Article

Evaluation of EELS spectrum imaging data by spectral components and factors from multivariate analysis

Siyuan Zhang[1,*] and Christina Scheu[1,2]

[1]Max-Planck-Institut für Eisenforschung GmbH, Max-Planck-Straße 1, 40237 Düsseldorf, Germany and [2]Materials Analytics, RWTH Aachen University, Kopernikusstraße 10, 52074 Aachen, Germany

[*]To whom correspondence should be addressed. E-mail: siyuan.zhang@mpie.de

Abstract

Multivariate analysis is a powerful tool to process spectrum imaging datasets of electron energy loss spectroscopy. Most spatial variance of the datasets can be explained by a limited numbers of components. We explore such dimension reduction to facilitate quantitative analyses of spectrum imaging data, supervising the spectral components instead of spectra at individual pixels. In this study, we use non-negative matrix factorization to decompose datasets from Fe_2O_3 thin films with different Sn doping profiles on SnO_2 and Si substrates. Case studies are presented to analyse spectral features including background models, signal integrals, peak positions and widths. Matlab codes are written to guide microscopists to perform these data analyses.

Key words: electron energy loss spectroscopy, spectrum imaging, multivariate analysis, non-negative matrix factorization, component-based supervision, oxide interface

Introduction

Since the early days of electron energy loss spectroscopy (EELS), multivariate analysis (MVA) has been introduced to process EELS data [1]. As advancing EELS acquisition routines generate increasingly bigger datasets [2,3], MVA has been widely acknowledged as a powerful tool to process EELS datasets, including spatially resolved spectrum imaging in two dimensions [4–11] and three dimensions [12–14], and sets of angular-resolved [15] or site-specific [16] spectra. With the development in MVA algorithms, principal component analysis (PCA) [17], independent component analysis [18], non-negative matrix factorization (NMF) [19] and geometric extraction methods [20] (such as vertex component analysis and Bayesien linear unmixing [21]) have been

applied to treat EELS data. These methods have shown success in noise reduction and have been applied to identify fine structures either around the absorption edge onset or in the valence EELS regime.

MVA algorithms decompose a spectrum imaging dataset into the linear combination of a few spectral components, as expressed by Eq. (1),

$$f_p(E) = \sum_i c_{p,i} f_i(E) \tag{1}$$

where f_p and f_i are the EELS spectra as functions of the energy loss E for individual pixels p and individual spectral components i, respectively, and $c_{p,i}$ are the coefficients for

the linear combination. Equation (1) is usually an approximation, neglecting the residual term after the linear combination. At each pixel p, f_p is expressed as a linear combination of i spectral components f_i with the coefficients $c_{p,i}$, which is also known as the score matrix. As the number of components i is usually much smaller than the number of pixels p, f_i is a sparse representation of f_p, reducing its dimension by p-i. Such dimension reduction has enabled a concise representation of spectrum imaging datasets by a number of spectral components with their corresponding coefficients at each pixel for each component.

Thus far, MVA has been widely applied to decompose spectrum imaging datasets [4–14]. However, interpretation on the MVA spectral components remains descriptive and qualitative. It is more efficient to capitalize on the dimension reduction and evaluate the spectral components instead of the spectrum at individual pixels. The computing power will not be saved by such an approach, as the MVA algorithms themselves take more time to perform. On the other hand, the man power to supervise and present the spectrum imaging can greatly benefit from a dimension reduction.

In this article, we demonstrate the component-based supervision to evaluate EELS spectrum imaging data, and bridge the gap between MVA analysis and conventional (pixel-based) EELS quantification routines. Evaluation of quantitative information from EELS spectra requires proper modelling. Model-based quantification, such as EELSModel developed by Verbeeck *et al.* [22], composes an EELS spectrum by element-specific edges and background from the preceding edges, which requires detailed information on the constituents in the sample. It is more common to analyse one absorption edge at a time using assigned energy windows in the vicinity of the edge for background subtraction and signal integration [23]. The selection of energy windows is a necessary supervision for EELS analysis. Enforcing such supervision on individual pixels is a laborious work, prone to bias, and hence rarely done in practice. Making use of the dimension reduction from MVA analysis, supervision on the component level becomes practical, as the linear combination coefficients $c_{p,i}$ from MVA algorithms are unsupervised and unbiased. We examine two case studies to evaluate the signal of an absorption edge and the chemical shift of a white line feature, respectively, all based on the component-based supervision.

Data Acquisition

We used thin film hematite (α-Fe_2O_3) photoelectrodes as materials systems for the case studies, where Sn dopants are introduced to enhance their photoelectrochemical activity for solar-driven hydrogen production [24]. The distribution of Sn dopants is discussed for three cases, uniformly doped Sn (denoted as Sn: Fe_2O_3), partially doped Sn (denoted as Sn/: Fe_2O_3), and without Sn dopants (Fe_2O_3). Two quantities are of interest from the EELS spectrum imaging datasets, the signal from the Sn-$M_{4,5}$ edges to evaluate the elemental distribution, and the chemical shift of the Fe-L_3 white line that shows the oxidation state of Fe.

Electron-transparent specimens were prepared by Ar^+ ion milling and investigated in a FEI Titan Themis microscope operated at 300 kV. Aberration correction of the probe forming lenses enables a scanning probe of 24 mrad convergence semi-angle and ~1 Å probe size. Scanning transmission electron microscopy (STEM) images were collected by the annular bright field (ABF) and the high angle annular dark field (HAADF) detectors that cover the range of 8–16 mrad and 73–352 mrad, respectively. STEM-EELS spectrum imaging was acquired using a Gatan Quantum ERS energy filter in the image-coupled mode with a 35 mrad entrance aperture. Dual EELS collection mode enabled the zero loss peak (ZLP) alignment at individual pixels. Spectra were acquired from relatively thin areas with thicknesses of 0.3–0.5 times the inelastic mean free path. Therefore, the spectra primarily consist of single-scattered inelastic events so that deconvolution with the help of the low loss spectra was not conducted. To minimize specimen damage, charging, and the subsequent drift, moderate dose (0.1 nA probe current and 0.1 s pixel acquisition time) was used to collect spectrum imaging data at a 1 nm × 1 nm pixel size.

Data Analysis

As shown in Fig. 1a, the thin film photoelectrode grown by atomic layer deposition is much thinner than the fluorine doped SnO_2 substrate. This makes quantification of Sn dopants in the hematite thin films by energy dispersive X-ray spectroscopy very challenging, as nonlocal X-ray excitation (e.g. due to stray electrons and fluorescence artefacts) from SnO_2 contributes to the signal. On the other hand, EELS only collect forward scattered electrons (up to 35 mrad in this study), so that the signal only contains local inelastic scattering events [23]. The most convenient edges to study are Sn-$M_{4,5}$ edge, which has a delayed edge onset from ~480 eV and overlaps with the O-K edge from ~530 eV, and the Fe-$L_{2,3}$ white lines with onset from ~710 eV, as shown in Fig. 1b.

First, we show an example on the conventional evaluation (analysed pixel by pixel) of a spectrum imaging across the SnO_2|Sn/:Fe_2O_3 interface shown in Fig. 1a. For simplicity, the same supervision was applied for each pixel here, including the edge integration window (500–525 eV to avoid overlapping with O-K), the background model (power law),

and the background fitting window (450–480 eV). Adjusting these parameters for individual pixels is not only exhaustive, but prone to bias. Due to the moderate dose, the spectrum at each pixel is noisy, as shown in Fig. 2a. Consequently, the power law modelling of the background returns noisy parameters, as shown in Fig. 2b, and hence the signal integral is also noisy. There is no finer detail in

the signal integral than the higher amount found in SnO_2 with respect to Sn/:Fe_2O_3.

Noise reduction through PCA is a well-established routine in the EELS community, which has been implemented in the Digital Micrograph software as the MSA plug-in [25]. Application of PCA routine to the dataset returns a scree plot, as shown in Fig. 2c, which aligns the spectral
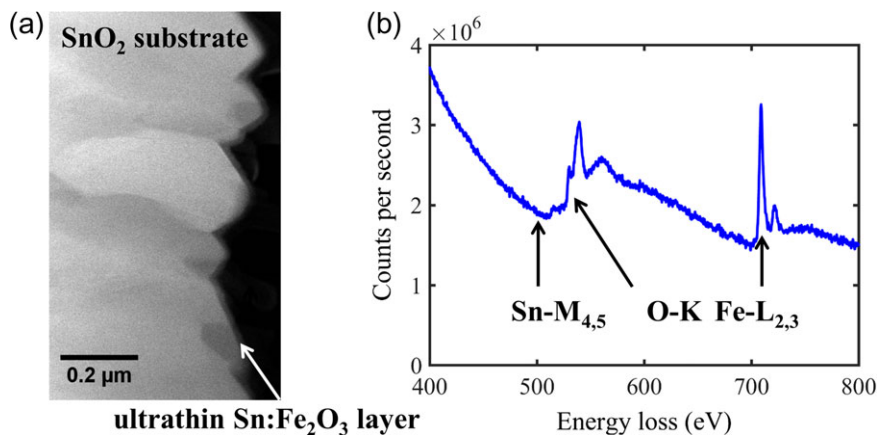
Fig.1. (a) HAADF-STEM micrograph of an ultrathin Sn:Fe_2O_3 photoanode on the fluorine doped SnO_2 substrate. (b) The sum core loss spectrum of an EELS spectrum imaging.
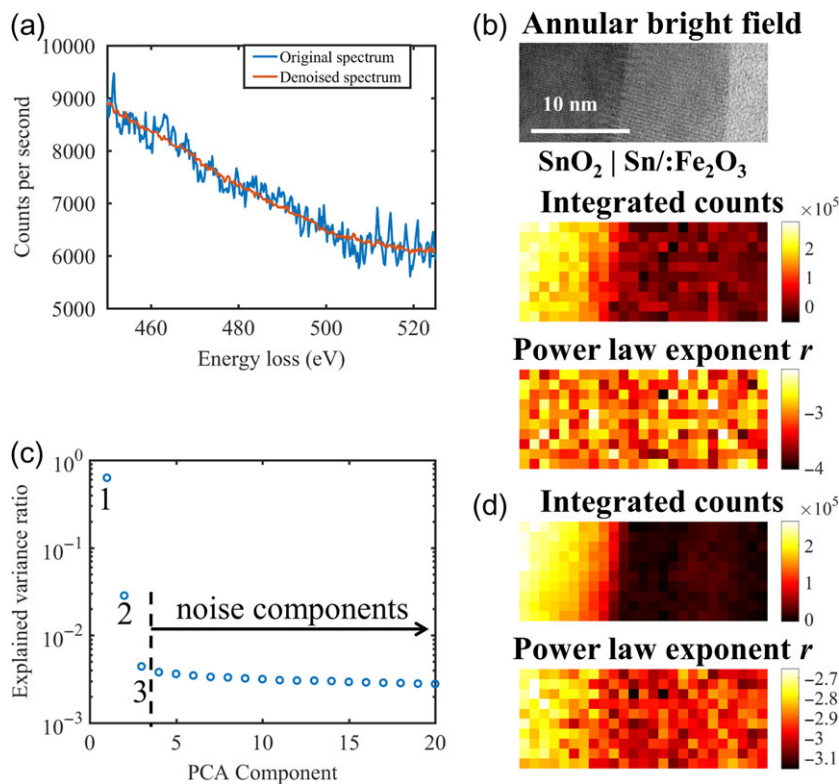
Fig. 2. (a) An EELS spectrum from the spectrum imaging dataset before and after the PCA denoise routine, (b) the area for spectrum imaging, the integrated counts of the Sn-$M_{4,5}$ edge, and the power law exponents for background modelling (pixel size is 1 nm), (c) PCA scree plot of the spectrum imaging to identify the threshold level of noise and (d) maps of the integrated counts and the power law exponents as in Fig. 2b after the PCA denoise routine.

components with respect to their explained variance. It is clear that two components stand out explaining the most variance from the dataset, whereas the third component is slightly higher than the following components showing a continuous decay of explained variance from the dataset. In this case, we treat the components starting from the fourth as noise components and set their $c_{p,i}$ coefficients to 0 to reconstruct the dataset from the first three components, resulting in a denoised spectrum imaging. Figure 2a shows a denoised spectrum at the same pixel to compare with the original one. As shown in Fig. 2d, the power law background fitting on the denoised spectrum imaging returns less noisy parameters and the signal integral is consequently clearer. It is evident from the denoised signal integral that the top half of the Sn/: Fe_2O_3 film has more Sn than the bottom half. The different Sn doping level in the Sn/: Fe_2O_3 film (top half doped, bottom half nominally undoped) was engineered in the atomic layer deposition process and can only be verified in this example after the PCA denoise routine.

By a sparse representation of an EELS spectrum, MVA analysis not only enables signal identification and noise removal, but also offers the opportunity to supervise on the individual components instead of individual pixels. This can be expressed by the following equation,

$$\sum_E I_p(E) = \sum_E \sum_i c_{p,i} I_i(E) = \sum_i c_{p,i} \sum_E I_i(E) \qquad (2)$$

where $I_p(E) = f_p(E) - f_p^{BG}(E)$ is the signal of the absorption edge obtained after background subtraction of the EELS spectrum $f_p(E)$ at each pixel, and $I_i(E) = f_i(E) - f_i^{BG}(E)$ is the signal of the absorption edge for each spectral component. As shown by Eq. (2), summation of the signal counts at individual pixels $I_p(E)$ over an energy range is a linear operation that can be done by linear combinations of such summation of the spectral components $I_i(E)$. As a result, to evaluate the signal counts, supervision on background modelling can benefit from the dimension reduction from all pixels $f_p^{BG}(E)$ to all spectral components $f_i^{BG}(E)$.

Physically, both the EELS signal and background are expected to be positive or 0. Moreover, it is intuitive to decompose the spectrum datasets into non-negative coefficients. The NMF approach decomposes the spectrum imaging into non-negative coefficients and non-negative spectral components, and thereby facilitating the interpretation of individual components. With the development of MVA algorithms, the options for decomposing EELS datasets are ever increasing [19,20]. The decomposed spectral components may all serve as input for further processing. In the following examples, we use the NMF algorithm as described in the scikit learn tool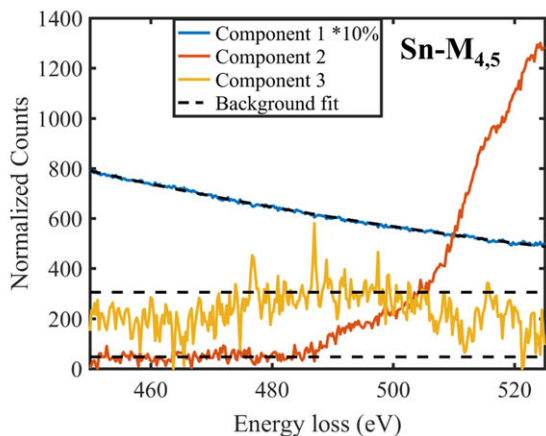kit [26], which is also used in the open source Hyperspy software [27]. We then focus on the post processing of the MVA spectral components to extract quantitative values for EELS analysis. The following evaluation routines are written in a Matlab package that offers a guided way to process EELS spectrum imaging datasets [28].

We continue with the aforementioned example to show the evaluation of Sn-$M_{4,5}$ edge counts from component-based analysis and background modelling. In addition to the dataset on the $SnO_2|Sn/:Fe_2O_3$ interface, another dataset on the $SnO_2|Fe_2O_3$ interface is included that contains spectra from pure Fe_2O_3 to compare with. The combined datasets are decomposed by NMF into spectral components (Fig. 3a) and their corresponding coefficient maps (Fig. 3b). The spectral components are normalized with respect to their corresponding coefficient maps that are set to have an average value of 1. The NMF component 1 has higher coefficients on the sample and follows a power law profile, which reflects the background. Component 2 has higher coefficients where Sn is abundant and reveals the profile of Sn-$M_{4,5}$ edge with a delayed onset. Component 3 has a noisy origin and does not show spectral characteristics from any edges of interest.
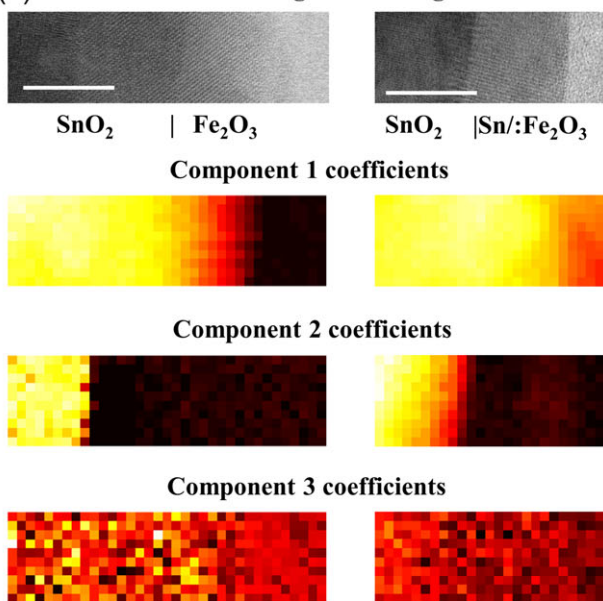
Going beyond the qualitative description, background modelling on individual components is required for quantitative analysis. Fitting the background of component 1 is straightforward, as it follows a power law (in the form of aE^{-r}) with exponent $r = 3.14$ over the modelled energy range. The signal integration for this component is 0. It is obvious that components 2 and 3 do not follow a power law function within any range of energy loss. Although, linear combination of power law functions has been used to model EELS backgrounds [29], such combinations do not return a power law function. In order to quantify the signal integral of the Sn-$M_{4,5}$ edge, the background of components 2 and 3 is modelled by functions that can be combined linearly, for example, linear functions. As shown in Fig. 3a, the background of component 2 was modelled as a constant before the edge onset, whereas the background of component 3 was also chosen as a constant before the signal integral window to minimize the effect of the preceding feature. The signal integrals of components 2 and 3 are calculated, and those of individual pixels are reconstructed using Eq. (2). The resulting evaluation of signal integral is very close to the evaluation from individual pixels, as shown by the line profile in Fig. 3c. The bottom half of the Sn/:Fe_2O_3 film (right spectrum imaging) has the same level of Sn counts as the pure Fe_2O_3 film (left spectrum imaging) and the vacuum area. Therefore, the amount of Sn between SnO_2 and the top half of Sn/:Fe_2O_3 is below the detection limit, in accordance with the designed film doping profile. In the top part of the Sn/:Fe_2O_3 film (right spectrum imaging), a Sn signal is detected. In this example, component-based
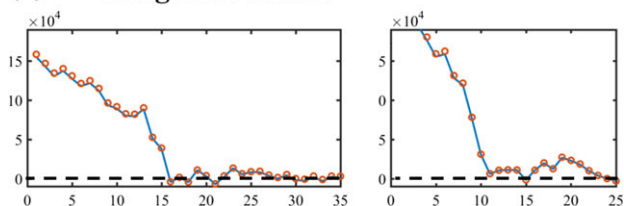
(a) Spectral components

(b) Annular bright field image

SnO$_2$ | Fe$_2$O$_3$ SnO$_2$ |Sn/:Fe$_2$O$_3$

Component 1 coefficients

Component 2 coefficients

Component 3 coefficients

(c) Integrated counts

Fig. 3. (a) Background fitting as a power law function $e^{28.14} \times E^{-3.14}$ for component 1, and as constants 47.09 and 305.07 for components 2 and 3, respectively, (b) areas of the spectrum imaging and their respective coefficient maps (scale bars are 10 nm, pixel size is 1 nm), and (c) the signal integration of a line profile derived from Eq. (2) (open circles) as compared with the values from fitting on individual pixels (lines). Dashed lines are eye guides for the baseline (0 integrated counts).

supervision reproduces the quantification from the conventional (pixel-based) routine, while the supervision on the background model for three components replaces the supervision on 650 pixels.

Despite having only three spectral components, supervision on the background subtraction is far from straightforward, as there is no definite function to model. In the example above, the background is modelled by the linear combination of a power law function and two constants, as shown in Fig. 3a. The power law exponent $r = 3.14$ of component 1 is close to the maximum exponent among individual pixels, as shown in Fig. 2d. Adding constants to such a background effectively reduces the power law exponent. In order to quantify the difference between power law functions, we make use of the fact that the multiplication of two power law functions returns a power law function. Instead of decomposing the spectrum imaging dataset into spectral components, the spectrum imaging can be factorized into spectral factors,

$$f_p(E) = \prod_i f_i(E)^{c_{p,i}}, \text{ or } \ln f_p(E) = \sum_i c_{p,i} \ln f_i(E) \quad (3)$$
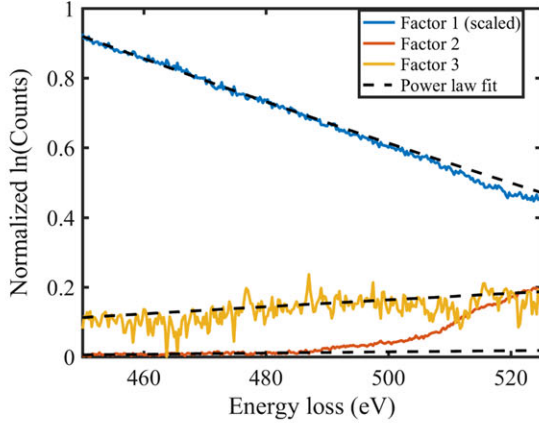
In order to realize factorization by means of linear MVA algorithms, a logarithmic operation needs to be taken, as formulated in Eq. (3). In comparison with Eq. (1), $f_p(E)$ and $f_i(E)$ are replaced by their logarithmics. The practical implementation includes taking the logarithmic of EELS spectrum $\ln f_p(E)$ as input for MVA algorithms, and taking the exponents of the output $\ln f_i(E)$ as spectral factors.

As an example, the EELS spectrum imaging shown in Fig. 3 was taken the logarithmic operation and underwent NMF decomposition to factorize into spectral factors (Fig. 4a) and their corresponding coefficient maps (Fig. 4b). As shown in Fig. 4a, background modelling on the spectral factors $\ln f_i(E)$ becomes more straightforward using a linear fit against $\ln(E)$ before the edge onset. In order to reconstruct the power law background of individual pixels, the form of the power law function is put into Eq. (3), which is then rearranged into Eq. (4),
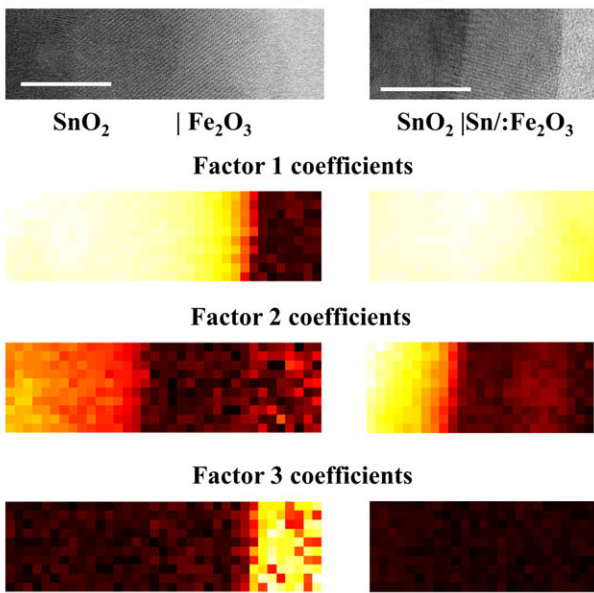
$$\ln(a_p E^{-r_p}) = \sum_i c_{p,i} \ln(a_i E^{-r_i}), \text{ where } \ln a_p$$
$$= \sum_i c_{p,i} \ln a_i, \text{ and } r_p = \sum_i c_{p,i} r_i \quad (4)$$

where a_i and r_i are the prefactor and the exponent of the power law function for each spectral factor, respectively, and a_p and r_p are the power law parameters at each pixel. In this example, the component-based supervision allows for a selection of different windows for the power law fitting of the background. As shown in Fig. 4a, spectral factors $\ln f_i(E)$ for $i = 1, 2$ are linear with respect to $\ln(E)$ in the window between 450 and 480 eV, before the Sn-M$_{4,5}$ edge onset. Factor 3, on the other hand, is not linear in that range, as a dip is observed around 465 eV, and the fitting window is chosen between 475 and 495 eV. From the
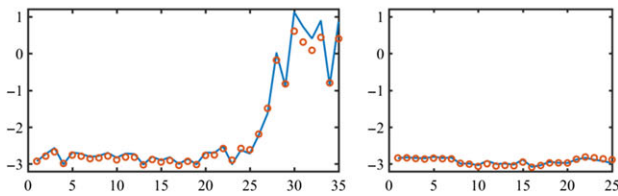
(a) Spectral factors

(b) Annular bright field image

SnO_2 | Fe_2O_3 SnO_2 |Sn/:Fe_2O_3

Factor 1 coefficients

Factor 2 coefficients

Factor 3 coefficients

(c) Power law exponent r

Fig. 4. (a) Background fitting parameters for the three spectral factors (factor 1 scaled down by $\ln(2000) = 7.6$): $(a_1, r_1) = (26.30, 2.91)$, $(a_2, r_2) = (-2.83, -0.48)$, $(a_3, r_3) = (-0.50, -0.08)$, (b) areas of the spectrum imaging and their respective coefficient maps (scale bars are 10 nm, pixel size is 1 nm), (c) the power law exponent of a line profile reconstructed from Eq. (4) (open circles) as compared with the values from fitting on individual pixels (lines).

coefficient map in Fig. 4b, it is shown that factor 3 mainly contributes to the area beyond the thin film (the glue line and vacuum) of the left spectrum imaging, which does not relate to the Sn-$M_{4,5}$ edge. The power law exponents for each pixel are then reconstructed from the three factors and

their coefficient maps according to Eq. (4), as shown in Fig. 4c. A comparison there with a pixel by pixel power law fitting using a fixed window shows a good agreement between the two approaches. Fitting the background from the spectral factors offers a more straightforward supervision by offering the same function of power law (or a linear function in the logarithmic form, see Fig. 4a) and the flexibility on the fitting window.
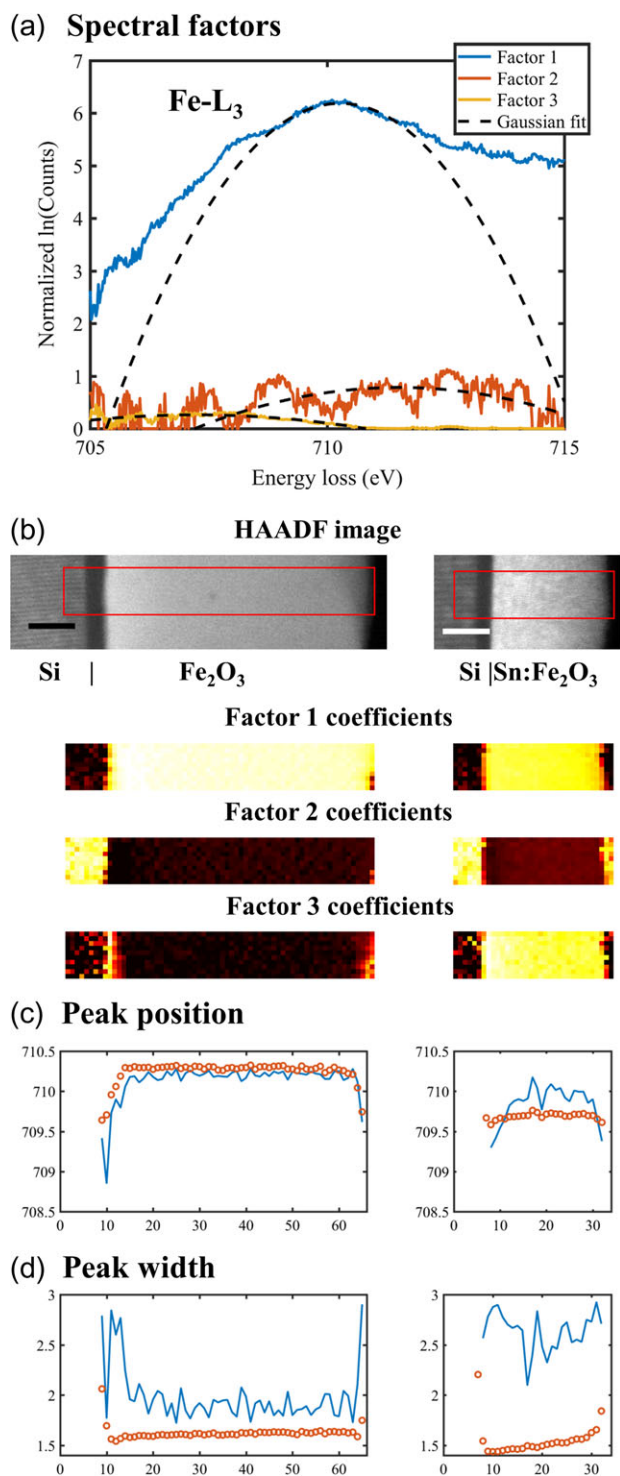
Besides modelling the power law background, the logarithmic formulation Eq. (3) can be useful to study other mathematic functions that can be factorized. For example, the Gaussian shape function can be modelled using the following expression,

$$\ln\left(a_p e^{-\frac{(E-\mu_p)^2}{2\sigma_p^2}}\right) = \sum_i c_{p,i}\ln\left(a_i e^{-\frac{(E-\mu_i)^2}{2\sigma_i^2}}\right), \text{ where}$$

$$\sigma_p^{-2} = \sum_i c_{p,i}\sigma_i^{-2}, \text{ and } \mu_p = \frac{\sum_i c_{p,i}\mu_i\sigma_i^{-2}}{\sum_i c_{p,i}\sigma_i^{-2}} \tag{5}$$

where a_i, μ_i and σ_i are the height, the centre, and the width of the Gaussian shape function for each spectral factor, respectively, and a_p, μ_p and σ_p are the Gaussian shape parameters at each pixel, which can be reconstructed using Eq. (5). We use the Fe-L_3 peak as a case study for Gaussian fitting to determine the chemical shift. As shown in Fig. 5b, two datasets are combined together, representing a scan across a Si|Fe_2O_3 interface for the left dataset and a scan across a Si|Sn:Fe_2O_3 interface for the right. The background of the Fe-L_3 peak was modelled by supervising individual spectral factors and reconstructing using Eq. (4), as introduced in the preceding example. Then, the Fe-L_3 peak signal is subjected to the NMF decomposition in the logarithmic formulation Eq. (3) into the spectral factors (Fig. 5a) and their corresponding coefficient maps (Fig. 5b).

A quadratic fitting on each spectral factor is performed to model a_i, μ_i and σ_i of the Gaussian peak. In comparison with the centre of factor 1, factor 2 represents a shift to the right ($\mu_2 > \mu_1$), whereas factor 3 represents a shift to the left ($\mu_3 < \mu_1$). It is shown in Fig. 5d that factor 1 comes from undoped and doped Fe_2O_3 of both spectrum imaging, factor 2 mainly from areas without Fe_2O_3, and factor 3 from the homogeneously doped Sn:Fe_2O_3 film (right spectrum imaging) as well as the interface and surface of undoped Fe_2O_3 film (left spectrum imaging). A chemical shift of the Fe-L_3 edge correlates to the bonding environment and especially the oxidation state of Fe atoms. As hematite (undoped Fe_2O_3) only consists of Fe^{3+}, a chemical shift to lower energy loss indicates a lower oxidation state of Fe, for example, reduction to Fe^{2+}[30].

(a) Spectral factors

(b) HAADF image

Factor 1 coefficients

Factor 2 coefficients

Factor 3 coefficients

(c) Peak position

(d) Peak width

Fig. 5. (a) Gaussian fitting parameters for the 3 factors $(a_1, \mu_1, \sigma_1) = (6.187, 710.274, 1.403)$, $(a_2, \mu_2, \sigma_2) = (0.783, 711.519, 3.447)$, $(a_3, \mu_3, \sigma_3) = (0.262, 707.236, 4.935)$, (b) areas of the spectrum imaging and their respective coefficient maps (scale bars are 10 nm, pixel size is 1 nm); (c) the centre and (d) the width of a line profile derived from Eq. (5) (open circles) are compared with the values from Gaussian fitting of individual pixels (lines).

A reconstruction of the Gaussian peak parameters using Eq. (5) then returns a left shift of the peak at the pixels where factor 3 has large coefficients, as shown by a line

profile in Fig. 5c. The quantified Gaussian parameters are compared with those fitted for individual pixels in Fig. 5c and 5d, where a discrepancy is evident. This is because modelling the spectral factors (Fig. 5a) as Gaussian functions has a big discrepancy. For example, factor 1 has a noticeable deviation from a Gaussian shape, showing a broader profile. As a result, the width derived from spectral factors is lower than the width modelled at individual pixels, as shown in Fig. 5d. Moreover, factor 2 has several peak features, whereas the modelled peak only takes their enveloped shape into account. Despite the conventional modelling of the white line features by Gaussian peaks, a spectrum imaging may be impossible to be factorized into Gaussian peaks. Indeed, many more types of functions have been applied to model a white line feature to model the chemical shift [30].

In order to propose a universal way to model the peak position, we return to the linear formulation Eq. (1). Instead of doing peak fitting, we approximate the centre and the width of the peak profiles by their statistical mean and variance, as shown by the following expressions,
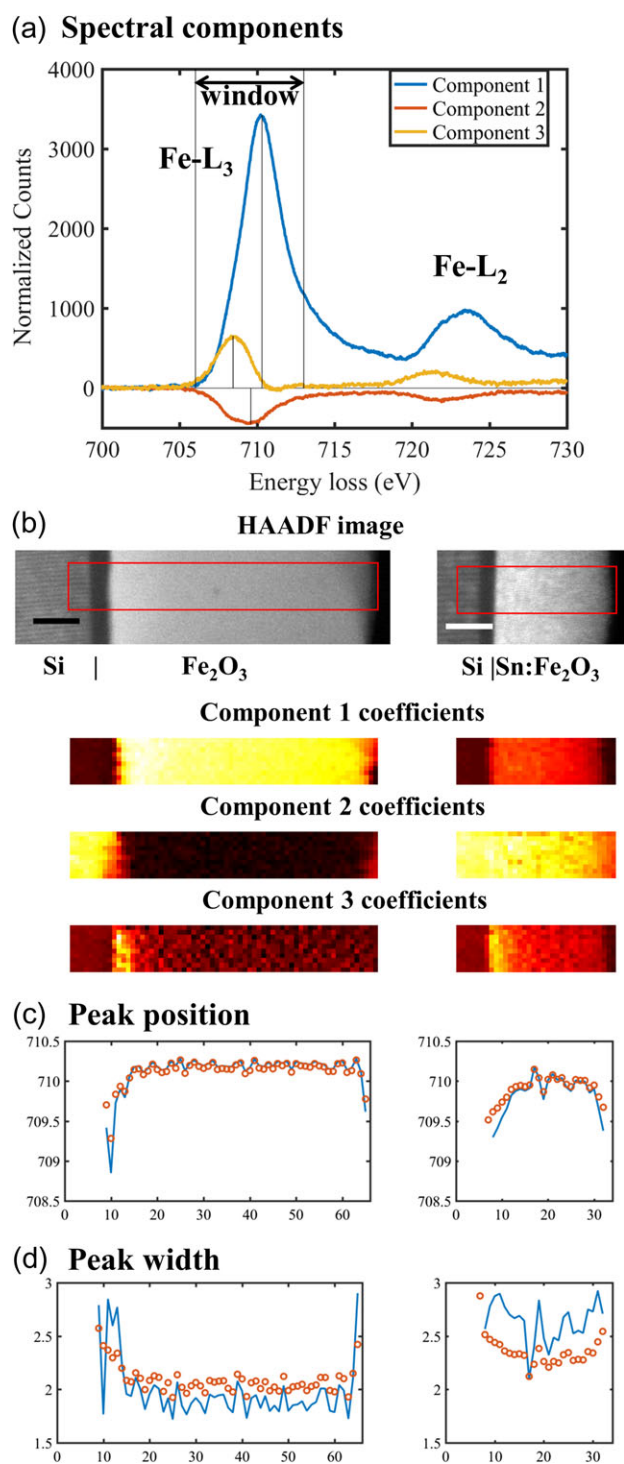
$$
\begin{aligned}
\mu_p &= \frac{\sum_E E I_p(E)}{\sum_E I_p(E)} \\
&= \frac{\sum_E E \sum_i c_{p,i} I_i(E)}{\sum_E \sum_i c_{p,i} I_i(E)} = \frac{\sum_i c_{p,i} \sum_E E I_i(E)}{\sum_i c_{p,i} \sum_E I_i(E)}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
\sigma_p^2 &= \frac{\sum_E (E - \mu_p)^2 I_p(E)}{\sum_E I_p(E)} = \frac{\sum_E E^2 I_p(E)}{\sum_E I_p(E)} - \left(\frac{\sum_E E I_p(E)}{\sum_E I_p(E)} \right)^2 \\
&= \frac{\sum_i c_{p,i} \sum_E E^2 I_i(E)}{\sum_i c_{p,i} \sum_E I_i(E)} - \frac{(\sum_i c_{p,i} \sum_E E I_i(E))^2}{(\sum_i c_{p,i} \sum_E I_i(E))^2}
\end{aligned}
\tag{7}
$$

where μ_p and σ_p are the statistical mean and variance at each pixel, and the other variables are defined in Eq. (2). Similar to Eq. (2), individual terms in the numerators and denominators of Eqs. (6) and (7) can be expressed as linear combination of quantities from the spectral components, including their first and second moments. Unlike the previous example (Fig. 5) where each spectral factor is fit to a Gaussian shape, the spectral components discussed here do not need to have a Gaussian shape to derive their moments for the reconstruction of the peak shape parameters at each pixel using Eqs. (6) and (7).

As an example, the Fe-L_3 peak signal from the preceding example underwent NMF decomposition in the linear formulation Eq. (1) to compute the spectral components (Fig. 6a) and their corresponding coefficient maps (Fig. 6b). The background of each spectral component is modelled by power law and linear fitting introduced in the example of Fig. 3. Figure 6a displays the spectral components after background subtraction, where negative values may show

(a) Spectral components

(b)

HAADF image

Si | Fe₂O₃   Si |Sn:Fe₂O₃

Component 1 coefficients

Component 2 coefficients

Component 3 coefficients

(c) Peak position

(d) Peak width

Fig. 6. (a) Spectra of the three components after background subtraction with supervised window selection, (b) areas of the spectrum imaging and their respective coefficient maps (scale bars are 10 nm, pixel size is 1 nm); line profiles of the Fe-L₃ (c) peak position and (d) peak width derived from individual components (open circles) are compared with the values from Gaussian fittings of individual pixels (lines).

up (e.g. component 2). The supervision on each spectral component is necessary to find an energy window so that the resulting centre of intensity coincides with their respective

peak positions, as shown in Fig. 6a. By using this approach, there is a much improved numerical agreement on the peak position with the Gaussian analysis at individual pixels, as shown in Fig. 6c. Moreover, as shown in Fig. 6d, the peak width reconstructed from the variance of individual components has the same trend as the Gaussian width determined from individual pixels. Component-based supervision on the window selection can be a reliable way to reconstruct the position (chemical shift) and the width of a peak feature and replace the supervision and curve fitting for individual pixels.

Conclusion

In summary, we have employed MVA analyses to reduce the dimension of EELS spectrum imaging datasets and thereby facilitating the supervision for further evaluation. This has been employed to model the integral counts as well as the centre and width of peak features. Using the logarithmic formulation introduced to factorize the dataset, parameters of power law background can be reconstructed, and deviation from a Gaussian peak is visualized. These utilities are implemented in a Matlab package. More utilities to model and analyse EELS spectrum imaging datasets can benefit from component-based supervision facilitated by MVA analysis.

Acknowledgements

Funding

References

1. Trebbia P, and Bonnet N (1990) EELS elemental mapping with unconventional methods I. Theoretical basis: image analysis with multivariate statistics and entropy concepts. *Ultramicroscopy* 34: 165–178.

2. Sigle W (2005) Analytical transmission electron microscopy. *Annu. Rev. Mater. Res.* 35: 239–314.

3. Maigné A, and Twesten R D (2009) Review of recent advances in spectrum imaging and its extension to reciprocal space. *J. Electron. Microsc.* 58 (3): 99–109.

4. Bosman M, Watanabe M, Alexander D T L, and Keast V J (2006) Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images. *Ultramicroscopy* 106: 1024–1032.

5. Lozano-Perez S, de Castro Bernal V, and Nicholls R J (2009) Achieving sub-nanometre particle mapping with energy-filtered TEM. *Ultramicroscopy* 109: 1217–1228.

6. Muto S, Yoshida T, and Tatsumi K (2009) Diagnostic nanoanalysis of materials properties by multivariate curve resolution applied to spectrum images by S/TEM-EELS. *Mater. Trans.* 50: 964–969.

7. de la Pena F, Berger M-H, Hochepied J-F, Dynys F, Stephan O, and Walls M (2011) Mapping titanium and tin oxide phases using EELS: An application of independent component analysis. *Ultramicroscopy* 111: 169–176.

8. Duchamp M, Lachmann M, Boothroyd C B, Kovács A, Haug F-J, Ballif C, and Dunin-Borkowski R E (2013) Compositional study of defects in microcrystalline silicon solar cells using spectral decomposition in the scanning transmission electron microscope. *Appl. Phys. Lett.* 102: 133902.

9. Eljarrat A, López-Conesa L, López-Vidrier J, Hernández S, Garrido B, Magén C, Peiró F, and Estradé S (2014) Retrieving the electronic properties of silicon nanocrystals embedded in a dielectric matrix by low-loss EELS. *Nanoscale* 6: 14971–14983.

10. Sugar J D, Gabaly F E, Chueh W C, Fenton K R, Tyliszczak T, Kotula P G, and Bartelt N C (2014) High-resolution chemical analysis on cycled LiFePO$_4$ battery electrodes using energy-filtered transmission electron microscopy. *J. Power Sources* 246: 512–521.

11. Rossouw D, Langelier B, Scullion A, Danaie M, and Botton G A (2016) Multivariate-aided mapping of rare-earth partitioning in a wrought magnesium alloy. *Scripta Mater.* 124: 174–178.

12. Yedra L, Eljarrat A, Arenal R, Pellicer E, Cabo M, López-Ortega A, Estrader M, Sort J, Baró M D, Estradé S, and Peiró F (2012) EEL spectroscopic tomography: Towards a new dimension in nanomaterials analysis. *Ultramicroscopy* 122: 12–18.

13. Nicoletti O, de la Peña F, Leary R K, Holland D J, Ducati C, and Midgley P A (2013) Three-dimensional imaging of localized surface plasmon resonances of metal nanoparticles. *Nature* 502: 81–84.

14. Haberfehlner G, Orthacker A, Albu M, Li J, and Kothleitner G (2014) Nanoscale voxel spectroscopy by simultaneous EELS and EDS tomography. *Nanoscale* 6: 14563.

15. Hu X, Sun Y, and Yuan J (2008) Multivariate statistical analysis of electron energy-loss spectroscopy in anisotropic materials. *Ultramicroscopy* 108: 465–471.

16. Tatsumi K, and Muto S (2009) Local electronic structure analysis by site-selective ELNES using electron channeling and first-principles calculations. *J. Phys. Condens. Matter.* 21: 104213.

17. Bonnet N, Brun N, and Colliex C (1999) Extracting information from sequences of spatially resolved EELS spectra using multivariate statistical analysis. *Ultramicroscopy* 77: 97–112.

18. Bonnet N, and Nuzillard D (2005) Independent component analysis: A new possibility for analyzing series of electron energy loss spectra. *Ultramicroscopy* 102: 327–337.

19. Shiga M, Tatsumi K, Muto S, Tsuda K, Yamamoto Y, Mori T, and Tanji T (2016) Sparse modeling of EELS and EDX spectral imaging data by nonnegative matrix factorization. *Ultramicroscopy* 170: 43–59.

20. Spiegelberg J, Rusz J, Thersleff T, and Pelckmans K (2017) Analysis of electron energy loss spectroscopy data using geometric extraction methods. *Ultramicroscopy* 174: 14–26.

21. Dobigeon N, and Brun N (2012) Spectral mixture analysis of EELS spectrum-images. *Ultramicroscopy* 120: 25–34.

22. Verbeeck J, Van Aert S, and Bertoni G (2006) Model-based quantification of EELS spectra: Including the fine structure. *Ultramicroscopy* 106: 976–980.

23. Egerton R F (2011) *Electron Energy-Loss Spectroscopy in the Electron Microscope*, (Springer, US). doi: 10.1007/978-1-4419-9583-4.

24. Dunn H K, Feckl J M, Müller A, Fattakhova-Rohlfing D, Morehead S G, Roos J, Peter L M, Scheu C, and Bein T (2014) Tin doping speeds up hole transfer during light-driven water oxidation at hematite photoanodes. *Phys. Chem. Chem. Phys.* 16: 24610–24620.

25. Watanabe M, Okunishi E, and Ishizuka K (2009) Analysis of spectrum-imaging datasets in atomic-resolution electron microscopy. *Microsc. Anal.* 23: 5–7. https://www.hremresearch.com/Eng/plugin/MSAEng.html. Please also refer to the HREM Research Inc. website.

26. Lin C-J (2007) Projected gradient methods for non-negative matrix factorization. *Neural. Comput.* 19: 2756–2779. http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.NMF.html. Please also refer to the scikit learn website.

27. de la Peña F, Ostasevicius T, Fauske V T, Burdet P, Jokubauskas P, Nord M, Sarahan M, Johnstone D N, Prestat E, Taillon J, Caron J, Furnival T, MacArthur K E, Eljarrat A, Mazzucco S, Migunov V, Aarholt T, Walls M, Winkler F, Martineau B, Donval G, Hoglund E R, Zagonel L F, Garmannslund A, Gohlke C, and Chang H-W (2017) hyperspy/hyperspy v1.2 [Data set]. *Zenodo.* http://doi.org/10.5281/zenodo.345099.

28. The Matlab code is updated at sites.google.com/site/siyuanresearch.

29. Cueva P, Hovden R, Mundy J A, Xin H L, and Muller D A (2012) Data processing for atomic resolution electron energy loss spectroscopy. *Microsc. Microanal.* 18: 667–675.

30. van Aken P A, and Liebscher B (2002) Quantification of ferrous/ferric ratios in minerals: new evaluation schemes of Fe L$_{23}$ electron energy-loss near-edge spectra. *Phys. Chem. Miner.* 29, 188–200.