Principal Component Analysis for STEM EDS images: Optimal collection parameters

Isac Nissim Lazar

Bachelor Thesis

Supervisor: Carl Martin Ek Rosén

Abstract

STEM EDS is a common technique used for compositional mapping in materials. The vasts amounts of raw data produced by STEM EDS is suitable for advanced analysis using methods in unsupervised machine learning. One powerful method is Principal Component Analysis (PCA) which can automatically discover significant chemical correlations in a sample. Although more powerful than classical analysis, there are limits to how weak chemical correlations PCA can detect. Faced with a limited data collection time, it is of great importance to know how data collection parameters in the STEM affects PCA performance. In this work, the effect of image size and X-ray counts has been investigated using simulated STEM EDS datasets. It was found that decreasing overall image size while increasing counts per pixel both increases the chances of discovering weak chemical correlations and improves the overall PCA accuracy. Furthermore, use of varimax factor rotations has been investigated as a way to improve interpretability of PCA results. Finally, a real sample originating from a tungsten-carbide cutting tool which had been turned in a titanium alloy was analysed using the presented methods. This analysis confirmed the existence of an expected pure tungsten phase but also lead to the discovery of an unexpected ridge-like structure in the adhered alloy material.

Contents

1 Introduction

1.1 Motivation and main objectives

So-called Energy Dispersive X-ray Spectroscopy (EDS) mapping in the Scanning Transmission Electron Microscope is a method that can be used to identify and quantify different chemical species in a sample [1]. The raw data obtained from EDS-maps, called spectrum images, is suitable for advanced analysis using algorithms in multivariate statistics and machine learning, which has previously been demonstrated [2][3]. These algorithms take advantage of the large redundancy of information contained in the

spectrum images and enable automatic identification of chemical species.

Although these algorithms improve the analysis of data, there are limits to how weak signals that can actually be identified and discerned. In most articles utilizing these algorithms for spectrum images, there is often no motivation behind the choice of collection parameters, such as ratio of pixels to total X-ray counts when collecting the EDS map in the microscope. Limiting values for the accuracy of a certain type of algorithm, Principal Component Analysis (PCA), and the relation to samples size, has recently been theoretically explored in [4]. The overall goal of this work is to find optimal collection parameters when using PCA for distinguishing chemical phases in EDS-maps. More specifically, the optimal relationships between sample size and total X-ray counts will be explored.

The method of EDS-mapping is used in a wide range of materials. Here, the focus will be on metal machining tools and their interaction with a workpiece material. Signals generated in particular from a sample consisting of a piece of used cemented carbide cutting tool (WC grains bonded together with Co) turned in a titanium alloy (Ti6Al4V). When using such a cutting tool, high temperatures arise which in turn lead to diffusion between the tool surface and the working material. New chemical species are formed at the contact interface which affect the service life of the tool, but the mechanism is still not fully understood. To optimize service life, an understanding of what and where new phases occur is required. Previous studies has found evidence of carbon depletion from the WC grains that leads to the new phases bcc-W and TiC being formed at the interface between grain and adhered Ti6Al4V material [5][6][7]. Present in the W-C phase diagram, is a W_2C inter-metallic phase [8]. The question arises then, is this phase also present in between the WC grain and the newly discovered bcc-W phase, and if so, can it be detected? In this work, it is explored under which conditions these new phases can be detected by the use of PCA.

2 Theory and background

2.1 Spectrum imaging

In a scanning transmission electron microscope, the electron beam is focused down to a small probe. Deflection coils then raster the electron beam across a thin sample. An electron detector sitting underneath the sample collects transmitted electrons into a set of pixels that build up an image. Sample composi-

tion and density are factors that affect the amount of transmitted electrons and are major sources of contrast in the electron image. This is the main signal output in a STEM [9].

A secondary signal that can be detected and which this work focuses on, are characteristic X-rays. The electrons in the beam have energies (100 keV-300 keV) that are well above of what is required to eject electrons in the core shells of the sample atoms. The atomic relaxation occurring after an ionisation event sends out an X-ray photon with an energy characteristic of the particular atomic species and electron transition. A single atomic species can send out photons of a couple of different energies, but the relative ratio of photons of different energies emitted is directly related to the probability of a specific atomic transition occurring. This makes a specific atomic species recognisable by it's X-ray energy "footprint". A distinct chemical species, that has a fixed ratio between its constituent atomic species, also has a recognisable X-ray footprint.

An X-ray spectrometer can detect and discern X-ray photons of different energies. A voltage is generated when a photon hits the detector. This voltage, which is proportional to the photon energy is then digitized and saved as a discrete value, called a *count* in digital bins called energy channels. Each energy channel will record counts from all photons within a small and specific energy interval. The range of energies recorded in a single energy channel is variable and is called the energy resolution. The number of channels is also variable but commonly set to multiples of 2^n. The histogram of collected photon counts versus energy channel is then presented in a *spectrum*. Simulated X-ray spectrum from a couple of different chemical species can be seen in figure 1. They contain 2048 energy channels with a resolution of 6 eV/channel. Due to various electronic noise, the sharp energy lines arising from discrete transitions in the atom, become smeared out over several energy channels in the collected spectrum, and a particular transition will show up as a Gaussian peak. In the end, we are talking about collections of characteristic *peaks* for different atomic species. These are named according to the characteristic element and electron transition. For example, C Kα is an X-ray energy that corresponds to the electron transition from the L shell to the K shell in a Carbon atom. A spectrum generated from a specific chemical species, such as WC, will contain peaks characteristic for both Carbon as well as Tungsten.

As such by putting an X-ray spectrometer inside the

STEM above the sample, an X-ray spectrum can be collected and saved for every (x,y) position that is scanned by the electron beam. The longer the electron beam stays in the same (x,y) position, its *dwell time*, the more characteristic X-ray photons are generated and collected as *counts*. Then, the raw data is organized into a 3-dimensional "data cube", where a spectrum is saved in each pixel[9]. An analogy is a digital RGB image where each pixel has three color channels: red, green and blue. In the case of X-ray spectrum, we are dealing with often 2048 "color" (energy) channels.

This data cube contains vasts amount of data. For example, an image with 512 by 512 pixels with a detector set for 2048 channels will generate a data cube with 512x512x2048 = 536 870 912 data points! The assumption in all multivariate statistical analysis methods is that far from all these data points are independent variables [10]. This is logical since we know that the spectrum collected in each pixel arises from characteristic peaks from only a few chemical species. Thus most of the image can instead be described by linear combinations of only a few independent components. The goal of MSA methods is to find these independent components. Going back to the RGB image analogy: imagine an image that consists only of varying amounts of the colors yellow and pink. Instead of describing each pixel as a combination of the three variables red, green and blue, we can describe the image in another basis with only two basis vectors: yellow and pink. These new basis vectors are linear combinations of red green and blue. The hope is that these new basis vectors represent something physical. In the following discussion, these new basis vectors will be called component vectors.

One thing to note going forward is that the methods analysed in this work does not take into consideration the relative position of each pixel. Instead, all pixels are treated as m measurements of n channels (or variables). Thus, the 3-dimensional data cube can instead be flattened and visualised as an m x n data matrix D, where each row is a measurement/sample of n channels.

2.2 Principal component analysis and Singular Value Decomposition

2.2.1 Linear model of data

As previously hinted, an underlying assumption of the EDS data is that the intensity of a characteristic X-ray peak is directly proportional to the abundance of the element from which it originates. Taking this

concept one step higher: the intensity of a collection of characteristic peaks (here called source spectra) originating from a distinct chemical species is also directly proportional to the abundance of that species in the sample. It is therefore logical to describe the spectrum in each pixel as some linear combination of a few source spectra. To clarify, source spectra as captured by a spectrometer are simply vectors containing the values of each energy channel. The task at hand is then to find these vectors corresponding to source spectra and their linear combination in each pixel from a dataset. Unfortunately, even though a dataset was originally created from source spectra, there exists an infinite amount of ways to describe the same dataset with linear combinations of other vectors that does not represent the source spectra. In other words, there is no unique representation of the data. What is assumed in the principal component analysis following is a spiked covariance model used by Nadler [11]. The implications of it was tested experimentally to fit well to spectrum imaging by Potapov [12]. In this model, each observed sample vector $\mathbf{x}$ originates from linear combinations of q *components* plus a noise term ϵ

$$\mathbf{x} = \sum_{j=1}^{q} u_j \mathbf{v}_j^T + \boldsymbol{\epsilon}^T \tag{1}$$

All $\{u_j\}_{j=1}^{q}$ component weights are random variables and uncorrelated to each other. All $\{\mathbf{v}_j\}_{j=1}^{q}$ component vectors are orthogonal to each other. The noise term $\boldsymbol{\epsilon}$ is a multivariate Gaussian noise vector with zero mean and fixed variance σ^2, i.e. the noise is assumed to be homoscedastic. The model's assumption of orthogonality between component vectors does not make much sense physically applied to spectrum images, where orthogonality between source spectra is in no way logical to suspect[3]. What it provides, is a unique way to linearly describe a dataset. For any given dataset there is only one set of component vectors and weights that fulfill these constraints.

2.2.2 Principal component decomposition

Principal component analysis is a factor analysis method that segments the data matrix D into the product of a m x n matrix T and a n x n matrix P as such

$$D = TP^T \tag{2}$$

[3]. In other words, each row (the spectrum at pixel i) $\mathbf{s}_i$ in the data matrix D is segmented into the sum

$$\mathbf{s}_i = \sum_{j=1}^{n} t_j \mathbf{p}_j^T \tag{3}$$

Where $\mathbf{p}_j$ is a n x 1 -dimensional *component* vector representing a column in P and t_{ij} is the abundance or weighting of that component in pixel i. Equation 3 looks very similar to equation 1, differing only in the missing noise term and the fact that usually $n >> q$. The goal is to find the underlying q principal components and to accomplish this PCA enforces the same constraints on the extracted components as in the spiked covariance model. Each column in T and P have to be orthogonal and orthonormal respectively. This has the added effect of making the decomposition unique. Furthermore, PCA orders the rows and and columns in T and P so that the first component explains most of the variance in D, the second component explains the second most variance and so on. Ideally then, the first q extracted components will include true signal variance and the rest $n - q$ will only explain the noise variance σ^2 and can be discarded from analysis. For the purpose of dimensionality and possibly noise reduction, a lower rank estimation of D can be constructed as

$$\mathrm{T}_p \mathrm{P}_p^T = \mathrm{D}_p \approx \mathrm{D} \tag{4}$$

Where subscript p denotes that only the first p columns in the matrices are kept. One final property of the PCA decomposition is that an estimation with the p first components in equation (4) is the best p-rank estimation of the data matrix that exists [3]. This implies that unless the mean has been subtracted from D before decomposition, the first extracted component will always represent the mean, since that is the rank 1 best estimation of D.

Imagine a completely noise-free (i.e, $\sigma = 0$) dataset, with $m >> 1$ samples that can theoretically be described using one component $(u, \mathbf{v})$ in the spiked covariance model. We assume that the first component in a PCA decomposition of this dataset is equal to v

$$\mathbf{p}_1 \equiv \mathbf{v} \tag{5}$$

And call it the *true* component. The reason for defining the spiked covariance model is now apparent. If we now add noise to the dataset and perform a PCA decomposition, we obtain the first component vector $\mathbf{p}_1^*$. Since we know it is unique, we can directly interpret it as a noisy estimate of $\mathbf{v}$.

2.2.3 Singular Value Decomposition and PCA implementation

One way to calculate the matrices T and P under the PCA constraints is by Singular Value Decomposition. This is an algorithm that decomposes D into three matrices

$$\mathrm{D} = \mathrm{U}\Sigma\mathrm{V}^T \tag{6}$$

The PCA matrices T and P can then simply be obtained as

$$\begin{aligned} \mathrm{T} &= \mathrm{U}\Sigma \\ \mathrm{P} &= \mathrm{V} \end{aligned} \tag{7}$$

U is called a left singular vector with dimensions m x n, V is a right singular vector with dimensions n x n. Both U and V are orthogonal. Σ is a diagonal matrix containing the singular values of D ordered after significance. That is, the singular values are the square root of the eigenvalues of $\mathrm{D}^T\mathrm{D}$ ordered in a descending order. V contains the corresponding eigenvectors of $\mathrm{D}^T\mathrm{D}$ [3]. The j:th largest eigenvalue of a noise-free dataset divided by the number of samples m will be denoted λ_j and corresponds to the underlying component signal variance of component j

$$\lambda_j = \frac{j\text{:th largest eigenvalue of } \mathrm{D}^T\mathrm{D}}{m} \tag{8}$$

This definition is appropriate for spectrum imaging, as it makes the explained variances λ independent on sample density (the number of samples m per scanned area) [13].

2.2.4 To subtract mean or not?

In general PCA literature, one is often advised to subtract the mean from all samples as a pre-processing step. This is called mean centering. In this case the mean would be the sum spectra divided by the number of pixels. It is advised against specifically in the case of spectrum imaging by Keenan et al.[10][3], and has been shown to not make a difference in the fit to the data. Although it is argued to improve the "interpretability" of the obtained components according to Potapov [13]. In this work, no mean centering will be performed. Instead varimax factor rotations will be used to enhance interpretability of obtained components.

2.2.5 Varimax factor rotation

As mentioned, the assumption of orthogonality between the extracted component vectors is not based on any real properties of X-ray spectrum of different chemical species, making the PCA extracted components rather abstract [3]. For example, a component vector and/or its corresponding weight may contain negative values, which would correspond to negative intensities, having no real physical interpretation. One way to bring the extracted components closer to physical reality is by the use of factor rotations [10][14][15]. These are based on the fact that,

given an invertible rotation matrix R, the extracted PCA matrices can be rotated as such

$$D = TP^T = T(RR^{-1})P^T = (TR)(R^{-1}P^T) = \ddot{T}\ddot{P}^T$$
(9)

While still providing the same fit to the data. Using a rotation matrix R, we can rotate the basis vectors in either the spatial domain (column vectors in T) or the spectral domain (column vectors in P) to "simpler structures". In practice, this can be viewed as enhancing contrast in one one of the domains. In this work, a varimax rotation is performed in the spatial domain. This basically assumes that the signal in a given pixel s_i, should be described by a linear combination of as few component vectors as possible. Imagine an orthogonal rotation matrix R. Given the left singular vector U, varimax tries to find the R that maximizes the variance of the squares of the elements in the matrix product UR. The new rotated factor matrices are obtained as

$$\ddot{T} = UR$$
$$\ddot{P} = V\Sigma R$$
(10)

Besides increasing contrast in the spatial domain, this also relaxes the orthognality constraint in the spectral domain. In other words, the columns in $\ddot{P}$, the rotated component vectors, no longer have to be orthogonal, hopefully now more resembling real spectra.

2.2.6   Theoretical PCA threshold

In unsupervised machine learning and PCA literature, one finds the notion of a phase transition, or "threshold" where the fraction of samples to measured variables needs to exceed a certain value in order to be able to find a component in a noisy dataset[16][17][18]. In the spiked covariance model used by Nadler, it was found that for PCA to be able to detect a component in a one-component noisy dataset even remotely, the following inequality has to be fulfilled.

$$\left(\frac{\sigma^2}{\lambda}\right)^2 < \frac{m}{n}$$
(11)

In other words, the term $\frac{m}{n}$ has to be bigger than the square of the ratio of noise variance σ^2 to component signal variance λ. We can approximate the absolute noise variance σ^2 in a dataset as

$$\sigma^2 = \frac{1}{mn}\sum_{i,j}^{m,n}(s_{ij} - s^{*r}_{ij})^2 \approx$$
$$\frac{1}{mn}\sum_{i,j}^{m,n}\text{Var}(s^*_{ij}) = \frac{1}{mn}\sum_{i,j}^{m,n}s_{ij}$$
(12)

[13] Where s^{*r}_{ij} is one realisation of the Poisson random variable s^*_{ij} in pixel i and energy channel j. The last equality follows from the fact that the variance of a Poisson random variable is equal to its expected value. The noise level σ^2 cannot be calculated on experimental data in this way since s_{ij} is not known. This is one advantage of using simulated datasets where each value s_{ij} is known.

Additionally, if we define the PCA extracted component vector from the noisy dataset as $\mathbf{v}^*$, and its dot product with the true component vector $\mathbf{v}$ as

$$R^2 = \frac{\mathbf{v}^*\mathbf{v}^T}{||\mathbf{v}^*||||\mathbf{v}||}$$
(13)

Nadler showed that, given the inequality in equation 11 is satisfied

$$R^2 = \frac{\frac{m}{n}\left(\frac{\lambda}{\sigma^2}\right)^2 - 1}{\frac{m}{n}\left(\frac{\lambda}{\sigma^2}\right)^2 + \frac{\lambda}{\sigma^2}}$$
(14)

If the inequality is not satisfied, the expected value of R^2 is zero, meaning no information of the true vector $\mathbf{v}$ is expected to exist in $\mathbf{v}^*$. This theory will be tested in the simulations following.

2.3   Poisson noise and scaling

The usual definition of noise in hyperspectral PCA literature assumes homoscedasticity[19][12][4]. This implies a noise term ϵ with constant variance σ^2 over all energy channels, as what is assumed in the linear model in equation 1. In reality, measurements involving counting events are corrupted by noise of a Poissonian nature [20]. This means that the variance of a random variable is equal to the expected value of that same variable. And so the assumption of homoscedasticity does not hold. In their article, Keenan and Kotula propose a solution for this problem by a transformation of the measured data into a space where the noise is approximately homoscedastic [20]. Subsequent PCA can then be performed under the assumption of homoscedastic noise. The derivation of the transformation is lengthy, but in short utilizes an estimate of the random variable variance as simply the mean of the sampled variable. The transformation can be expressed as

$$\widetilde{D} = G^{-1/2}DH^{-1/2}$$
(15)

Where D is the original unfolded mxn data matrix with m pixels and n energy channels. G is a mxm matrix with a mx1 vector $\mathbf{d}_m$ along its diagonal. $\mathbf{d}_m$ is obtained first by summing all counts in each spectrum, and then dividing the elements by the number

of energy channels n. Similarly, H is a nxn matrix diagonalised by $\mathbf{d}_n$ that consists of the sums of all counts in each energy channel over all measurements divided by m. After PCA is performed on the scaled dataset, the obtained factors $\widetilde{T}$ and $\widetilde{P}$ can be inversely scaled back to the original space as

$$T = G^{1/2}\widetilde{T}$$
$$P = \widetilde{P}H^{1/2} \tag{16}$$

The matrices T and P does not technically have orthogonal columns anymore but they "can be considered principal components in the sense that they provide a basis that maximally accounts for the chemically related variation in the data" [20]. For subsequent varimax rotation, the factors can be reorthogonalised first according to an fPCA algorithm presented by Keenan [3]. Alternatively, another SVD can simply be performed on the dataset reconstructed with p components

$$T_p P_p^T = D_p = U\Sigma V^T \tag{17}$$

3 Methods

3.1 Spectrum and spectrum image simulation

In order to assess the accuracy of the presented analysis methods, one must know the exact composition of a sample. Since this is very hard to obtain experimentally, simulated samples are better suited. Using the free microanalysis software DTSA-II [21], Monte-Carlo simulations of interactions between high energy electrons and various materials was used to obtain characteristic X-ray spectra. These spectra were then linearly combined to build noise-free data sets according to various models. To simulate the conditions in a STEM, the electron beam was set to 300 keV. Each characteristic spectrum was obtained by selecting a Monte Carlo simulation of a 100 nm thick and 1000 um wide block of the specified material on a substrate of None (=vacuum) in the software. The total electron dose, defined as time · current was set to 0.1 nAs. All available X-ray generation modes were selected: Characteristic primary, Characteristic secondary, Brehmstrahlung primary and Brehmstrahlung secondary. To detect X-rays generated in the MC-simulation, a simulated SDD detector with a 10 mm² detector area and 2048 energy channels was used. For every characteristic spectra a few different resolutions were used. 5, 6, 10 and 20 eV/channel. In table 1, the densities specified for the MC simulations are presented. The resulting noise-free spectra using

Table 1: Materials densities used in spectrum simulations

Material	Density (g/cm^3)
TiC	4.95 [22]
Ti6Al4V	4.43 [23]
W	19.26 [24]
WC	15.65 [25]
W$_2$C	17.23 [26]

6 eV/ch can be seen in figure 1. These will be referred to as source spectra. The values in the energy channels are not integers and instead interpreted as the expected number of counts, as opposed to actual number of counts, and these values scale linearly with electron dose.

3.1.1 Two component gradient model

This model simulates a sample were the composition changes from chemical species s1 to s2 linearly along the x-axis. An image grid m with x x y pixels, given any source spectra arrays s1 and s2, the model is generated as

```
for row in range(1, y):
 for column in range(1 to x):
  k = column/x
  m[row, column] = k*s1 + (1-k)*s2
```

A visualisation of the generated model is presented in figure 2.

3.1.2 Cutting tool interface model

A model of the interface between the used cutting tool with adhered alloy material can be seen in figure 3. The theorised W$_2$C species increases gradually from WC over the span of around 30 nm.

3.1.3 Counts and adding noise

Each noise-free dataset was first scaled by some constant so that the mean counts per pixel was equal to one, but keeping the relative differences between pixels. Datasets were then scaled to the desired mean counts per pixel $\bar{c}$. Noise was then added using a Poisson random number generator. If the value of channel j in pixel i is denoted s_{ij} in the noise-free data set, then the corresponding value s_{ij}^* in the noisy dataset is drawn from a Poisson distribution with parameter s_{ij} as such

$$s_{ij}^* \sim Po(s_{ij}) \tag{18}$$

A realisation of this using a spectrum of only WC can be seen in figure 4b.
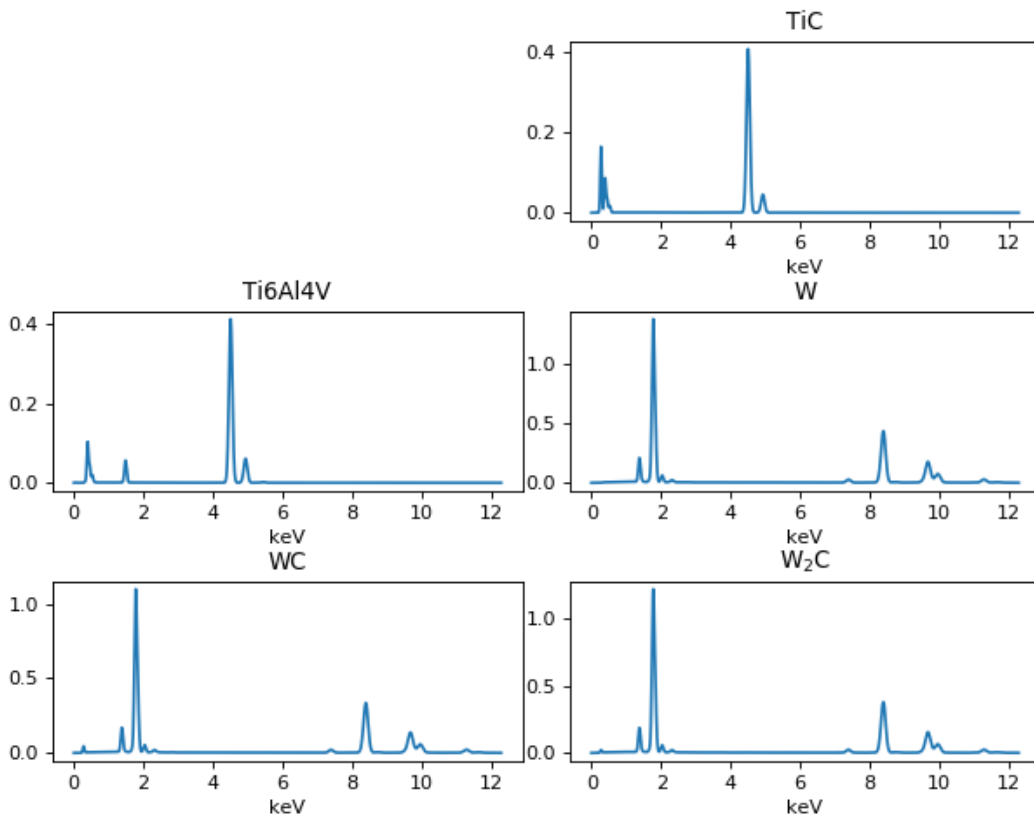
6

Figure 1: Noise free spectra generated by Monte-Carlo simulations in the DTSA-II software. The y axis can be interpreted as the expected number of counts. The detector is configured for 6 eV/ch.

4 Results and Discussion
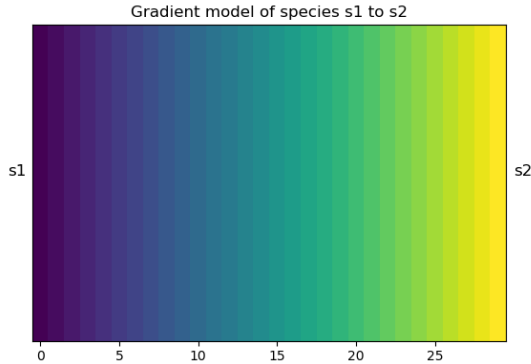
4.1 Two component system simulations. WC and W_2C

Figure 2: Gradient model. The source spectra s1 increases linearly as a function of x-position from right to left. The opposite is true for source spectra s2.
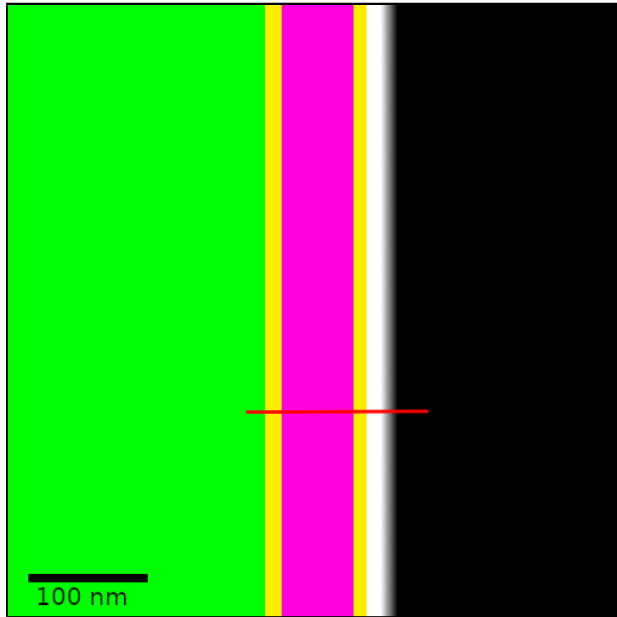

Figure 3: Model of used cutting tool interface with adhered Titanium alloy. Ti6Al4V (Green), TiC (yellow), W (pink), W_2C (white), WC (black). Red line shows where a linescan was sampled.

Because the two spectra from WC and W_2C are very similar and may be the most difficult to distinguish in classical analysis, a model built with these two spectra are of great interest to analyse. A particularly low variance model is constructed with W_2C to the left gradually changing to WC to the right according to the two component gradient model described in Methods. This could for example simulate the signal coming from an area where a WC and W_2C grain lies on top of each other with varying relative thicknesses. The model consists of 10x50 = 500 pixels and the results of PCA on a noise-free dataset generated from this model can be seen in figure 5.

It is clear that the first principal component simply represents the mean spectrum. It also has a very large explained variance of $\lambda_1 = 3.26 \cdot 10^6$. Component 2 essentially pronounces the differences between each spectrum to the mean and it has an explained variance of only $\lambda_2 = 85.3$. The carbon C K_α peak is clearly identified and positive. Looking at the component 2 weighting map shows that the carbon content is correctly shown to gradually increase closer to the WC phase and decrease closer to W_2C. 4 of the 5 tungsten peaks are negative implying anticorrelation to carbon. They increase gradually towards W_2C. Interestingly, the W M_β seems to correlate with carbon content. This may be an effect arising from WC having lower density than W_2C. Density affects the X-ray intensity in that higher density materials absorb more low energy X-rays before they are able to escape from the material. This may therefore be the best fit to explain this effect.

4.1.1 Pixels versus mean counts-per-pixel

Next, tests were performed on a WC-to-W_2C model with added Poisson noise. To directly test for a given total counts C, whether it is better to sample a few pixels m, with high mean counts per pixel $\bar{c}$ or vice versa, a batch test was performed. Initially the gradient model was used to create the dataset with WC and W_2C as before but with dimension 10x50000 pixels. m samples was then randomly selected from the initial dataset and noise was subsequently added. Total counts was always $C = 10^7$ but divided on the m samples. So a large number of samples m means a low mean counts-per-pixel $\bar{c}$ and vice versa. This procedure is thought to simulate varying the number of pixels per physical area selected in the microscope
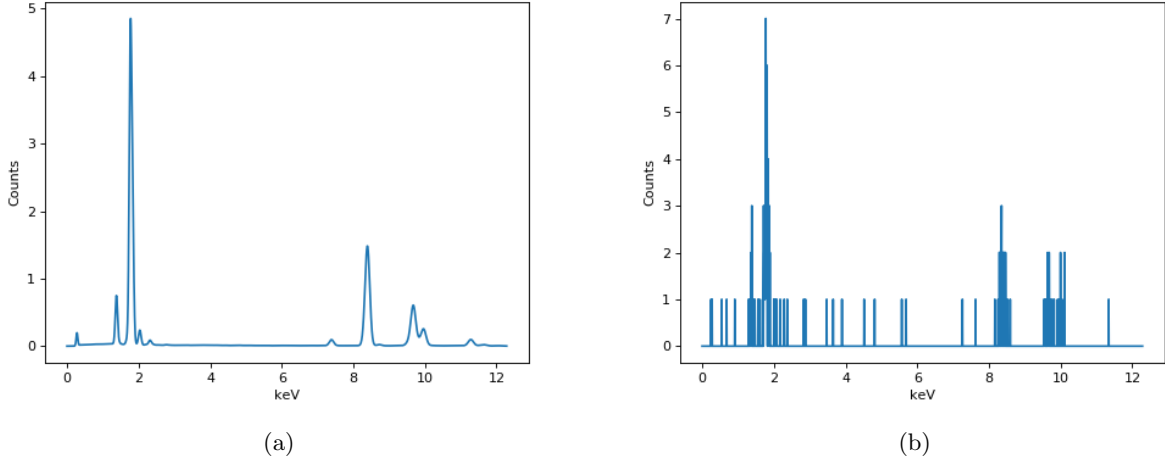
Figure 4: 200 count WC spectrum a) Pure WC spectrum, b) WC spectrum with added Poisson noise

or the effect of binning post acquisition. Binning is an operation where the values of all energy channels of adjacent pixels are added together, effectively increasing the mean count per pixel but decreasing spatial resolution. For every dataset, PCA was performed and the accuracy R^2 of the extracted noisy component $\mathbf{v}_2^*$ to its equivalent true component $\mathbf{v}_2$ was recorded.

The simulation results can be seen in figure 6. It is observed that increasing the number of samples m while reducing the mean counts-per-pixel reduces the accuracy R^2 of the extracted principal component vector $\mathbf{v}_2^*$. This also approximately follows the theoretical values calculated with equation 14. Whats more, the predicted threshold in equation 11 seems to predict the observations very well. All datapoints with values below the threshold, seems to have components vectors that are almost orthogonal to the true ones. Looking at a datapoint well above the threshold, at $m = 1000, \bar{c} = 9991$ as seen to the lower left in figure 6. This extracted component seems to have a well defined carbon peak as in the pure component $\mathbf{v}_2$ in figure 5, as well as peaks representing W M and L peaks. In contrast, the extracted component vector for a datapoint below the threshold has no pronounced carbon peak. Additionally, an equal amount of positive and negative spikes in close energy proximity can be seen were the W peaks are expected. This isotropic appearance may be a clue for when a component vector only represents noise. Indeed, the third component extracted from this noisy dataset which can be seen in figure 7, which should only represent noise, since it is extracted from a 2 component system, has spikes with generally no preferred direc-

tion.

An attempt to explain the observed relationship between m and $\bar{c}$ now follows.

First an investigation of the noise level as defined in equation 12 is required. With this definition the noise level of an dataset will increase linearly with counts. If we multiply the pure dataset with a constant $\bar{c}$ to increase the mean count per pixel, and then add Poisson noise, mathematically, $\tilde{s}_{ij}^* \sim Po(\bar{c}s_{ij})$, then

$$\tilde{\sigma}^2 = \frac{1}{mn}\sum_{i,j}^{m,n}\text{Var}(\tilde{s}_{ij}^*) = \frac{1}{mn}\sum_{i,j}^{m,n}\bar{c}s_{ij} \propto \bar{c} \qquad (19)$$

This relationship was also confirmed experimentally.

Another way to define the true eigenvalue of the $j:th$ component can be defined as

$$\lambda_j = \text{Var}(u_j)||\mathbf{v}_j||^2 \qquad (20)$$

referring back to the spiked covariance model [27][11]. Here $\text{Var}(u_j)$ signifies how component vector $\mathbf{v}_j$ varies across the sample. Consider scaling a component with $\bar{c}$, by multiplying with each element in $\mathbf{v}_j$

$$\tilde{\lambda}_j = \text{Var}(u_j)||\bar{c}\mathbf{v}_j||^2$$
$$= \text{Var}(u)\bar{c}^2||\mathbf{v}_j||^2 \propto \bar{c}^2 \qquad (21)$$

Referring back to the threshold equation 11 and the theoretical estimated dot product R^2 in equation 14, we see that λ and σ^2 always appear together as a fraction of each other $\frac{\lambda}{\sigma^2}$. From equation 19 we know that the noise level scales linearly with counts, $\sigma^2 \propto \bar{c}$
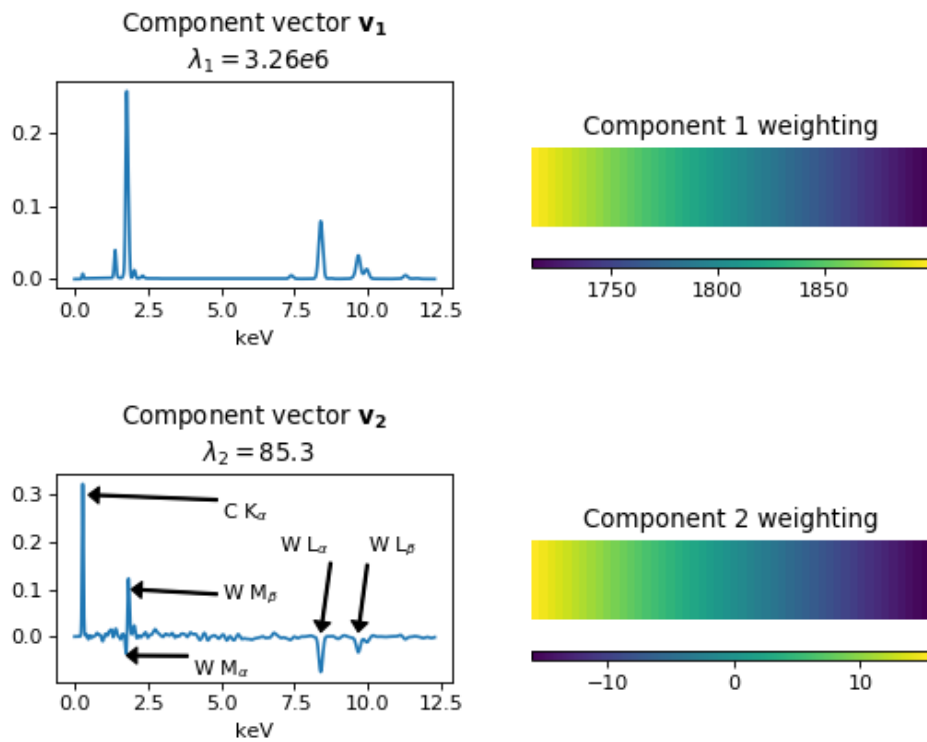
9

Figure 5: Component vectors (left column) and respective weightings (right column) for the simulated noise-free W_2Cc-to-WC dataset.
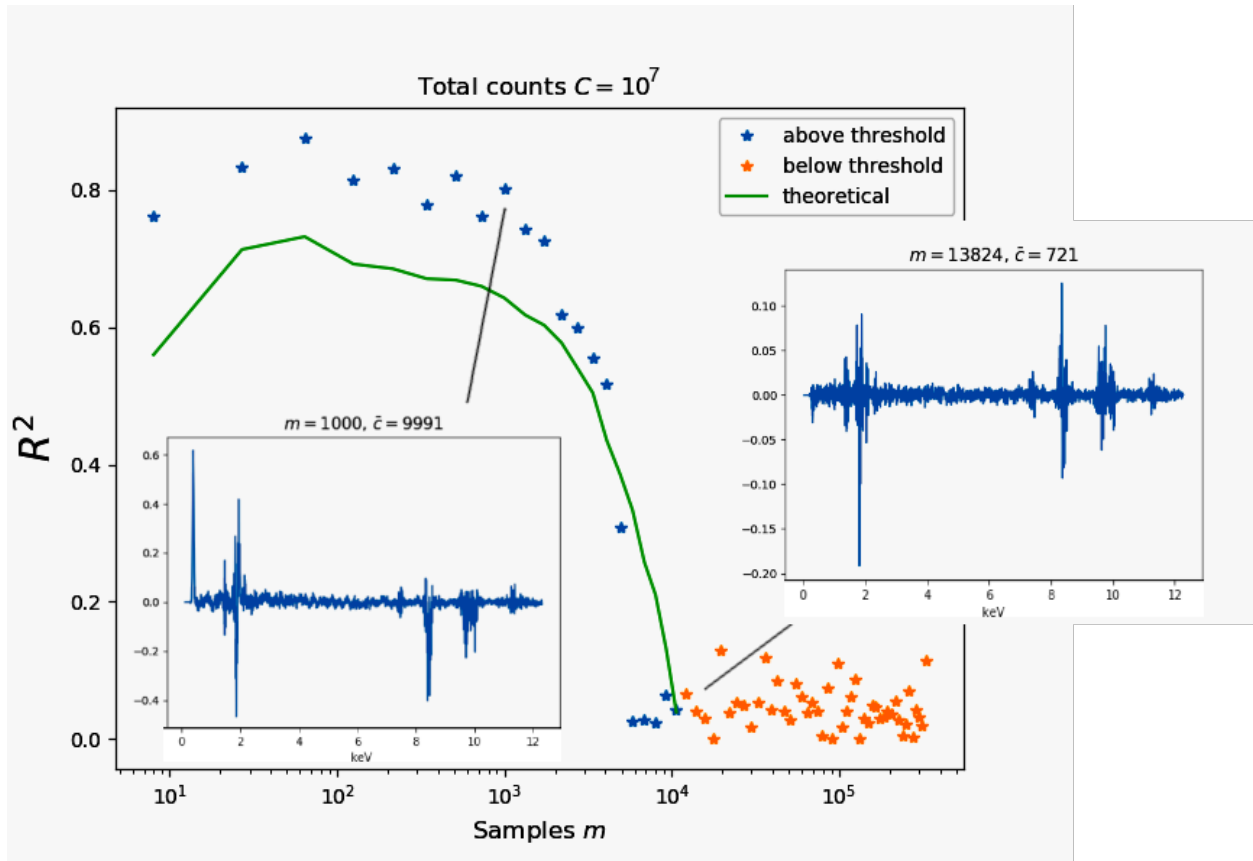
Figure 6: Batch test of the dot product between the extracted noisy component vector $\mathbf{v}_2^*$ to the true component vector $\mathbf{v}_2$ in the W_2C-to-WC model. Total counts was at a constant 10 million while varying the the sample size. The plot can also be viewed as increasing the mean counts per pixel $\bar{c}$ going from right to left along the x-axis. An example of an extracted component at above the threshold at $m = 1000$ is displayed. A clear carbon peak can be seen. In the example of an extracted component vector below the threshold at $m = 1324$ no distinguished carbon peak can be seen. The spikes present in the components below the threshold have an isotropic appearance with almost equal positive and negative amplitudes.

11
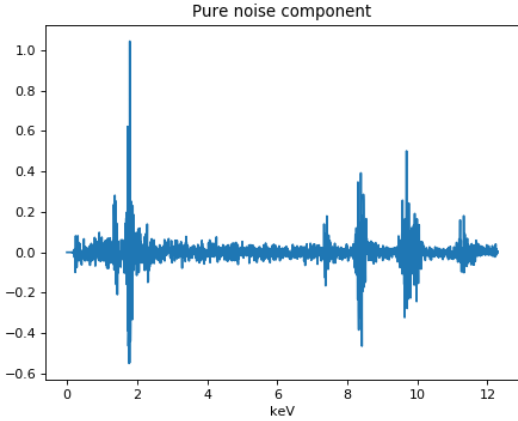
Figure 7: A component vector extracted from the simulated W_2C-to-WC model that contains only noise. Spikes with isotropic appearance can be seen.

these parameters was generated for further investigation. PCA was performed before adding noise to extract values for σ^2 and λ and these noise-free results have been presented earlier in figure 5. Results of PCA on the noisy dataset can be seen in 8. The noise level was calculated to $\sigma^2 = 9.23$ according to equation 12. The components have been reorthogonalised as described in section 2.3. Component 1 looks unchanged compared to the true component $\mathbf{v}_1$ in figure 5, whereas component 2 can clearly be viewed as a noisy estimation of $\mathbf{v}_2$ and its weighting. It is observed that the explained variance of component 2 in the noisy dataset has increased to $\lambda_2^* = 120$ due to the added noise variance, approximately following the prediction of [11]

$$\lambda_2^* = \left(1 + \frac{n}{m}\frac{\sigma^2}{\lambda_2}\right)(\lambda_2 + \sigma^2) = 136.4$$

.

The threshold in equation 11 is satisfied and calculating the predicted discrepancy between noisy and pure component as

$$R_2^2 = \frac{\frac{m}{n}\left(\frac{\lambda_2}{\sigma^2}\right)^2 - 1}{\frac{m}{n}\left(\frac{\lambda_2}{\sigma^2}\right)^2 + \frac{\lambda_2}{\sigma^2}} = 0.658$$

yields a bit lower estimate than the real measured value of

$$\frac{\mathbf{v}_2^* \mathbf{v}_2^T}{||\mathbf{v}_2^*||||\mathbf{v}_2||} = 0.783$$

and so the fraction of noise to component variance will always scale according to

$$\frac{\sigma^2}{\lambda} \propto \frac{\bar{c}}{\bar{c}^2} = \frac{1}{\bar{c}} \tag{22}$$

Extending this, given total counts of $C = m\bar{c}$, assuming $\sigma^2 = k_\sigma \bar{c}$, and $\lambda = k_\lambda \bar{c}^2$ with some constants (k_σ, k_λ), the threshold equation 11 can be rewritten as

$$\frac{k_\sigma}{k_\lambda}\left(\frac{m}{C}\right)^2 < \frac{m}{n} \tag{23}$$

The two terms in this equation equal each other at some critical value m_{crit} in where an increase in m will always result in being below the threshold. Additionally, the theoretical dot product in equation 14 can be rewritten as

$$R^2 = \frac{\left(\frac{k_\lambda}{k_\sigma}\right)^2 \frac{C}{n} - \frac{m}{C}}{\left(\frac{k_\lambda}{k_\sigma}\right)^2 \frac{C}{n} + \frac{k_\lambda}{k_\sigma}} \tag{24}$$

If the total counts C is large enough, if $m \longrightarrow 0$ then $R^2 \longrightarrow 1$. This means, that for a given total counts C, increasing the number of pixels (and decreasing the count per pixel accordingly) will reduce the accuracy of the extracted component explaining the appearance of the plot in figure 6. This form of the equation also implies that the accuracy increases with total counts C.

From the results in the batch test, suitable values that ensure being above the threshold, in a WC-W_2C model dataset was determined to be 10 x 50 = 500 samples and 9.53 million total counts. A dataset with

In the hope of obtaining components representing real spectra, varimax factor rotation was performed in the spatial domain as described in section 2.2.5. The results of rotations on the pure and noisy PCA components can be seen in figure 9 and 10 respectively. The orthogonality constraint on the component vectors has been lifted and they all now represent something more similar to physical spectra. The weighting for the two components are now also comparable in amplitude. The expectation that the components would perfectly represent the two source spectra from WC and W_2C however is not met. The component vectors differ only slightly. If the components would have represented the source spectra, then the weighting of one component would always be zero in positions were the other reaches its maximum. Instead the pure WC spectrum present at the right edge is achieved by subtracting a small amount of component vector 1 from component vector 2.

It seems like in this case, the non-rotated PCA results, although abstract, are perhaps easier to draw conclusions from.
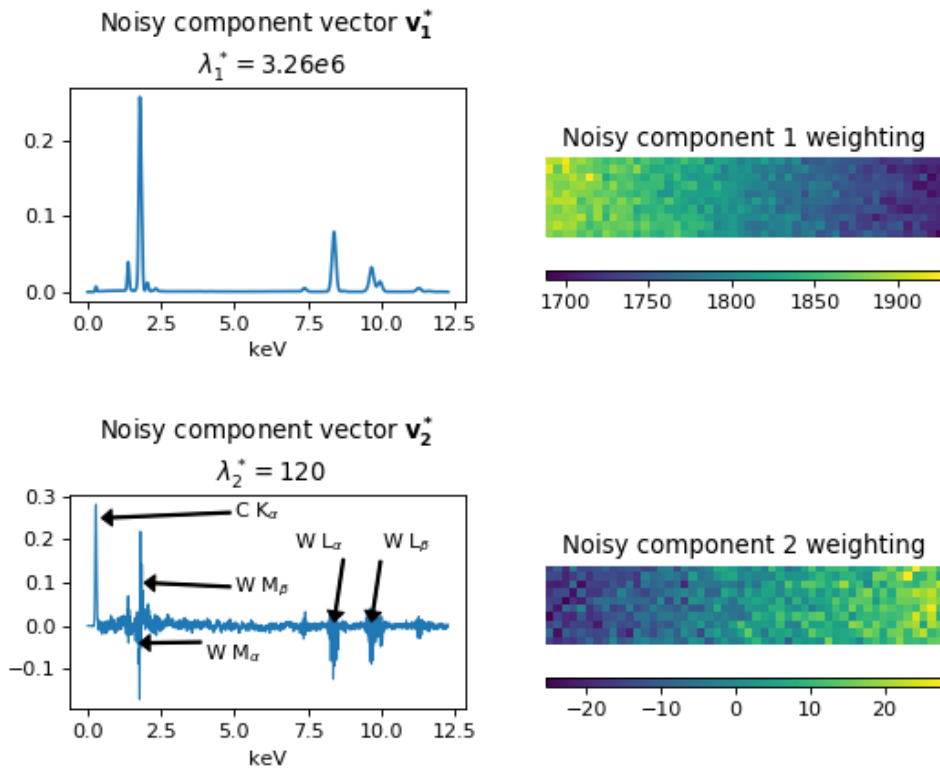
Figure 8: Component vectors (left column) and respective weightings (right column) for the simulated W_2Cc-to-WC dataset with added noise on a total of 9.53 million counts.
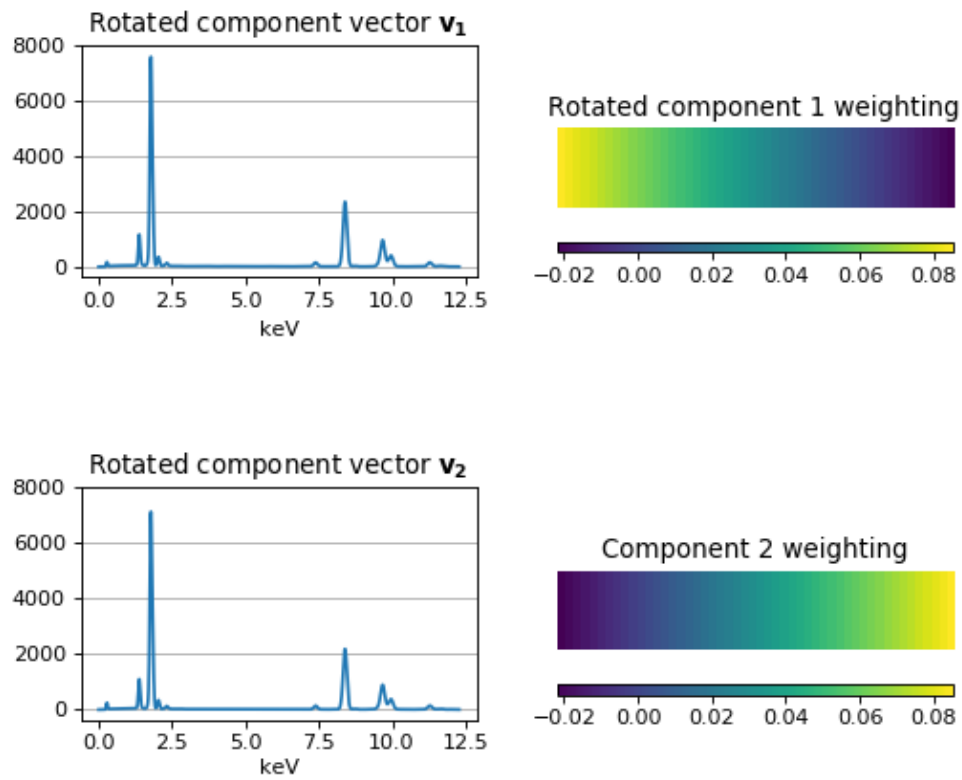
13

Figure 9: Rotated component vectors (left column) and weightings (right column for the simulated noise-free W_2C-to-WC dataset.
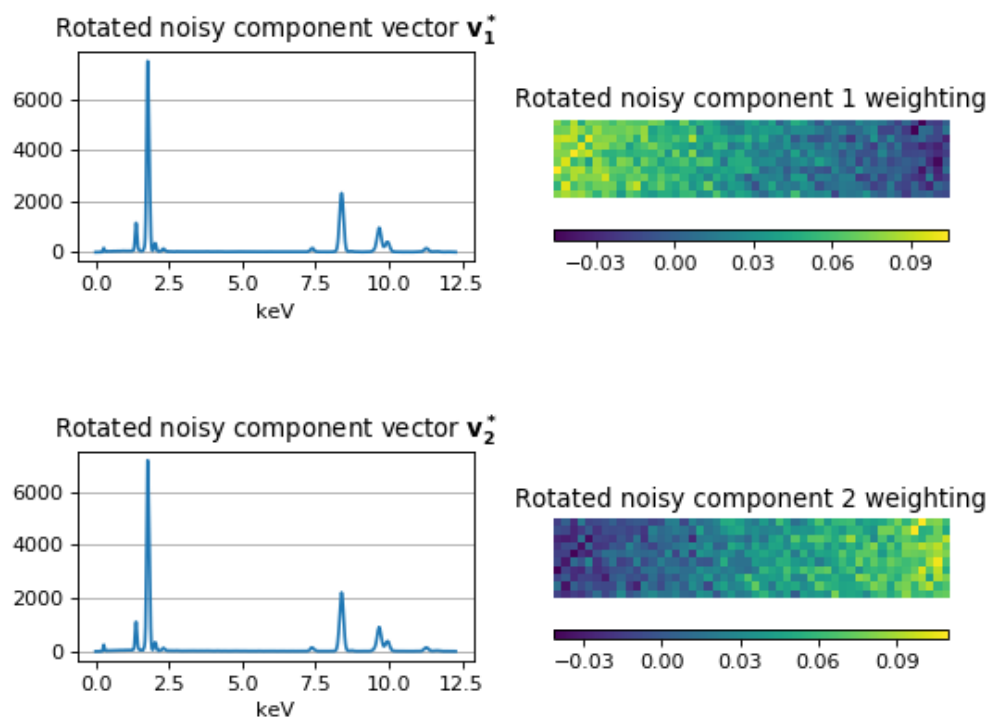
Figure 10: Rotated component vectors (left column) and weightings (right column for the simulated W_2C-to-WC dataset with added Poisson noise on 9.53 million total counts.

4.2 Multi component systems simulations

4.2.1 Interface

Since the interface model (figure 3) captures all the variance across the x-axis, and based on the results found in section 4.1.1 that counts per pixel should be preferred over sample size, a decision was made to only analyse this model by a linescan over the interface. In practice, this means a dataset consisting of only a 150 pixels wide line across the interface was analysed. This line is marked red in in figure 3. This line captures all variance in this model, while at the same time having the maximum amount of counts-per-pixel $\bar{c}$ possible for any chosen total counts C. PCA on the pure dataset is presented in figure 11. The first component exactly represents the W spectrum and has a easily interpretable weighting. The second component seems to represent the TiC and Ti6Al4V spectrum added together. It has a high weighting to the far left in the Ti6Al4V region, where carbon should not be present. This carbon is therefore cancelled by component 3. Component vector 3 has a large positive carbon peak and the weighting is negative in the Ti6Al4V region. This way of thinking about the components as adding and subtracting from each other is necessary and may eventually lead to the right conclusions about sample composition but is very time-consuming. Component 5 is a perfect example of when PCA yields a completely uninterpretable result even on noise-free data as a consequence of the orthogonality constraints.

Performing varimax rotation on the obtained factors produces remarkable results. The first three component vectors seem to exactly represent the three source spectra W, TiC and Ti6Al4V respectively as can be seen in figure 12a. Looking at their respective weightings in figure 12b, they are also mapped out spatially along the line exactly as the source spectra are in the underlying model. The last two components $\mathbf{v}_4$ and $\mathbf{v}_5$ exhibits the same behaviours as in the two-component system. Both of them are a mixture of the WC and W_2C source spectra differing only slightly. $\mathbf{v}_4$ has a slightly higher carbon peak than $\mathbf{v}_5$. The W_2C area to the far right on the line is achieved by a large weighting of C4 with a small amount of C5 subtracted. The weightings of components C4 and C5 are thus also the only ones containing negative values.

Scaling this dataset to 10 million total counts and adding noise and performing rotated PCA yields the results in figure 13. Essentially a noisy variant of figure 12. The weighting of component 5 is extremely noisy but can still be distinguished. Interestingly, noise seems to only be reflected in the component weightings, not the vectors.

In conclusion, on this model, performing varimax rotation on the PCA components drastically improves interpretability, both on pure and noisy data, except for areas with a gradient of WC to W_2C. It may be the case that generally, systems containing source spectra that are only present mixed in gradients are not as well suited for varimax spatial rotation. Since all pixels in the gradient are linear combinations of more than one source spectra, this goes against the underlying assumption made for this kind of rotation. This would also confirm the conclusion made by Keenan [14], where he argues that the minimum requirement for a successful rotation, is that each underlying chemical component needs to have at least one *selective pixel*. That is, at least one location where all except one chemical component are equal to zero. In the other limit where all pixels are selective, varimax factor rotation will produce vectors that are the best approximation possible in a least-square sense to the underlying source spectra [28]. In this simulated interface model, we know that there are pixels with pure W_2C (pixel 100-110) and pure WC (pixel 105-150), but mixing also occurs in pixels in the gradient area. And so the system is somewhere in between the two limits. Therefore, a logical assumption would be that the more selective pixels a system has, the better varimax factor rotation can approximate the underlying chemical source spectra. This raises the question if binning operations, where signals from neighbouring pixels are combined for an increased S/N ratio, while improving overall PCA accuracy, lowers the chances of discovering the real underlying chemical source spectra using factor rotations.

4.3 Experimental results

4.3.1 Interface

An experimental electron image of a used WC-Co cutting tool turned in Ti6Al4V can be seen in figure 14. The adhered Ti6Al4V material can be seen with dark contrast to the left, and the brighter contrast WC grains to the right. Figure 15 shows a magnified image of the interface with a WC grain. A suitable area to perform a linescan over the interface is marked in yellow in this figure. Unfortunately, the EDS software does not allow exporting of the raw data for a true linescan, only EDS maps. To get around this, EDS mapping was performed in the yellow area corresponding to 50 x 466 pixels. A total of $1.657 \cdot 10^7$
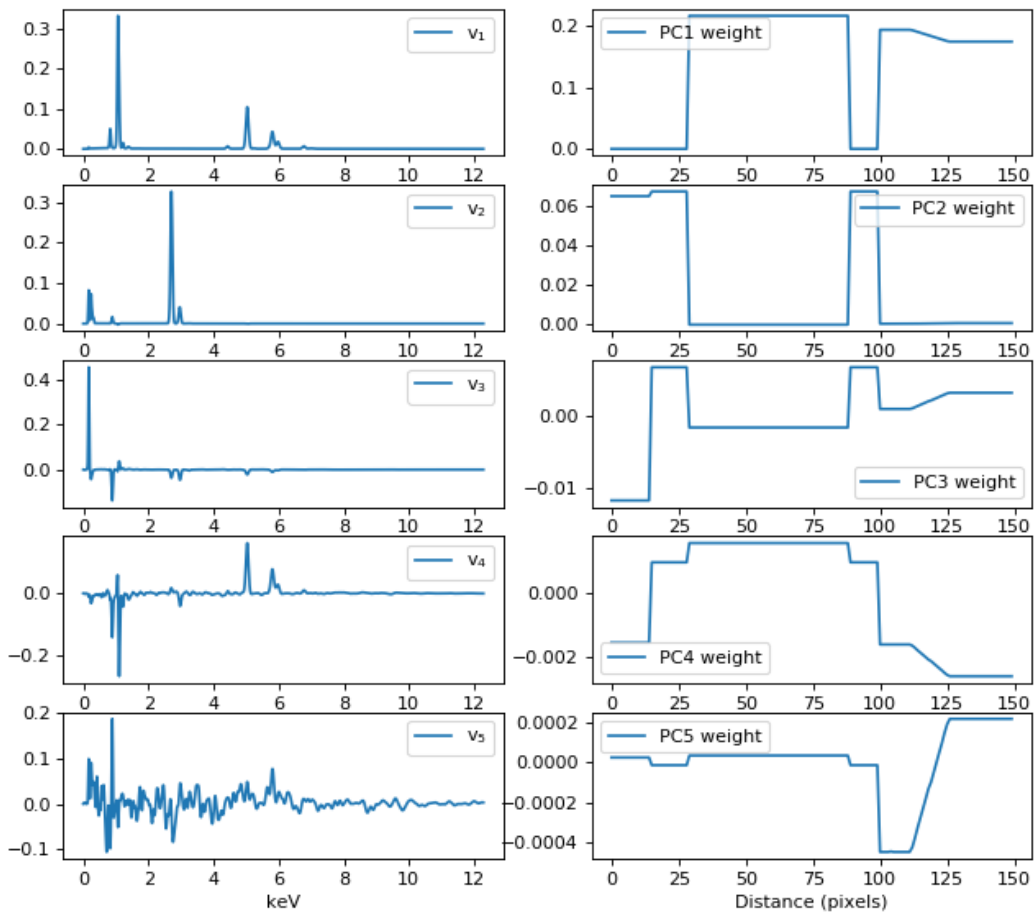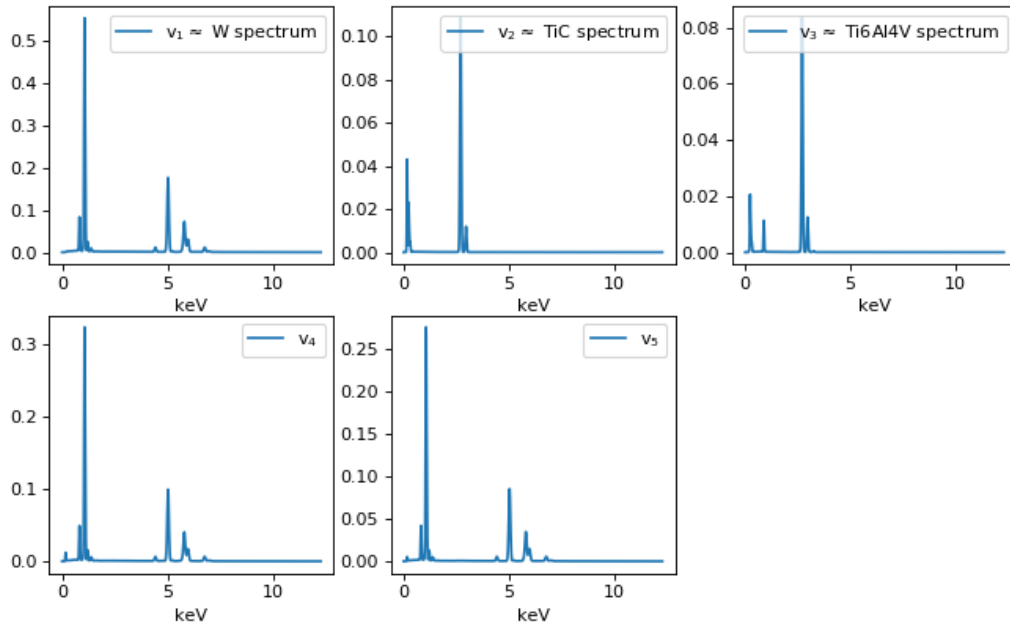
Figure 11: PCA results of noise-free linescan (red line in figure 3) over the simulated interface. Component vectors 1 to 5 are shown to the left and their respective weightings to the right
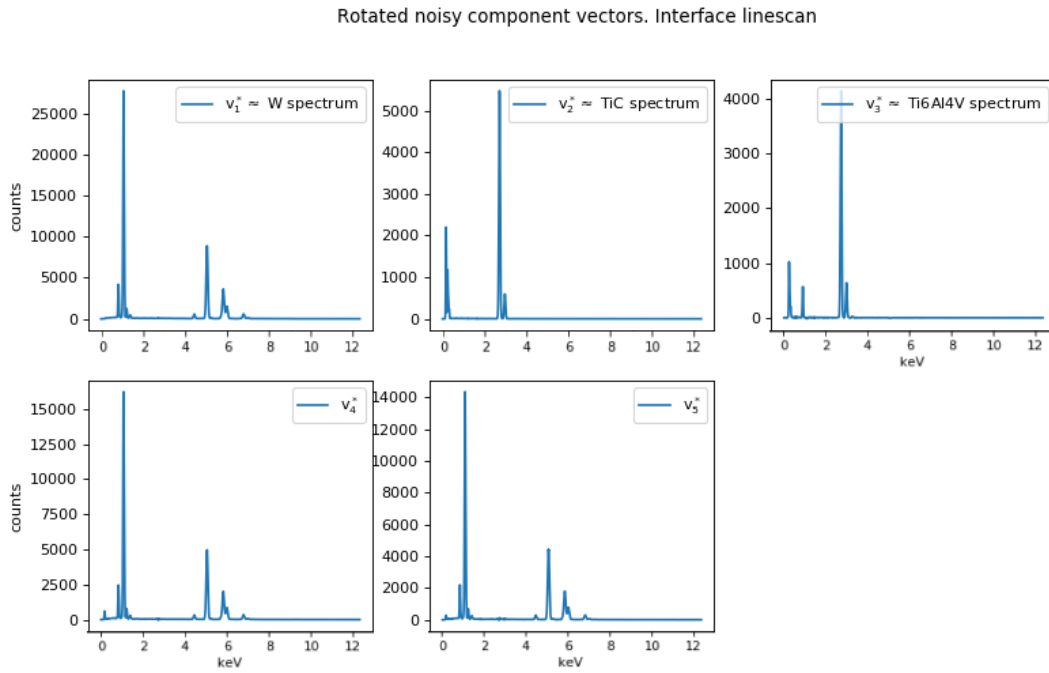
(a) Rotated PCA component vectors from the linescan over the simulated interface in figure 3. Components 1 to 3 can be seen to exactly represent source spectra used to construct the model
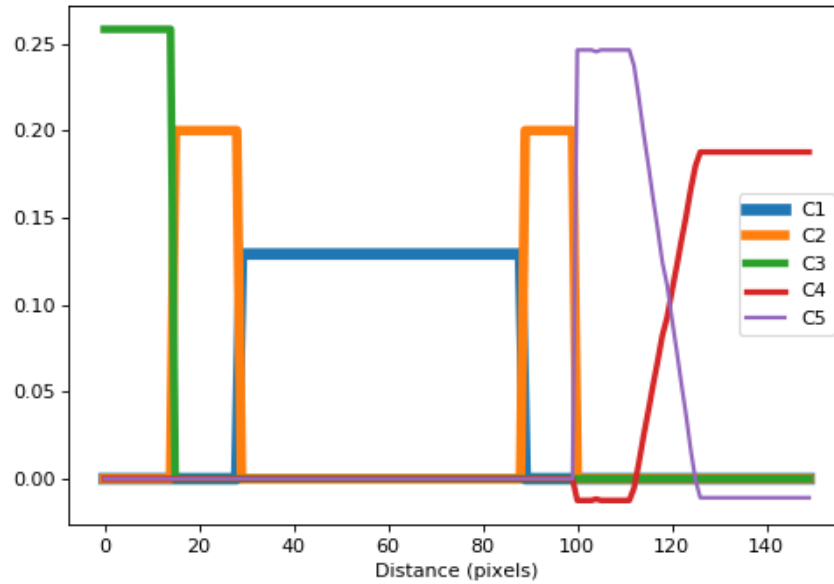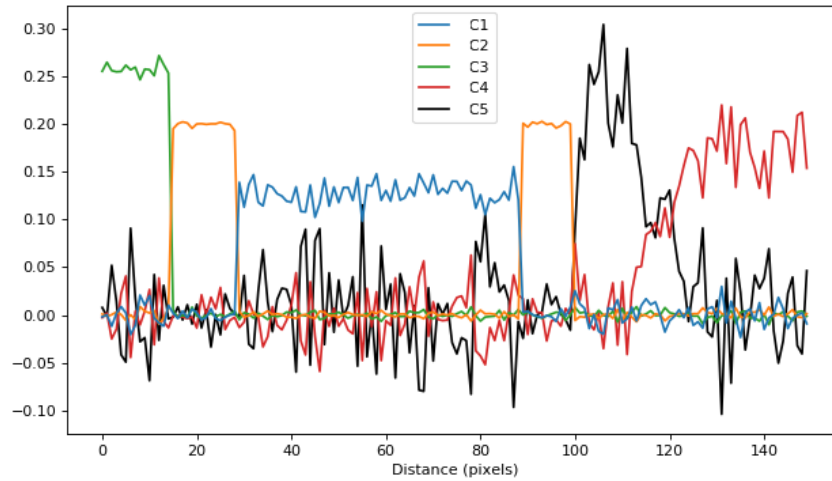

(b) Weightings of the rotated components whose corresponding component vectors can be seen in (a).

Figure 12: Results of rotating the principal components vectors and weights obtained from the linescan over the simulated interface.

Rotated noisy component vectors. Interface linescan

(a) Rotated PCA component vectors from the linescan with added noise over the simulated interface in figure 3. Noise does not seem to affect the component vectors who look identical to the ones in figure 12 (a).

(b) Weightings of the rotated components whose corresponding component vectors can be seen in (a).

Figure 13: Result of performing factor rotations on a noisy dataset consisting of a linescan over the simulated interface with
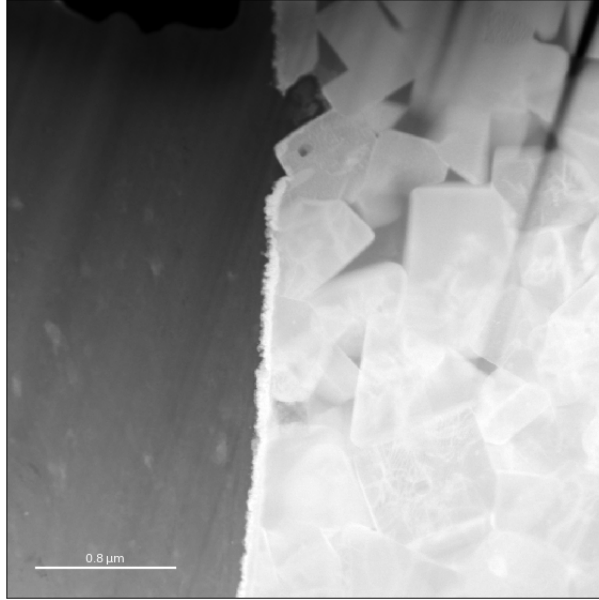
Figure 14: STEM HAADF image. Overview of used cuttingtool interface. WC grains to the right in bright contrast. Adhered Ti6Al4V material in low contrast to the left
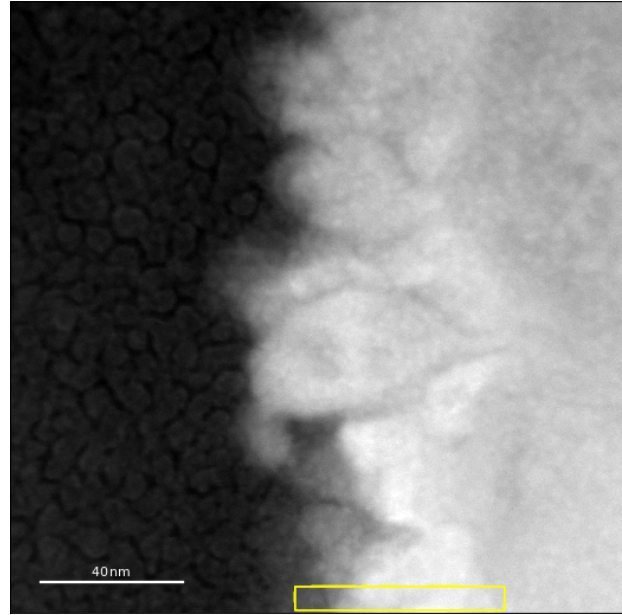

Figure 15: STEM HAADF image of interface between WC grain and adhered Ti6Al4V material. Area used for linescan is marked in yellow

counts was collected. The sum map can be seen in figure 16 that show the sum of all channels in each pixel. The data was subsequently binned to a line of 1 x 150 pixels. This procedure hopefully does not introduce significant artifacts to the data. The sum spectrum, i.e. the sum of all pixels in each channel can be seen in figure 17. It is evident that the W M_α is the highest peak and dominates the image. The second highest is Copper K_α which is expected since the sample holder was made of copper and is always present as a system generated peak as a function of electron scattering [9]. A classical window analysis where a characteristic peak intensity is viewed over all pixels can be seen in figure 18. The main W M_α peak is the overall strongest signal and can be seen to increase from the left with a maximum at around pixel 70, after which slowly decreasing a bit. This supports the existence of a bcc-W region in the middle of the interface. In this sample, the Cu signal seems to follow the overall intensity of the sum map. In fact it seems to be a direct linear function of the total counts in each pixel on the line. It was found that for the total counts in pixel i, x_i, the Copper K_α signal y_i very precisely follows the relationship

$$y_i = 0.01486x_i - 275 \qquad (25)$$

This linear fit to the sum signal can be seen in figure 19. This means that the Cu signal varies exactly as

Figure 16: Map of the sum of all energy channels summed for the EDS map obtained in the the yellow area in figure 15
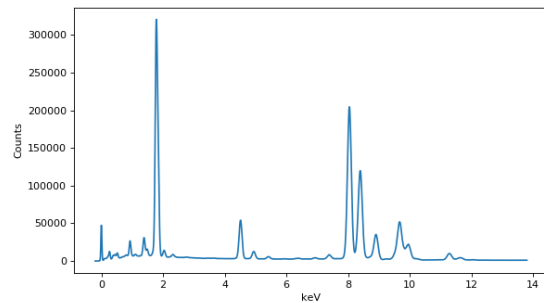

Figure 17: Sum spectrum of EDS map obtained in the yellow area in 15
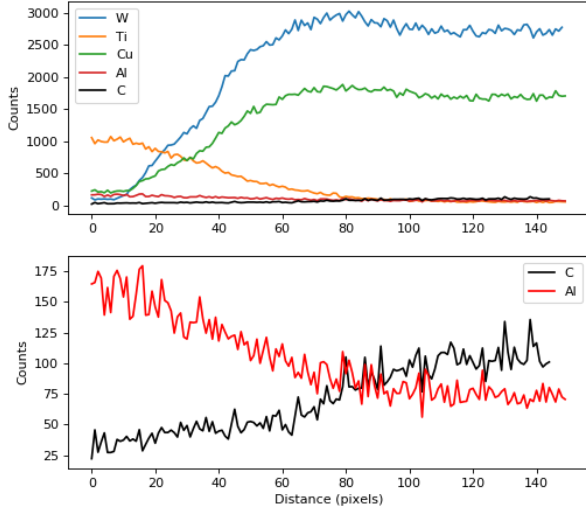
20

Figure 18: Energy window analysis of the line over the interface. Lower image shows a magnification of the weaker signals C and Al.
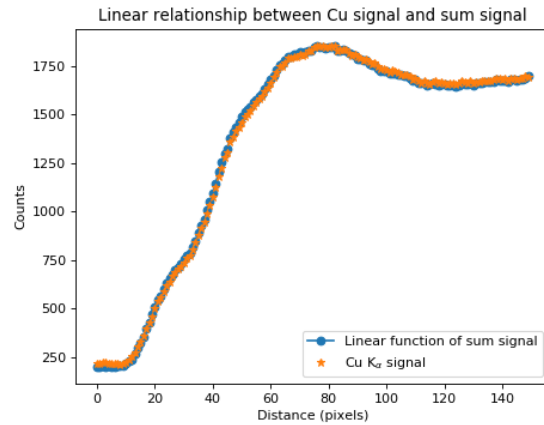

Figure 19: Plot showing the linear relationship between the Cu signal and total counts in each pixel. Blue line shows a linear fit of the sum signal to the Cu Kα peak.

the mean across the sample. This will hopefully mean that the unavoidable Cu signal does not increase the rank of the data matrix, i.e. it will not show up as its own independent principal component. This behaviour was confirmed during simulations for a component that varies according to the mean.

Before analysis of PCA results was made, the number of components to treat as significant, i.e. the rank of the data, needed to be determined. This was done by discarding component vectors with isotropic appearances as described in section 4.1.1 along with the "knee" method described in [29]. The knee method is performed on a plot of the explained variances $\lambda_1^*, \lambda_2^*..., \lambda_n^*$ plotted in a descending order. This analysis resulted in the consideration of 4 significant components i.e. the rank of the data matrix was determined to be 4.

An attempt was made to use the more robust rank estimator developed by Kritchman and Nadler [27]. During trials of this algorithm on noisy simulated datasets scaled for Poisson noise, it predicted the right number of factors only on datasets were n was comparable in size to m. Therefore this method was deemed unreliable. It should be noted that it always provided a very good estimate of the noise level on simulated data. Some minor modifications to this algorithm may be all that is needed for a good rank estimator for STEM EDS data.

PCA was performed on the line data and the resulting components are shown in figure 20. The first component is as usual recognised as the mean spectrum, PC1 is the sum spectrum in figure 17 divided by the number of pixels. Here we can see that the the middle area contributes the most in increasing the overall counts, corresponding well to the sum map. PC2 show the anti-correlation of W and Cu to Ti, Al and V. Presumably, this component corresponds to the Ti6Al4V signal and its difference to the mean.

PC3 has a very large positive carbon peak anticorrelated to W and Ti. This suggests a carbon component that varies independently of W. If W and C was always together in the form of WC (i.e. correlated), a C peak would have shown up in PC2 together with W, as anti-correlating to Ti6Al4V. Considering this and the weighting of PC3 supports the theory of a bcc-W region from pixel 50 to 75. Furthermore we can see this component increasing a bit into the Ti6Al4V region to the far left. This may indicate that some of the carbon has diffused from WC, into the Ti6Al4V alloy.

PC4 shows basically only an anticorrelation of W to Cu. It is really hard to see what this is representing, especially looking at its complex weighting. One key observation though is that this is an additional component dominated by W that varies independently of the others.

Factor rotation was then performed on the obtained PCA components on the interface linescan and the results are shown in figure 21. The component vectors now look more like real spectra, as a direct con-
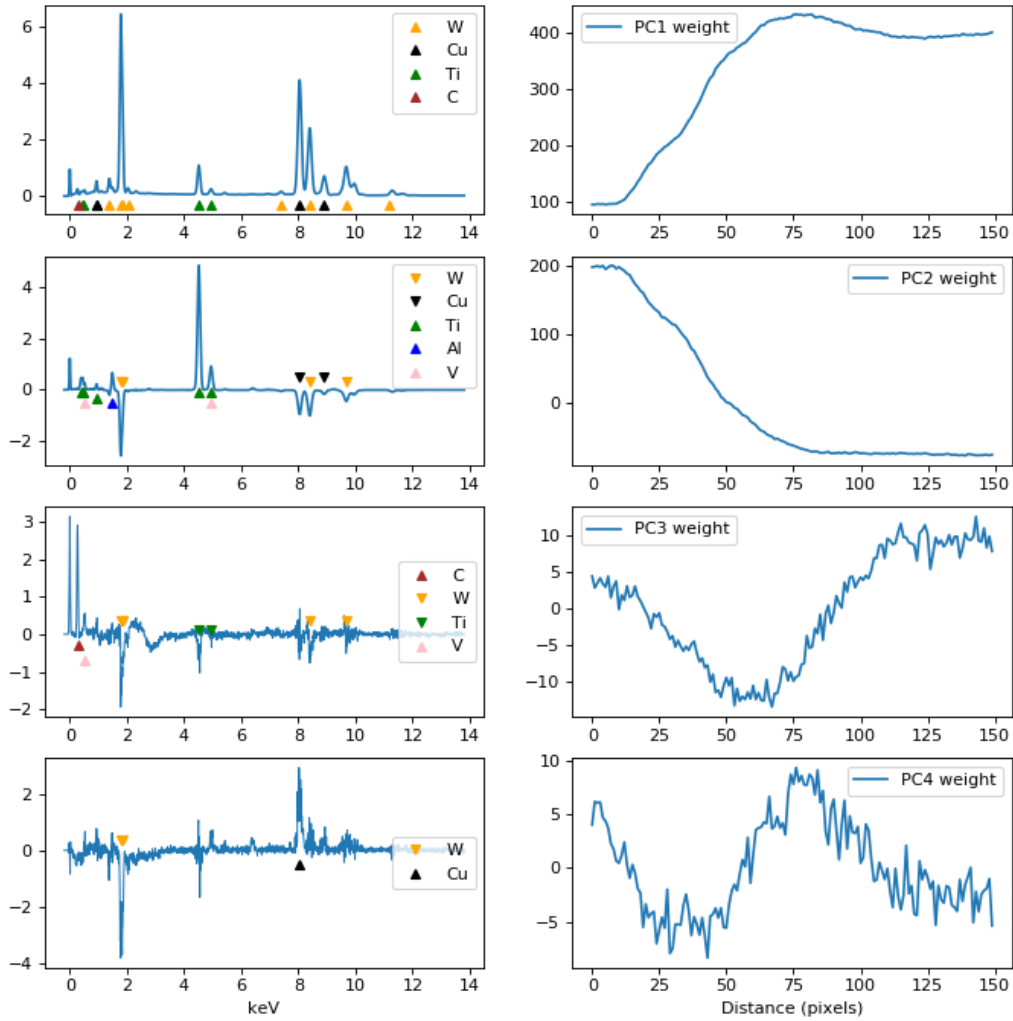
Figure 20: PCA component vectors (left) and corresponding weights (right) obtained from experimental data from the the linescan over the interface. Characteristic peaks for different elements are indexed with arrows corresponding to their signs

sequence of the orthogonality constraint being lifted. Unfortunately, this has also led them to all look quite similar. There is a similar looking cluster of W and Cu peaks in C1, C3 and and C4. Also, the largest peak in C1, C3 and C4 is the W M_α peak. Looking at the C1 vector it seems to be the one most closely representing the WC spectrum (mixed with Cu). The only difference to C3 seems to be a slightly larger C peak. In fact all component vectors have C K_α peaks, where C1 has the biggest followed by C3, C4 and C2. Looking at their weightings, C1 seems to increase and stabilise into a weighting of about 0.15 the right. At the same time all other component weights goes to zero. Basically, to the far right, C1 can describe the data all by itself and should be interpreted as coming from WC. In fact, overall there are 4 locations on the line were a single component describes all the data all by itself, i.e. when the weights of all other components are zero. These points also coincide with the maxima of each component. These are found following the horizontal purple line in figure 21 b. C1 is representative of the data at pixel 149, C2 at pixel 7, C3 at pixel 73 and C4 at pixel 41. In conclusion then, going from left to right across the interface we can first observe an initial component rich in Ti (C2) changing into a region with a component with reduced Ti but enriched in C and W (C4). This changes into a component with essentially no Ti but high in W and a small increase in C (C3). The line finally ends with a component rich in W and even more C (C1).

If the underlying structure resembles the simulated model, the collected counts should be enough to distinguish between W, WC and W_2C. Conclusive evidence for any W_2C phase could not be found by relating to the results from simulations. However, it was found that carbon anticorrelates with tungsten in some places (PC3 in figure 20) which is strong evidence for areas where C is not in the form of WC. Furthermore, the finding of 4 significant components suggests that some of the theorised new species exists here, if one assumes the Cu signal does not increase the rank of the data. In other words, 2 of the three chemical species (W, W_2C and TiC) may be present. This is not at all a trivial conclusion if one were to look only at the plot of a carbon energy window over interface as in figure 18.

Next, a larger EDS map of the whole area in figure 15 was acquired and the sum map can be seen in figure 22. Initially this data cube was 512 x 448 pixels ($m = 229376$). On the raw data, the rank of the matrix was determined to be 2. The knee method actually estimated the rank to 3, but by performing

PCA and looking at the isotrophy of the component vectors, component vector 3 consisted only of a extremely small spike in the positive and negative direction, and was completely indistinguishable from components 4, 5, 6,..n in appearance and, therefore most likely represented noise.

Finding only 2 significant components across this very large and complex area suggested that many components probably had signal variances below the noise level and were therefore impossible to retrieve initially. Knowing that increasing the counts per pixel affects noise eigenvalues and signal eigenvalues differently, binning to a range of different sizes was performed while observing the resulting eigenvalues of D^TD, λ_j^*. A plot can be seen in figure 23. Indeed, it is clear that by binning, two additional eigenvalues, hopefully originating from signal variance, break away from the rest. From this, the conclusion was that eigenvalues 1 to 4 behave as signal variances, and eigenvalues 5 and onward behave purely as noise. Keeping the 4 first signal eigenvalues above the noise level, while still retaining adequate spatial resolution, a dataset binned to $m = 1554$ (37 x 42 pixels) was chosen for PCA.

The resulting PCA components and weightings can be seen in figure 24. The first component represents the mean once again and can be seen being the weighted strongest to the right, following the strong signal from W.

The second component shows the anticorrelation of Ti and Al to the mean and its weighting suggests that this component reperesents the Ti6Al4V alloy. The third component is very interesting. An anticorrelation between Ti and V is evident. Ti and V are classically hard to separate since the Ti K_β peak exactly overlaps the V K_α peak. Since peaks coming from the same atomic species would theoretically never anticorrelate to eachother, and the fact that the V K_β is also visible is evidence of Ti and V varying independently of eachother. Correlating positively with V in component 3 is Al and surprisingly also Fe. Ti is also seen correlated with C. Looking at the components weighting, a distinct ridge can be seen crossing vertically over the whole image. PC3 can then be interpreted as follows: on the ridge lies a higher concentration of V, Al and Fe. An alternative view is this: to the left of the ridge, where the weighting of PC3 is negative, can be seen as region enriched with Ti and C.

PC4 shows a strong positive carbon peak, and some anticorrelation to Ti and W. The greatest impact of this component can be seen to the right in the image.
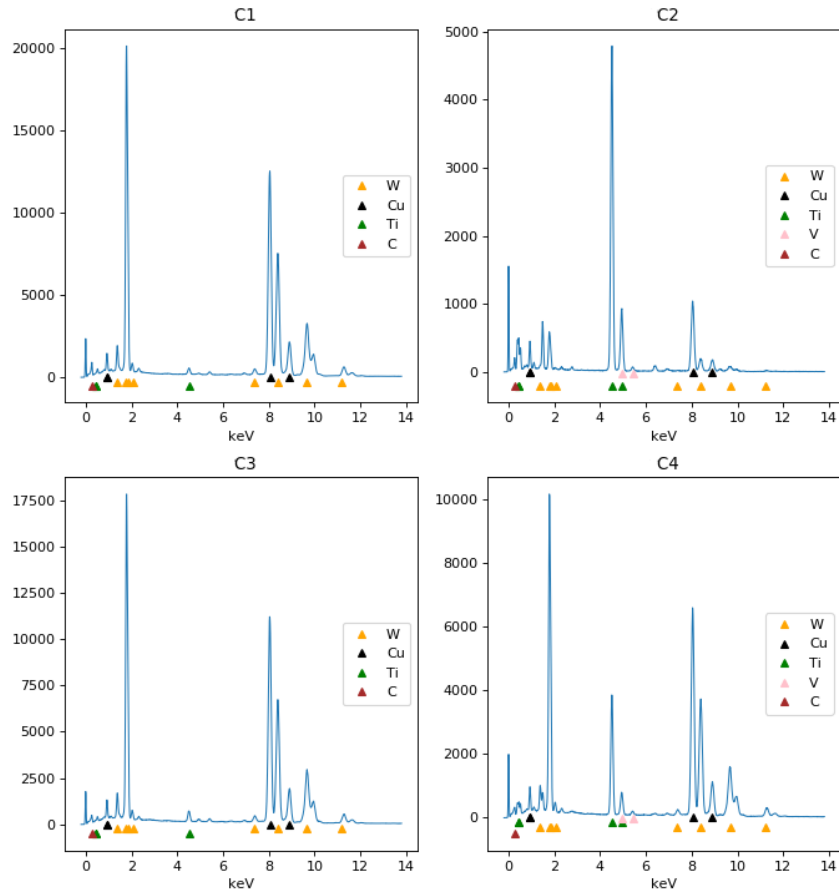
(a) Rotated component vectors from experimental data obtained for the linescan over the interface. Characteristic peaks for different elements are indexed with arrows corresponding to their signs
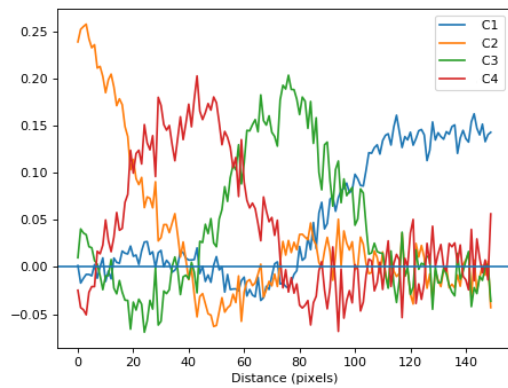
(b) Rotated component vectors for the corresponding component weights in (a). Purple line helps to show were weights are close to zero.

Figure 21: Results of rotating the PCA components obtained from experimental data from the the linescan over the interface.
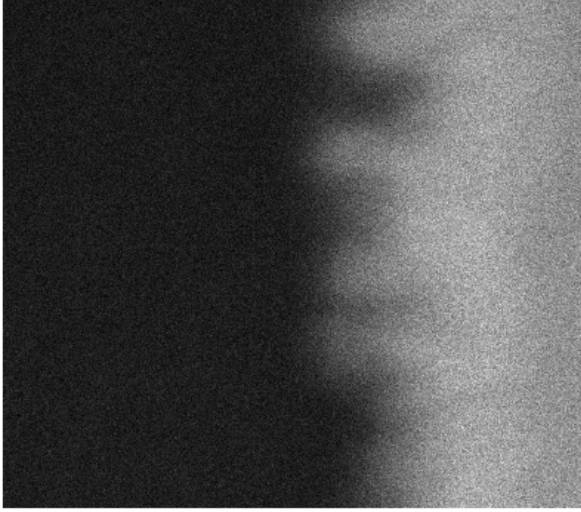
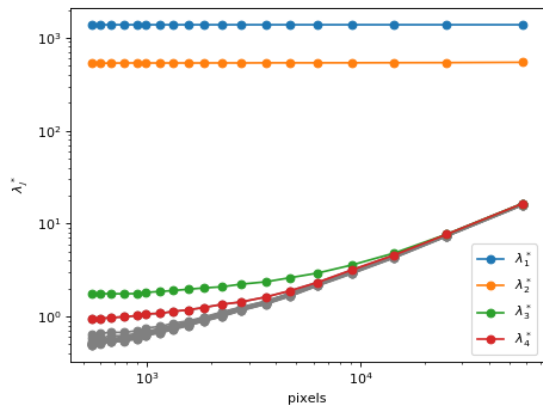Figure 22: Sum map of EDS map obtained over the whole area in figure 15

Figure 23: Plot showing the effect of rebinning on the eigenvalues of $D^T D$, where D is the flattened EDS map in figure 22. Increasing the binning factor means going from right to left in the figure. In descending order, λ_1, λ_2, λ_3, λ_4 are plotted in dark blue, orange, green and red respectively. The rest of the eigenvalues are clustered tight together and seem to decrease the same way. They are therefore consider to only represent noise.

The weighting starts out as positive while quickly becoming negative, meaning carbon is subtractded from the darkest region in the image. This supports the theory of a C depletion of WC at the interface.

Factor rotation gives the results in figure 25. The first component shows something that most closely resembles the W spectrum so far (ignoring the Cu peaks). It has a small contribution of Ti but essentially no C. C3 is very similar to C1 but it has the C peak and its W peaks are of slightly lower intensity. Therefore, the conclusion is that C1 represents W and C3 represents WC. The difference between C2 and C4 is in contrast not as easy to explain. C4 seems to have some W mixed in with a bit more C. C2 also has some Al. Clearly they both represent something with a lot of Ti but exactly what is hard to tell. The ridge structure is also apparent in both of them, weighted negative in C2 and positive in C4.

In conclusion, on this bigger map of the interface, W and WC species have been identified at the interface were they were expected. A ridge not visible in the electron micrograph was found where there is some anticorrelation mainly between Ti and V and possibly Fe. This ridge could be the subject of further work, investigating the presence of TiC or perhaps a VC phase.

5 Conclusions and outlook

In this work, the limits of PCA on hyperspectral EDS maps has been investigated. Previously, the effect of sample size on the accuracy of PCA has been evaluated which predict an increase in accuracy for the blunt increase in the number of pixels[4][12]. However these models assume a constant noise level. If phased with a limited collection time, it was found in this work, both theoretically and on simulated datasets, that counts should preferably be collected in as few pixels as tolerable to be able to resolve low variance components. However, rebinning pixels in a dataset post acquisition seem to have the same positive effect and can substantially help in discovering low variance signals. It may also be beneficial to investigate the benefits of binning the spectrum data in the energy axis as well [30]. Although not included here, limited testing on simulated datasets found no substantial difference for energy binning.

It was shown on one of the simulated datasets that varimax factor rotations in the spatial domain can make the PCA components easier to interpret as physical spectra. However, on experimental datasets, even if the components look like physical spectra,
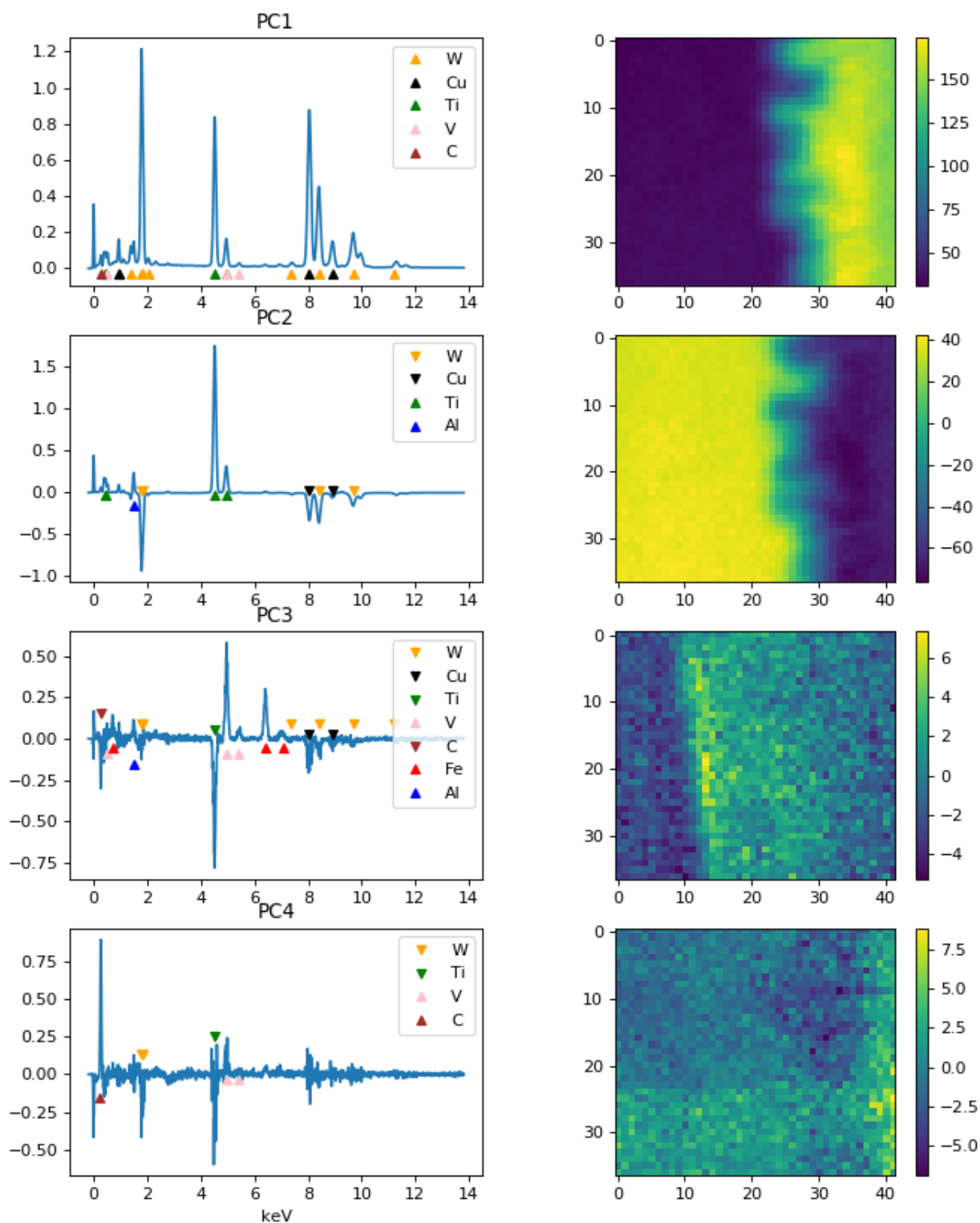
Figure 24: Results of PCA performed on the experimental EDS map over the interface. Characteristic peaks for different elements are indexed with arrows corresponding to their signs on the components to the left. The mapping to the right includes a colorbar that specifies the weighting. Note that the weights can have negative values. A distinct ridge structure can be seen in the weighting of PC3, which is not visible in the HAADF image in figure 15.
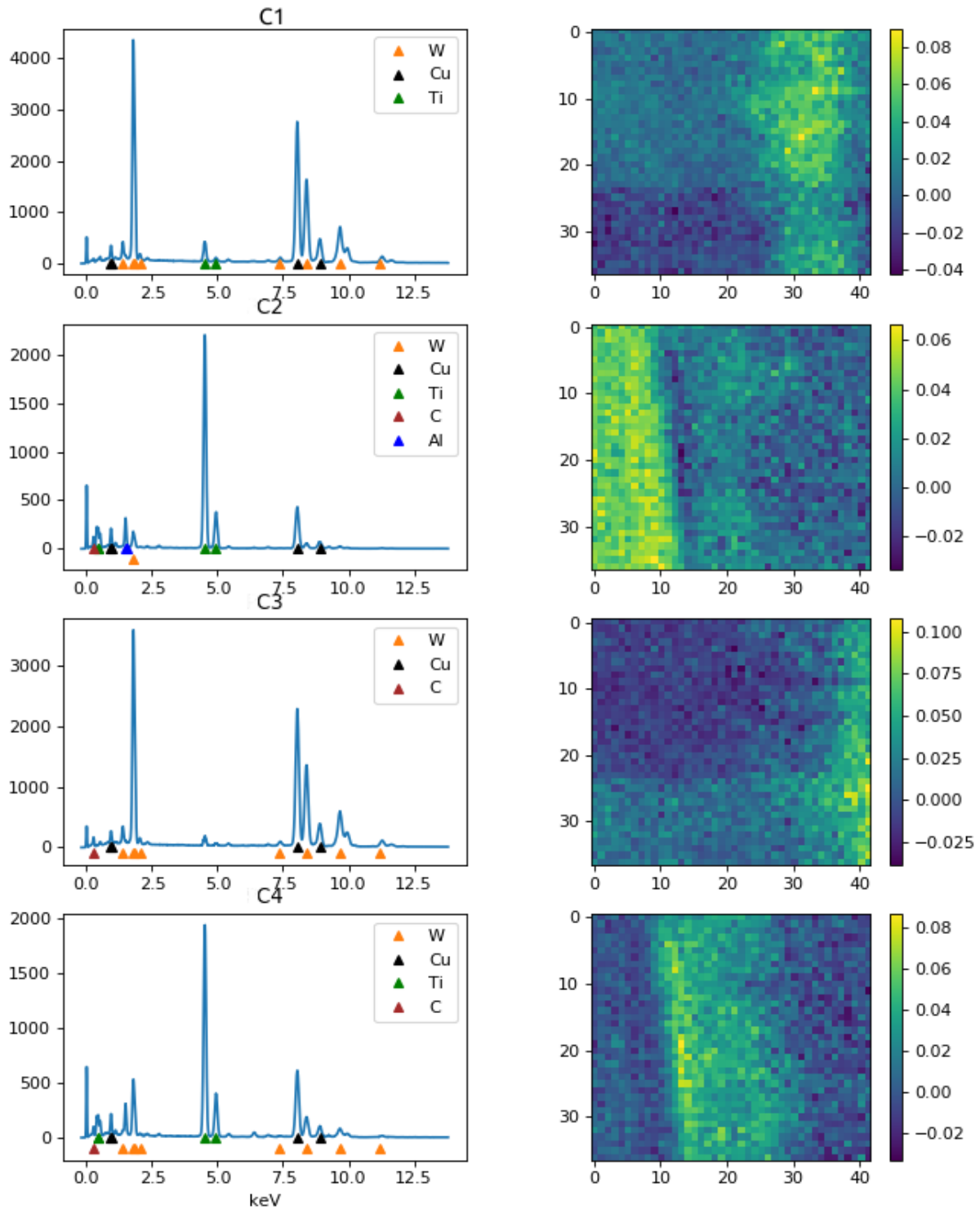
Figure 25: Results of the rotated PCA components performed on the experimental EDS map over the interface. Note that even though the peaks in the component vectors are all positive, the mapping of the weight shows some negative values.

mixing of peaks from several elements can occur that may not correspond to actual chemical species. Presumably, these kinds of rotations work better on datasets with many selective pixels, that is, many locations where only one chemical component is present. This may be a relationship suitable for further investigation.

For the cuttingtool sample. No conclusive evidence for a W_2C phase could be found, even when the according to simulations required amount of counts was collected. Some areas where Ti anticorrelates to V was found. This implicitly implies that by using PCA, it is possible to distinguish V from Ti on EDS maps which is difficult to do classically due to major peak overlaps.

As the rotated factors contain mixes of several elements, a next step may be to perform cluster analysis. So called fuzzy KMeans clustering can be performed to separate out spectra from individual atomic species although this method requires a lot of *a priori* knowledge about the sample [31]. Alternatively, varimax rotations in the spectral domain is another method that can be tried to maximize the differences between component vectors [15].

Finally, referring back to equation 20, if it is discovered that the value of the explained variances in a dataset depends almost exclusively on how much the source spectra differ from each other, it may be possible to, through simulations, predict the counts needed to discover suspected chemical species in a sample purely based on how their source spectra differ from each other, without knowledge about sample geometry.

References

[1] J. J. Friel and C. E. Lyman, "Tutorial review: X-ray mapping in electron-beam instruments," *Microscopy and Microanalysis*, vol. 12, no. 1, p. 2–25, 2006.

[2] S. Muto and M. Shiga, "Application of machine learning techniques to electron microscopic/spectroscopic image data analysis," *Microscopy*, vol. 69, no. 2, pp. 110–122, 11 2019. [Online]. Available: https://doi.org/10.1093/jmicro/dfz036

[3] M. R. Keenan, "Multivariate analysis of spectral images composed of count data," in *Techniques and Applications of Hyperspectral Image Analysis*, H. Grahn and P. Geladi, Eds. Wiley, 2007, ch. 5, pp. 89–126. [Online]. Available: https://www.wiley.com/en-se/Techniques+and+Applications+of+Hyperspectral+Image+Analysis-p-9780470010860

[4] J. Spiegelberg and J. Rusz, "Can we use pca to detect small signals in noisy data?" *Ultramicroscopy*, vol. 172, pp. 40 – 46, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399116302480

[5] S. Odelros, B. Kaplan, M. Kritikos, M. Johansson, and S. Norgren, "Experimental and theoretical study of the microscopic crater wear mechanism in titanium machining," *Wear*, vol. 376-377, pp. 115 – 124, 2017, 21st International Conference on Wear of Materials. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S004316481730248X

[6] B. Kaplan, S. Odelros, M. Kritikos, R. Bejjani, and S. Norgren, "Study of tool wear and chemical interaction during machining of ti6al4v," *International Journal of Refractory Metals and Hard Materials*, vol. 72, pp. 253 – 256, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0263436817306972

[7] J. García, M. Lattemann, E. Coronel, and I. Sadik, "Interaction between cemented carbide and ti6al4v alloy in cryogenic machining," 06 2017.

[8] K. Korniyenko and Materials Science International Team, MSIT®, "The C-W phase diagram: Datasheet from MSI Eureka in SpringerMaterials," copyright 2007 MSI, Materials Science International Services GmbH, Stuttgart. [Online]. Available: https:

//materials.springer.com/msi/phase-diagram/docs/sm_msi_r_10_010004_01_full_LnkDia0

[9] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy.* Springer, Boston, MA, 2009.

[10] "X-ray and eels imaging," in *Transmission Electron Microscopy - Diffraction, Imaging, and Spectrometry*, B. Carter and D. B. Williams, Eds. Springer International Publishing, 2016, ch. 16, pp. 439–466.

[11] B. Nadler, "Finite sample approximation results for principal component analysis: A matrix perturbation approach," *Ann. Statist.*, vol. 36, no. 6, pp. 2791–2817, 12 2008. [Online]. Available: https://doi.org/10.1214/08-AOS618

[12] P. Potapov, "On the loss of information in pca of spectrum-images," *Ultramicroscopy*, vol. 182, pp. 191–194, 2017. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85021964404&doi=10.1016%2fj.ultramic.2017.06.023&partnerID=40&md5=bcde149e60238c244a9e73bdc7b9020d

[13] ——, "Why Principal Component Analysis of STEM spectrum-images results in uninterpretable loadings?" *Ultramicroscopy*, vol. 160, pp. 197 – 212, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304399115300577

[14] M. R. Keenan, "Exploiting spatial-domain simplicity in spectral image analysis," *Surface and Interface Analysis*, vol. 41, no. 2, pp. 79–87, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sia.2949

[15] V. S. Smentkowski, S. G. Ostrowski, and M. R. Keenan, "A comparison of multivariate statistical analysis protocols for tof-sims spectral images," *Surface and Interface Analysis*, vol. 41, no. 2, pp. 88–96, 2009. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sia.2973

[16] A. Buhot and M. B. Gordon, "Phase transitions in optimal unsupervised learning," *Phys. Rev. E*, vol. 57, pp. 3326–3333, Mar 1998. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.57.3326

[17] T. L. H. Watkin and J. P. Nadal, "Optimal unsupervised learning," *Journal of Physics A: Mathematical and General*, vol. 27, no. 6, pp. 1899–1915, mar 1994. [Online]. Available: able: https://doi.org/10.1088%2F0305-4470%2F27%2F6%2F016

[18] N. Ipsen and L. Hansen, "Phase transition in pca with missing data: Reduced signal-to-noise ratio, not sample size!" in *Proceedings of Machine Learning Research*, vol. 97. International Machine Learning Society (IMLS), 2019, pp. 5248–5260, 36th International Conference on Machine Learning, ICML 2019 ; Conference date: 10-06-2019 Through 15-06-2019.

[19] N. Faber, M. Meinders, P. Geladi, M. Sjöström, L. Buydens, and G. Kateman, "Random error bias in principal component analysis. part i. derivation of theoretical predictions," *Analytica Chimica Acta*, vol. 304, no. 3, pp. 257 – 271, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/000326709400585A

[20] M. R. Keenan and P. G. Kotula, "Accounting for poisson noise in the multivariate analysis of tof-sims spectrum images," *Surface and Interface Analysis*, vol. 36, no. 3, pp. 203–212, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sia.1657

[21] Nicholas W. M. Ritchie , "NIST DTSA-II." [Online]. Available: https://cstl.nist.gov/div837/837.02/epq/dtsa2/

[22] "TiC crystal structure: Datasheet from "pauling file multinaries edition – 2012" in springermaterials," copyright 2016 Springer-Verlag Berlin Heidelberg & Material Phases Data System (MPDS), Switzerland & National Institute for Materials Science (NIMS), Japan. [Online]. Available: https://materials.springer.com/isp/crystallographic/docs/sd_0261188

[23] "Titanium Ti-6Al-4V (Grade 5), Annealed Bar," Aug 2020, [Online; accessed 4. Aug. 2020]. [Online]. Available: http://www.matweb.com/search/DataSheet.aspx?MatGUID=10d463eb3d3d4ff48fc57e0ad1037434

[24] "W crystal structure: Datasheet from "pauling file multinaries edition – 2012" in springermaterials," copyright 2016 Springer-Verlag Berlin Heidelberg & Material Phases Data System (MPDS), Switzerland & National Institute for Materials Science (NIMS), Japan. [Online]. Available: https://materials.springer.com/isp/crystallographic/docs/sd_0250184

[25] "WC crystal structure: Datasheet from "pauling file multinaries edition – 2012" in springerma-

terials," copyright 2016 Springer-Verlag Berlin Heidelberg & Material Phases Data System (MPDS), Switzerland & National Institute for Materials Science (NIMS), Japan. [Online]. Available: https://materials.springer.com/isp/crystallographic/docs/sd_0261436

[26] "W2C crystal structure: Datasheet from "pauling file multinaries edition – 2012" in springermaterials," copyright 2016 Springer-Verlag Berlin Heidelberg & Material Phases Data System (MPDS), Switzerland & National Institute for Materials Science (NIMS), Japan. [Online]. Available: https://materials.springer.com/isp/crystallographic/docs/sd_0260535

[27] S. Kritchman and B. Nadler, "Determining the number of components in a factor model from limited noisy data," *Chemometrics and Intelligent Laboratory Systems*, vol. 94, no. 1, pp. 19 – 32, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169743908001111

[28] C. A. Bernaards and R. I. Jennrich, "Orthomax rotation and perfect simple structure," *Psychometrika*, vol. 68, no. 4, pp. 585–588, Dec 2003.

[29] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a "kneedle" in a haystack: Detecting knee points in system behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops*, 2011, pp. 166–171.

[30] P. Kotula and M. Van Benthem, "Revisiting noise scaling for multivariate statistical analysis," *Microscopy and Microanalysis*, vol. 21, no. S3, p. 1423–1424, 2015.

[31] C. L. Stork and M. R. Keenan, "Advantages of clustering in the phase classification of hyperspectral materials images," *Microscopy and Microanalysis*, vol. 16, no. 6, p. 810–820, 2010.

6 Appendix A: Algorithms and Python implementation

Rescale data for Poisson noise

Transforms data into new space where noise is uniform. D is an unfolded mxn image. Returns the scaled data D_ along with the matrices sqrt(G) and sqrt(H). After performing matrix decomposition on D_ and obtaining matrices T_ and P_ you should invert scaling with T = root_G*T and P = root_H*P.

```python
def rescalePoisson(D):
    import numpy as np

    m = np.shape(D)[0]
    n = np.shape(D)[1]
    aG = np.squeeze(np.sum(D, axis=1))/n
    bH = np.squeeze(np.sum(D, axis=0))/m
    root_G = np.sqrt(aG)[:, np.newaxis]
    root_H = np.sqrt(bH)[np.newaxis, :]
    #ignore divide by zero and set NaNs to 0
    with np.errstate(divide="ignore", invalid="ignore"):
        D_ = D/(root_G*root_H)
        return np.nan_to_num(D_), root_G, root_H
```

Varimax factor rotation

Implemented from the function "varimax" in the R language, adapted for python here `https://en.wikipedia.org/wiki/Talk:Varimax_rotation`. Phi is the factor matrix to be rotated and R is returned as the calculated rotation matrix. In this article Phi = T, but this function can also be used to calculate a rotation matrix based on P, i.e. in the spectral domain.

```python
def varimax(Phi, gamma = 1.0, q = 20, tol = 1e-6):
    import numpy as np
    p,k = Phi.shape
    R = np.eye(k)
    d=0
    for i in range(q):
        d_old = d
        Lambda = np.dot(Phi, R)
        temp1 = np.dot(Lambda, np.diag(np.diag(np.dot(Lambda.T,Lambda))))
        temp2 = np.dot(Phi.T,np.asarray(Lambda)**3 - (gamma/p) * temp1)
        u,s,vh = svd(temp2)
        R = np.dot(u,vh)
        d = np.sum(s)
        if d/d_old < tol:
            break
    return R
```