p peak characteristic of a pure silicate may arise at the expense of the zircon peak. In our analysis, all pixels were fitted with contributions from both species in the Si 2p region. At pixels with intensity in the Zr 3d spectrum, weighted contributions from both Si species could be observed in most cases, examples of which are provided in the Supporting Information. Among these spectra, it was unclear whether the silicate peak was due to a chemical shift of the zircon, fine intermixing of silicate with zircon, or spatial cross-talk with proximate silicate masses. Further study of the zircon surface and the possibility of its chemical transformation is warranted. Such analysis should be made using concentrated zircon fractions.

In view of the above, recall the loading of Varimax factor 4 from large-area spectroscopy in Fig. 6. In the Si 2p region, a small peak at approximately 102 eV is indicated. In addition, a small negatively correlated Si peak appears at a higher energy consistent with silicate. It appears that anti-correlated variation in the abundance of these two chemical states has resulted in a

competition between them, which has effectively nullified the Si 2p intensity in the loading. This type of anti-correlation is consistent with variation in the apparent chemical shift of zircon.

5. Conclusions

Multivariate factor analysis of XPS spectra was used to study HMCs prepared from Athabasca oil sands bituminous froth treatment tailings. In large-area spectroscopy mode, Varimax rotation of the PCA noise-reduced spectral loadings matrix correctly identified the elemental correlations consistent with the primary mineral chemistries. The Varimax factor scores were interpreted in terms of several mineral associations. Known physical and chemical properties of the HMC minerals were used to benchmark our multivariate approach and thereby validate Varimax rotation in the interpretation of factors from complex mineral spectra.

In spectral imaging mode, PCA-based noise reduction and synthetic peak fitting algorithms were used to generate a spectral reconstruction of the imaging dataset. This dataset was used to render RGB mappings of the spatial distribution of selected elements. Orthogonal rotation of the spatial basis of the reconstructed dataset resulted in component images and spectra consistent with the primary mineral chemistries. Furthermore, it revealed surface components consistent with pyrite weathering. Orthogonal rotation of the spatial basis was effected by means of Varimax rotation of the PCA spectral loadings matrix that was column-scaled by the diagonal matrix of singular values resulting from SVD.

Acknowledgements

The authors wish to thank to Daniel Tyo for sample preparation and operation of the ASE. We also thank the Reviewers for helpful and instructive input.

Supporting Information

Supporting information may be found in the online version of this article.

References

[1] *Titanium Corporation Annual Report*, Edmonton, **2010**.
[2] M. Owen, R. Tipman, *CIM Bulletin* **1999**, *92*, 65-73.
[3] Z. Ciu, Q. Liu, T. H. Etsell, J. Oxenford, J. Coward, *Can. Metall. Quart.* **2003**, *42*, 383-392.
[4] H. A. W. Kaminsky, T. H. Etsell, D. G. Ivey, O. Omotoso, *Miner. Eng.* **2008**, *21*, 264-271.
[5] Q. Liu, Z. Cui, T. H. Etsell, *Fuel* **2006**, *85*, 807-814.
[6] R. G. Reeves, U.S. Patent No. 7341658 B2, Mar 11, **2008**.
[7] B. F. J. Manly, *Multivariate Statistical Methods: A Primer*, 3rd ed., Chapman and Hall/CRC: Boca Raton, **2004**.
[8] H. F. Kaiser, *Psychometrika* **1958**, *23*, 187-200.
[9] M. R. Keenan, *Surf. Interface Anal.* **2009**, *41*, 79-87.
[10] M. R. Keenan, U.S. Patent No. 7840626, Nov 23, **2010**.
[11] E. Vicenzi, M. Velbel, *Microsc. Microanal.* **2010**, *16*, 918-919.
[12] V. S. Smentkowski, S. G. Ostrowski, M. R. Keenan, *Surf. Interface Anal.* **2009**, *41*, 88-96.
[13] K. Artyushkova, J. E. Fulghum, *Surf. Interface Anal.* **2004**, *36*, 1304-1313.
[14] J. Walton, N. Fairley, *J. Electron. Spectrosc. Relat. Phenom.* **2005**, *148*, 29-40.
[15] M. C. Biesinger, B. R. Hart, R. Polack, B. A. Kobe, R. S. C. Smart, *Miner. Eng.* **2007**, *20*, 152-162.
[16] The calculation in Eqn (1) is limited to vectors corresponding to the non-zero singular values of **Σ** (economy SVD). This limits either **U** or **V** to column-orthonormality, as one of them may be a rectangular matrix depending on the dimensions of **X**. If the null space was evaluated in addition (full SVD) then matrices **U** and **V** are always square and fully column and row orthonormal.
[17] Y. Cai, Y. Pan, J. Xue, Q. Sun, G. Su, X. Li, *Appl. Surf. Sci.* **2009**, *255*, 8750-8760.
[18] E. Ahlberg, K. S. E. Forssberg, X. Wang, *J. Appl. Electrochem.* **1990**, *20*, 1033-1039.
[19] F. Cardarelli, *Materials Handbook: A Concise Desktop Reference*, Springer-Verlag: London, **2008**.
[20] C. S. Hurlbut, C. Klein, *Manual of Mineralogy*, 19th ed., John Wiley & Sons: New York, **1977**.
[21] J. L. Jabor, A. C. Roberts, *Am. Mineral.* **1995**, *80*, 1073-1077.
[22] H. H. Harman, *Modern Factor Analysis*, 3rd ed., University of Chicago Press: Chicago, **1976**.
[23] B. Hart, M. Biesinger, R. S. C. Smart, *Miner. Eng.* **2006**, *19*, 790-798.
[24] B. Plante, M. Benzaazoua, B. Bussière, M. C. Biesinger, A. R. Pratt, *Appl. Geochem.* **2010**, *25*, 1830-1844.
[25] J. Walton, N. Fairley, *Surf. Interface Anal.* **2004**, *36*, 89-91.
[26] M. R. Keenan, P. G. Kotula, *Surf. Interface Anal.* **2004**, *36*, 203-212.
[27] D. E. Peebles, J. A. Ohlhausen, P. G. Kotula, S. Hutton, C. Blomfield, *J. Vac. Sci. Technol. A* **2004**, *22*, 1579-1586.
[28] B. J. Tyler, G. Rayal, D. G. Castner, *Biomaterials* **2007**, *28*, 2412-2423.
[29] Y. V. Shchapova, S. L. Votyakov, M. V. Kuznetsov, A. L. Ivanovskii, *J. Struct. Chem.* **2010**, *51*, 657-662.

Table 1. Orthogonality of components (θ) after Varimax rotation, in degrees.

factor	pyrite	Gypsum	Ti-bearing	zircon	quartz
pyrite	0	63.0	63.0	73.7	67.2
gypsum	-	0	85.0	83.1	86.0
Ti-bearing	-	-	0	77.1	79.4
zircon	-	-	-	0	82.8
quartz	-	-	-	-	0

Figure 1. Matrix block diagram of the spectral data reconstruction after PCA noise reduction. Matrix truncation marked × in the indicated dimensions retains n_p factors of significance.

Figure 2. Optical microscopy (100× - 200×) of the primary minerals observed in Athabasca oil sands heavy-mineral concentrates: a) gypsum, b) pyrite, c) quartz, d) ilmenite, e) leucoxene, f) zircon (left), and in extinction by 90° rotation of a polarization analyzer (right).

Figure 3. Mineral spectra determined by selection of pixel clusters in elemental ternary diagrams generated from SEM-EDX mapping images. Spectrum counts are additive over the selected pixels of interest.

Figure 4. Overlay of all data from large-area spectroscopy in the Ti 2p and Zr 3d XPS regions. Spectra have been background subtracted and concatenated by region. The solid line corresponds to the spot-averaged mean spectrum. Inset: typical spectrum from the series.

Figure 5. PCA scree plot: variance (left) and cumulative variance (right) of the principal components determined by SVD of the autoscaled dataset $\mathbf{X}_{sc}$ from large-area spectroscopy.

Figure 6. Varimax (filled) and PCA (solid) spectral loadings of the most significant factors following SVD of the autoscaled data from large-area spectroscopy. Loadings have been rescaled according to Eqns (9) and (10).

Figure 7. Factor score plots after Varimax rotation of the spectral basis in large-area spectroscopy. Minerals correspond to the loadings in Fig. 6. Circled areas refer to physical associations discussed in the text.

Figure 8. PCA scree plot: eigenvalues corresponding to the principal components (1-30 of 567) of the autoscaled spectral imaging data. The first 8 factors were retained for reconstruction of the data matrix.

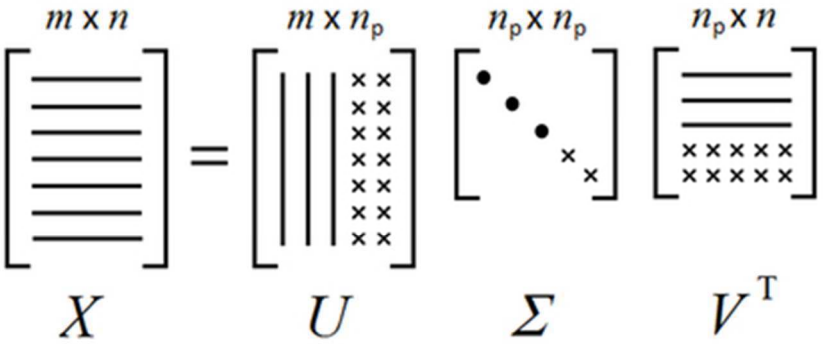
Figure 9. Pixel data (circles), PCA noise reduced spectrum (red) and synthetic peak-fitted reconstruction (blue). Ti 2p and Zr 3d data arise from separate pixels in the image plane. Data has been background subtracted against the PCA spectrum.

Figure 10. RGB mapping of the elemental distribution of Ti (red), Zr (green), and Si (blue) in HMC. Indicated points refer to physical associations of mineral species, see text for details. Map area has been cropped near center to enlarge feature size.

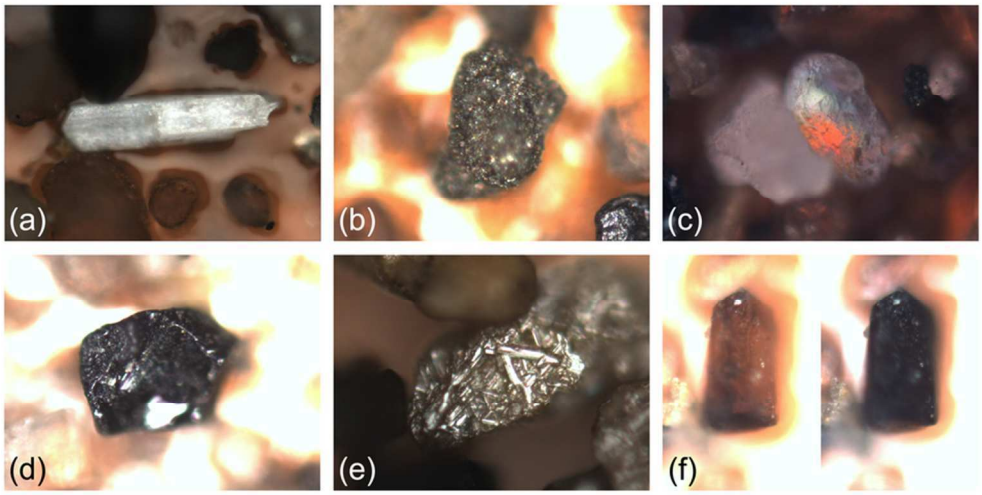
Figure 11. Varimax component images and spectra illustrating the spatial distribution of elemental correlations representing the indicated mineral chemistries. The iron hydro(oxides) and sulphates are consistent with surface species resulting from pyrite weathering.

Figure 12. Top**:** SEM-EDX mapping of the sulphur $K\alpha_1$ line at 2.3 keV after off-peak background subtraction. Solid masses refer to pyrite formations whereas the distributed intensity indicates a surficial element comprised of sulphate. Bottom: corresponding SEM image in back-scattering mode.

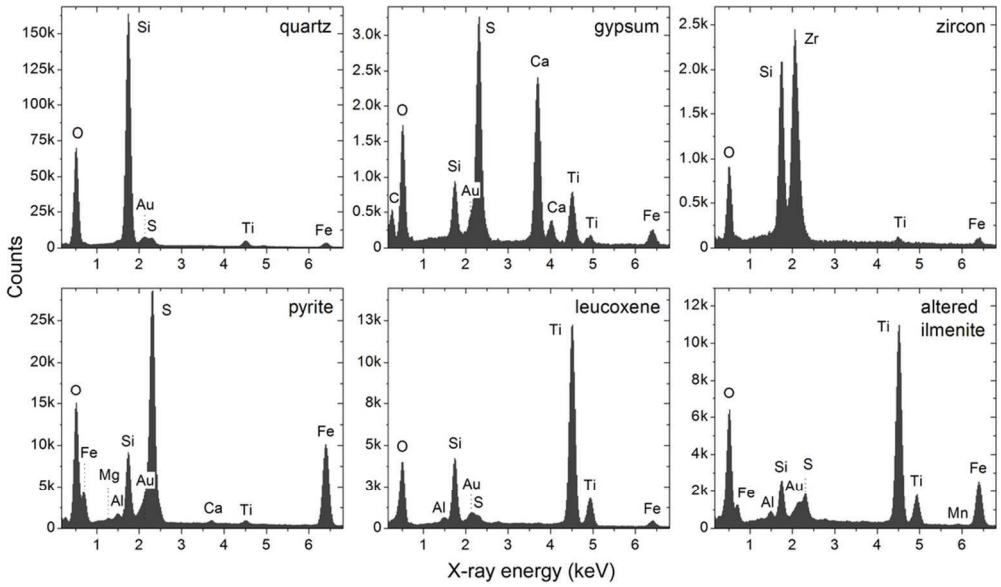
Figure 13. Component image (a), and spectrum (b) of Varimax factor 5 corresponding to zircon. The spectra in (c) and (d) arise from two separate pixels within the particle identified by the arrow. The line colouring (red and blue) corresponds to the individual pixels.



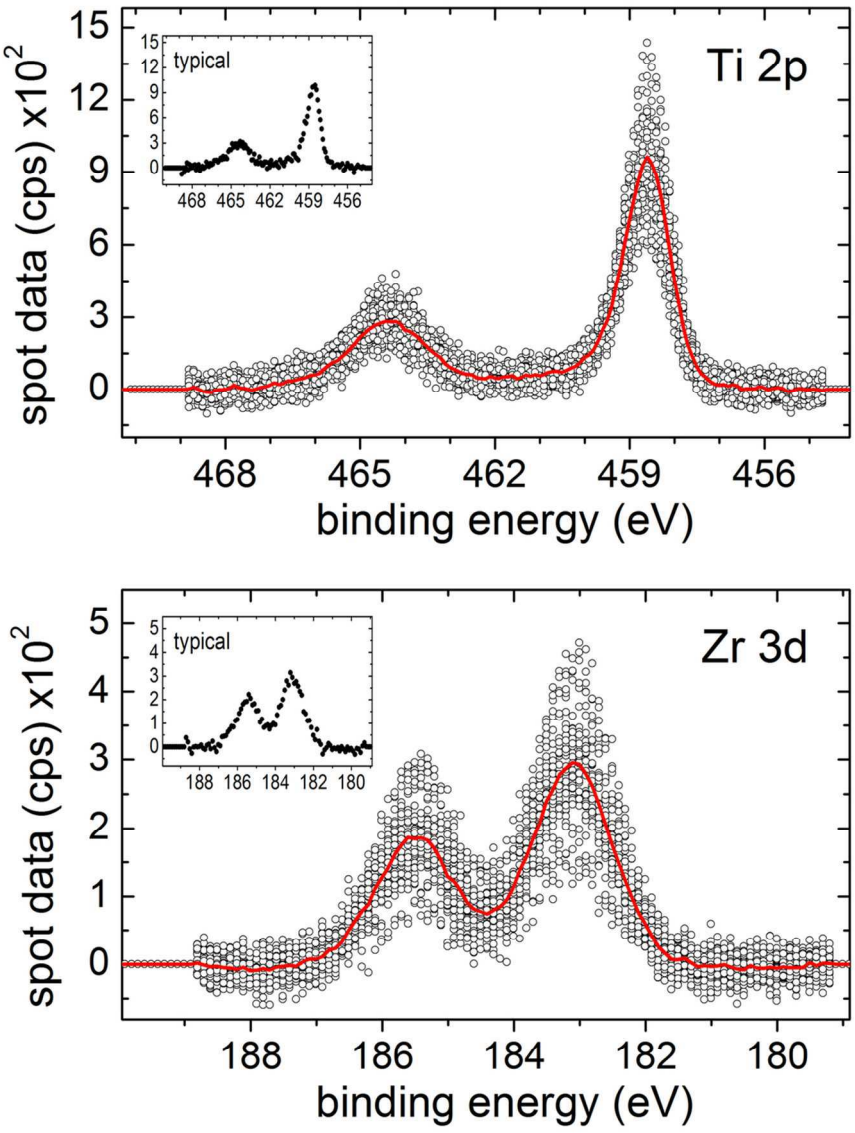
35x14mm (300 x 300 DPI)



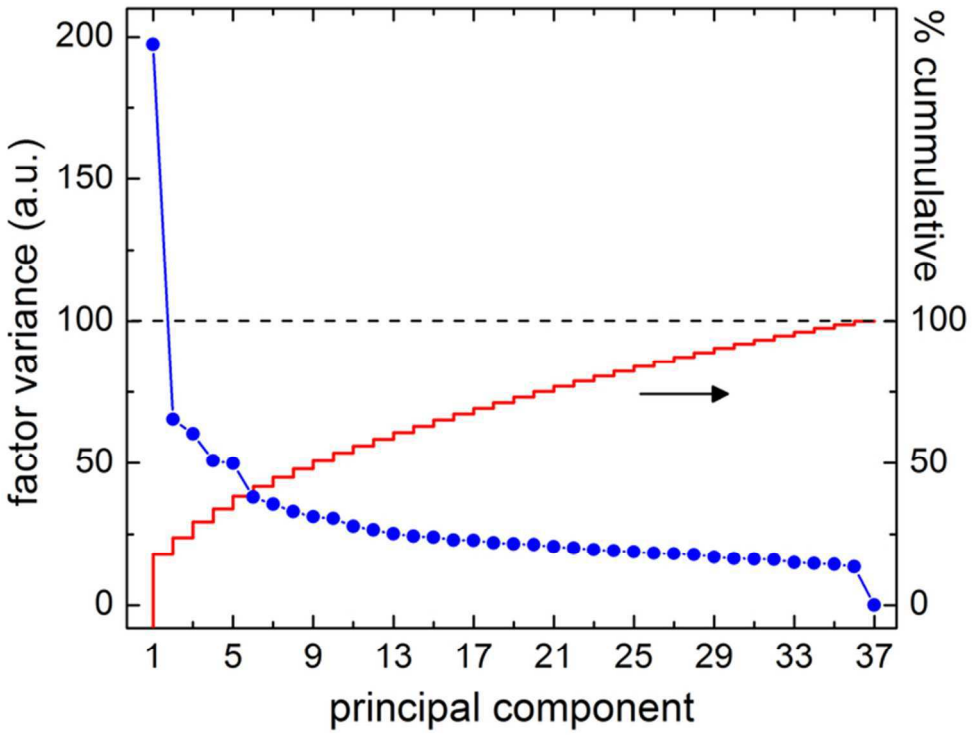
82x42mm (300 x 300 DPI)



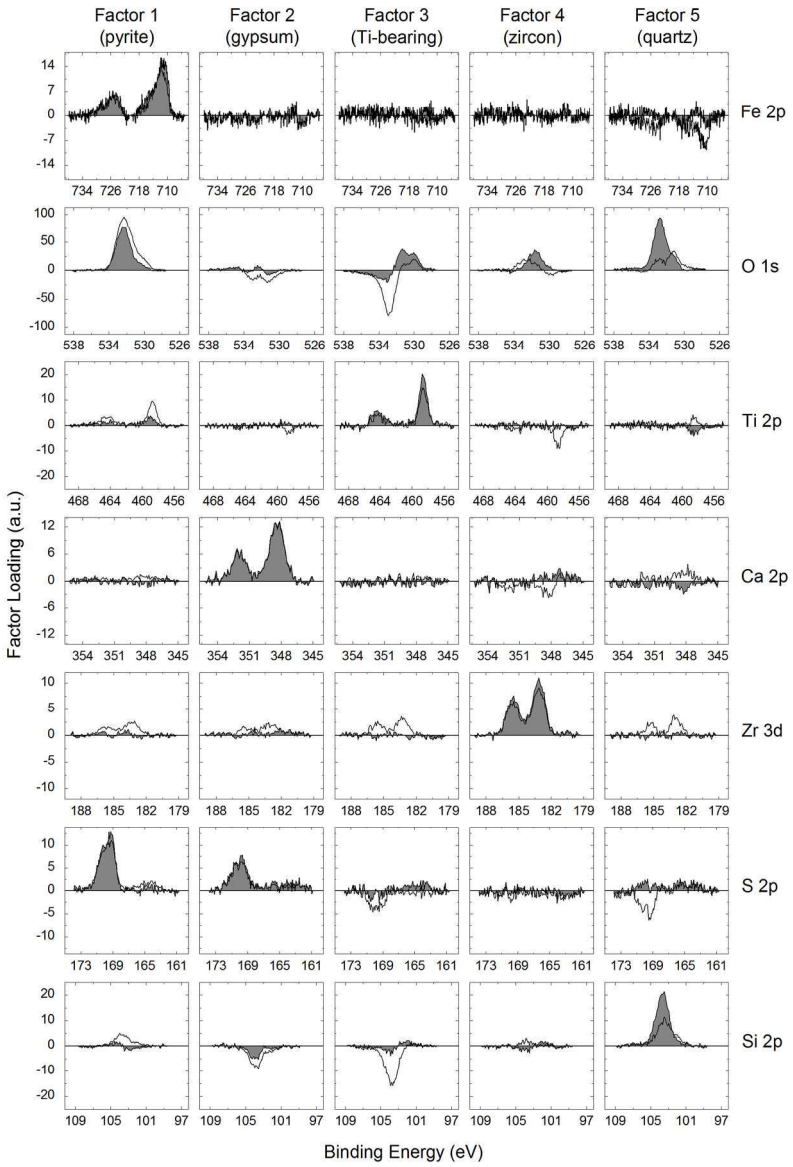
96x56mm (300 x 300 DPI)



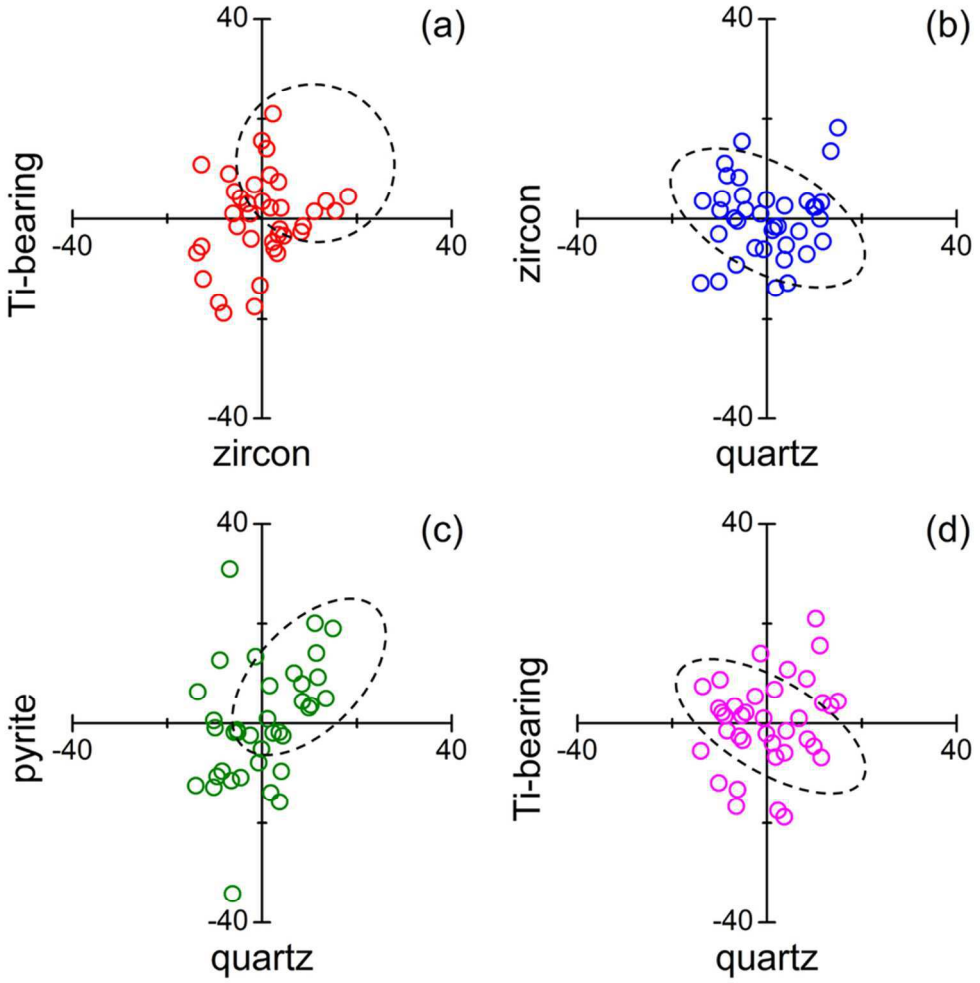
98x128mm (300 x 300 DPI)



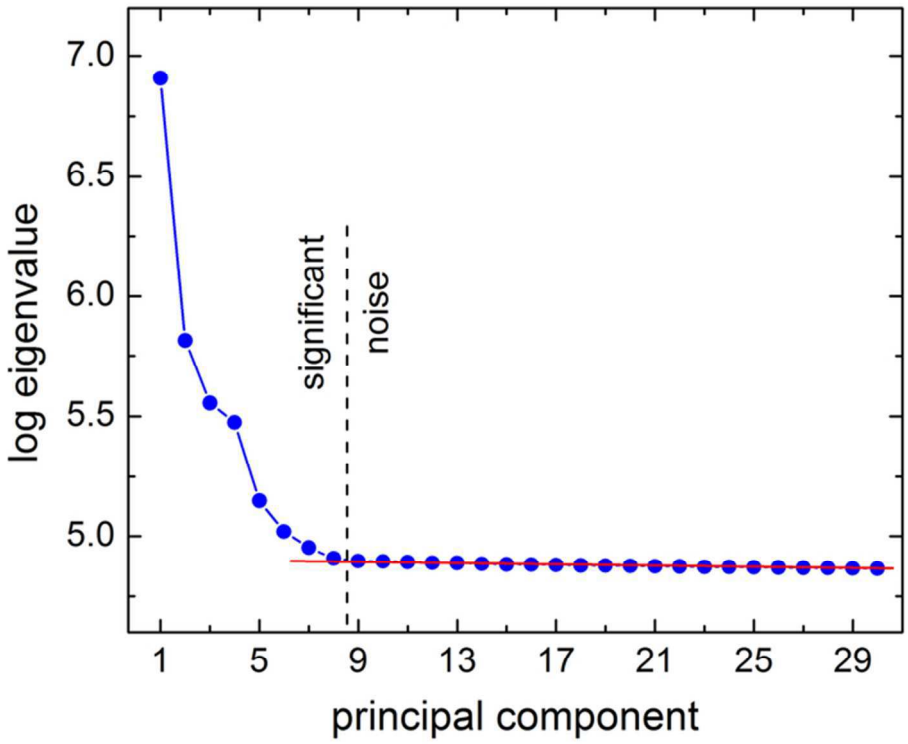
65x50mm (300 x 300 DPI)



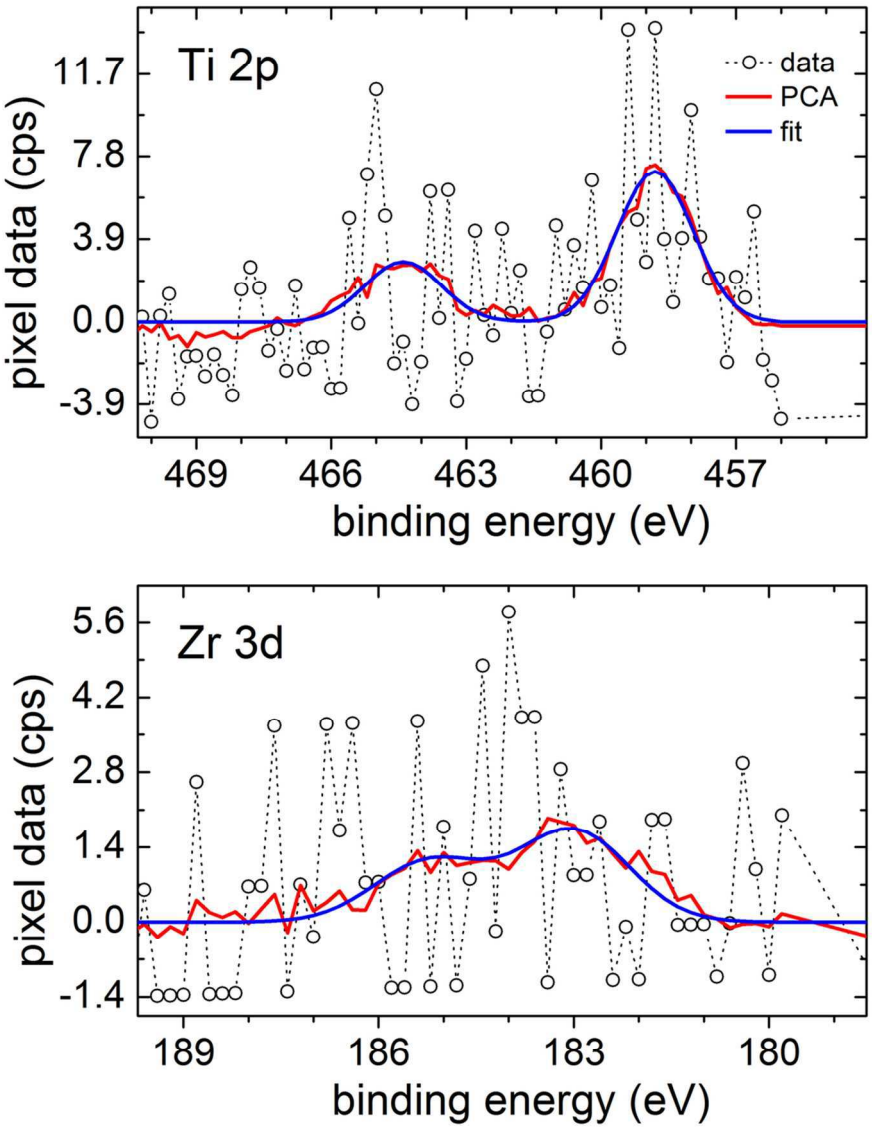
209x306mm (300 x 300 DPI)



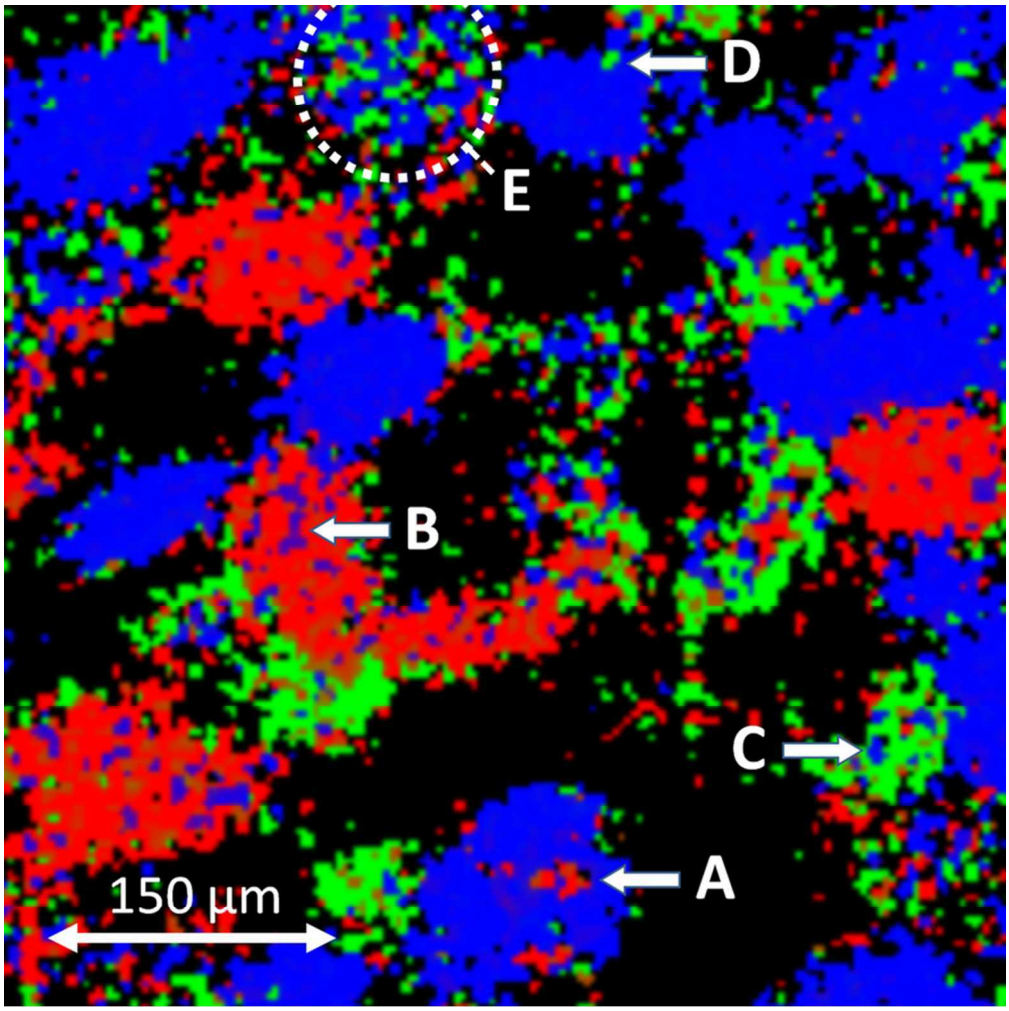
85x86mm (300 x 300 DPI)



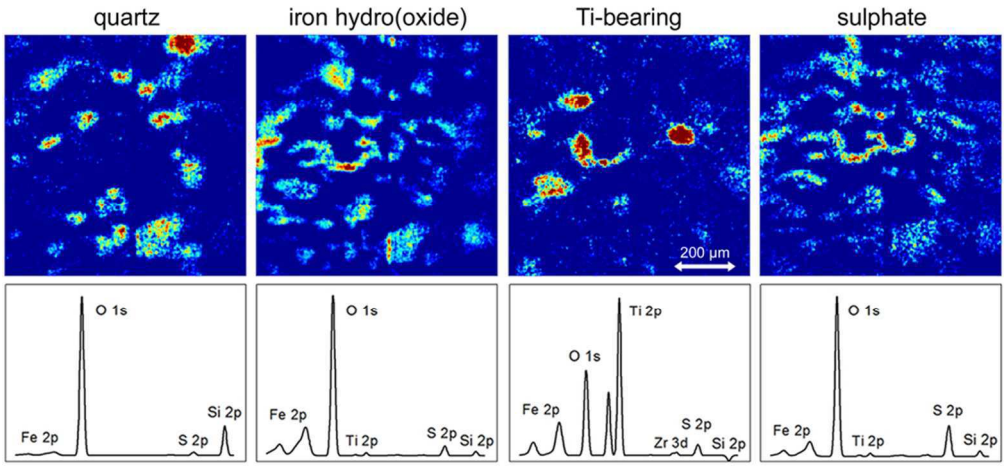
65x51mm (300 x 300 DPI)



98x128mm (300 x 300 DPI)

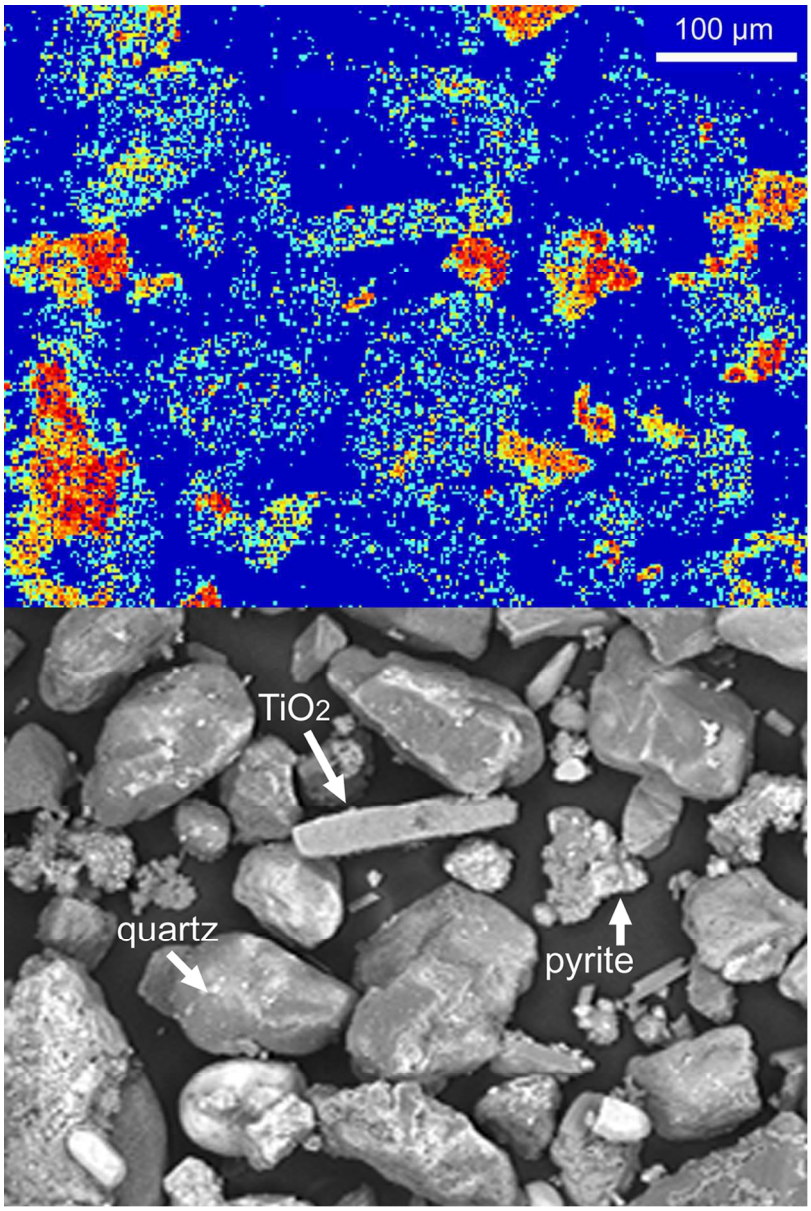


74x74mm (300 x 300 DPI)

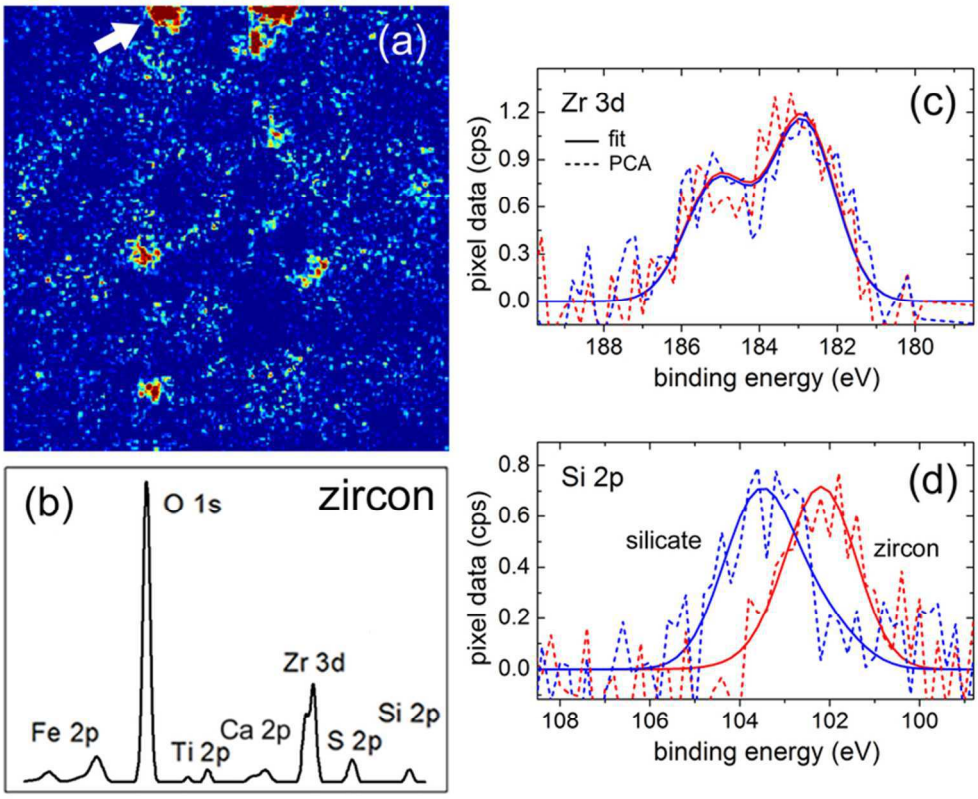


Varimax Factor Loading

85x44mm (300 x 300 DPI)



97x145mm (300 x 300 DPI)



70x58mm (300 x 300 DPI)