Ultramicroscopy 125 (2013) 35–42

Contents lists available at SciVerse ScienceDirect

Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Statistical consequences of applying a PCA noise filter on EELS spectrum images

Stijn Lichtert, Jo Verbeeck*

EMAT, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

ARTICLE INFO

Article history:
Received 11 June 2012
Received in revised form
17 October 2012
Accepted 21 October 2012
Available online 27 October 2012

Keywords:
EELS
Spectrum imaging
Principal component analysis
Cramér–Rao lower bound

ABSTRACT

Principal component analysis (PCA) noise filtering is a popular method to remove noise from experimental electron energy loss (EELS) spectrum images. Here, we investigate the statistical behaviour of this method by applying it on a simulated data set with realistic noise levels. This phantom data set provides access to the true values contained in the data set as well as to many different realizations of the noise. Using least squares fitting and parameter estimation theory, we demonstrate that even though the precision on the estimated parameters can be better as the Cramér–Rao lower bound, a significant bias is introduced which can alter the conclusions drawn from experimental data sets. The origin of this bias is in the incorrect retrieval of the principal loadings for noisy data. Using an expression for the bias and precision of the singular values from literature, we present an evaluation criterion for these singular values based on the noise level and the amount of information present in the data set. This criterion can help to judge when to avoid PCA noise filtering in practical situations. Further we show that constructing elemental maps of PCA noise filtered data using the background subtraction method, does not guarantee an increase in the signal to noise ratio due to correlation of the spectral data as a result of the filtering process.

1. Introduction

Before the era of aberration-corrected electron microscopes, typical EELS experiments consisted of a few dozen spectra with integration times up to several seconds for core loss spectra. This method was limited experimentally by specimen drift and total acquisition time if applied to higher resolution experiments (nm scale). However with improved aberration-corrected microscopes with high brightness guns, the limitation has shifted from the detection method towards the sample [1–3]. Due to high beam currents focused on a very small region of the sample, acquisition times are considerably reduced (several ms for core loss spectra) and often limited by beam damage. Improved stability in terms of sample drift, electromagnetic shielding and faster acquisition facilitate the ability to collect more spectra. EELS data cubes with 100×100 pixels and a CCD spectrometer range of 2048 bins are no exception. This increase in the amount of data, often with poor counting statistics, demands a different approach in the retrieval of information. Typically several basis spectra are repeated throughout the sample and small variations are the main subject of interest. A few examples include atomic resolution bond structure determination in graphene [4], valence mapping in oxides [5–10] and single atom spectroscopy [11]. More applications of EELS in aberration-corrected microscopes are mentioned in recent review papers [12,13] and by Egerton [14].

To exploit this repetitive nature of the EELS data cube, multivariate statistical analysis techniques have been used. Principal component analysis (PCA) is the most widely used technique, and has been applied on EELS data sets to improve the signal to noise ratio of data sets with low counting statistics [15,16]. The mathematical basis of this technique lies in the singular value decomposition of the EELS data cube D. Here we will follow the notation introduced by Keenan and Kotula [17]. In EELS we unwrap two spatial dimensions to obtain the data matrix D with dimension $p \times N_E$, where p is the total amount of spatial positions and N_E the amount of pixels in each spectrum. The SVD algorithm allows the decomposition of this matrix into

$$D = U\Sigma V^T = TV^T, \tag{1}$$

where singular values Σ provide an ordering for the rows of T (scores) and V (loadings). We can approximate the data set by retaining only the principal components with singular values above some threshold. The number of retained components is most often determined by the scree plot, where only components with a singular value above the scree slope are used for reconstruction [5,14,18].

Additional steps have been introduced to overcome problems caused by the Poisson nature of the noise in EELS experiments.

* Corresponding author. Tel.: +32 32653249; fax: +32 32653318.
E-mail address: jo.verbeeck@ua.ac.be (J. Verbeeck).

0304-3991/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ultramic.2012.10.001

A weighting method with the average spectrum and image has been proposed by Keenan et al. [17] to treat high and low signals on equal footing. In this case, we rescale D to a weighted version:

$$\tilde{D} = gDh, \tag{2}$$

with g and h diagonal matrices containing the inverse of the square root of the average image and average spectrum in D. The SVD decomposition then becomes

$$\tilde{D} = \tilde{U}\tilde{\Sigma}\tilde{V}^T. \tag{3}$$

This method has been exploited successfully on high resolution EELS data resulting in atomic resolution elemental maps [19–21], even when the SNR before PCA was not sufficient to distinguish atom positions in all elemental maps [22,23].

The main goal of PCA in EELS applications is denoising the data, thus eliminating part of the original signal after reconstruction with a limited set of principal components. The rational behind this approach is that the retained components contain the signal while the rejected components contain only noise. This assumption is however too optimistic as it was recently shown by Cueva et al. [24] that considerable artefacts can be introduced even when taking more components than indicated by the scree plot.

In order to clarify the fundamental reasons for these artefacts, we will investigate the influence of PCA data reduction on the relevant parameters in an EELS spectrum imaging experiment, e.g. a chemical concentration map. Instead of experimental data sets, we will use a computer generated data set as this allow us to create many different noise realizations of the same experiment while the 'true' noise free spectrum is known. This allows us to accurately separate signal and noise and to study how the noise filtered result differs from the true signal. We make use of the theoretical prediction for such bias by Faber et al. [29] as a criterion to judge when PCA should be avoided. We will also study the mechanism why the SNR improvement of elemental maps is not as substantial as can be expected as reported also by Dudeck et al. [22]. Compared to examples on experimental data, our numerical approach has the benefit of creating a much better understanding of why and when artefacts occur in denoising EELS data with PCA.

2. Results

2.1. Example: hexagonal boron nitride simulation

A series of phantom EELS data sets of a hexagonal boron nitride structure was generated based on HAADF and STEM-EELS observations of similar structures [4,25]. In this way we know the true parameters (elemental concentrations) of the data and we can do repetitive experiments to retrieve the statistical behaviour of PCA application. The EELS spectra for the K edges of both boron and nitrogen were calculated using FEFF as described in [26]. A data set with 30×26 pixels represents four sheets of stacked single boron nitride sheets with a sampling step of 0.24 Å/pixel. In the simulated volume (Fig. 1) all atomic columns, except for the labelled ones, contain two boron and two nitrogen atoms. Extra atoms were added at positions 1 (extra nitrogen) and 2 (extra boron). The weight maps of the elements are generated using Gaussian peaks centred on the atomic positions with a height relative to the atomic content and equal width for boron and nitrogen. A constant power-law background and normally distributed noise with a standard deviation σ was assumed. Representative EELS spectra for both high and low noise standard deviations of $\sigma = 2 \times 10^2$ and $\sigma = 5 \times 10^1$ are shown in Fig. 2. For comparison, the mean pixel value of the average spectrum is 1.2×10^3. In order to test weighted PCA (wPCA), we use Poisson noise instead with an average of 34 and 545 electrons per pixel, leading to a similar signal to noise ratio as for the normally distributed noise. In the remainder of this paper we will simply label these two noise levels as 'low noise' and 'high noise' referring to Gaussian noise for PCA results and Poisson noise for wPCA results.

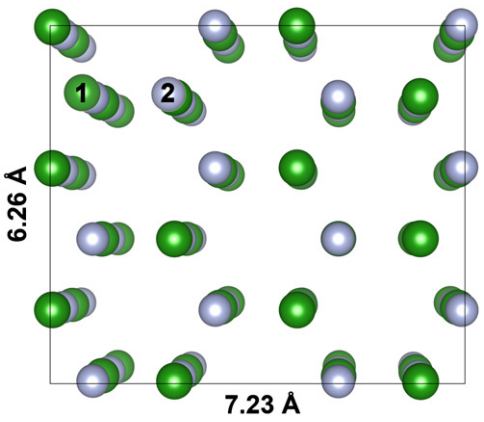

Fig. 1. The simulated data set consists of a stacked sequence of hexagonal boron nitride sheets. All atomic columns contain both two B and N atoms, except for the columns labelled by 1 and 2. At these atomic positions, an extra N or B atom is added respectively.

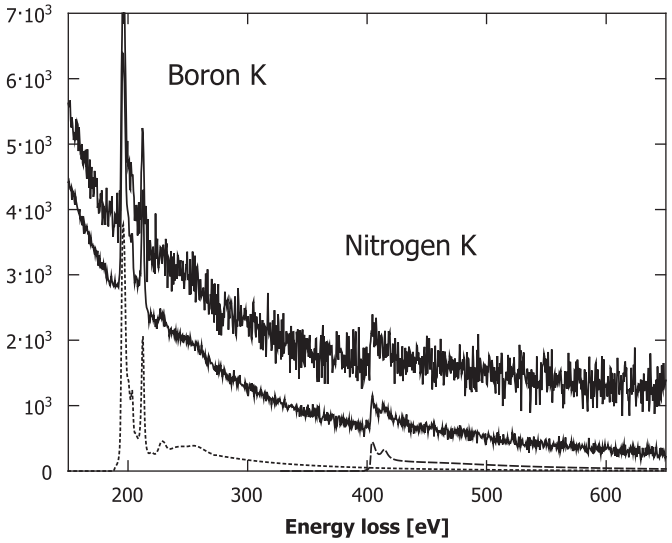

Fig. 2. A data set based on calculated spectra for the boron K (dotted line) and nitrogen K (dashed line) edge was used for simulations. Typical spectra are shown containing normally distributed noise with a standard deviation σ of 5×10^1 and 2×10^2. The latter was shifted by 10^3 for clarity.

We apply both PCA and wPCA on these phantom EELS data sets in order to study the potential introduction of artefacts in the reconstructed data. The nitrogen elemental map is calculated for two different noise levels by least squares fitting of model spectra (Fig. 3) to the filtered data. An elemental map on a single noise realization of the untreated phantom data set is shown in the first row of Fig. 1 indicating the effect of noise on the N elemental map. In order to study the systematic bias we average over 20 different noise realizations of the same phantom. The result for the untreated data set is shown in the second row of Fig. 3 showing an identical average output for the N map, indicating that no bias is present. In other words, if the elemental map is created by least squares fitting with the correct model (we know the model in this case) the result is unbiased as expected from the properties of a maximum likelihood fitter. If however we use PCA or wPCA

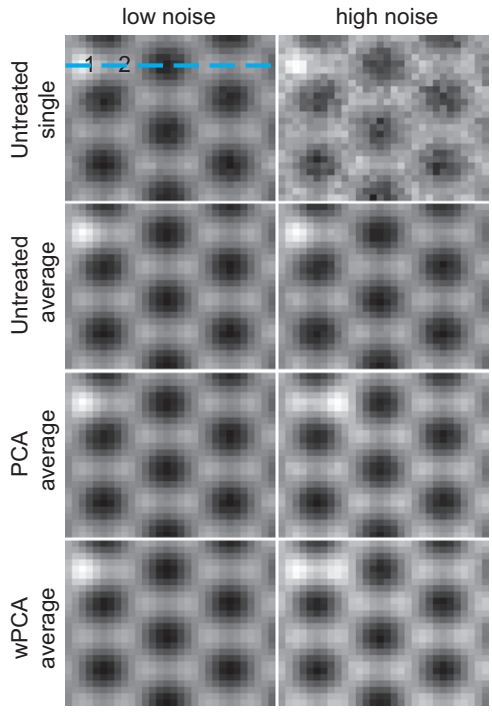

Fig. 3. N elemental maps obtained from least squares fitting results for two different noise levels. The first row shows the result using untreated data for a single realization of the noise. The second row shows the average result for 20 repeated noise realizations giving a better idea of the bias. The next two lines show the average results on PCA and weighted PCA treated data retaining 10 components. For the low noise case (left column) no clear differences between the raw (second row) and the reconstructed data with either PCA or wPCA is observed indicating that the use of PCA is appropriate here. A significant bias however is observed for the atomic columns labelled by 1 and 2 for the high noise data (right column). The elemental map for the averaged raw data correctly indicates a higher N concentration at position 1, this is not reproduced by the reconstructed data set for PCA (position 2 is higher than 1) and neither for the weighted PCA result (position 2 is similar to 1). In both cases, PCA introduces a bias that would lead to the wrong conclusion about the defect of interest in the data set. Clearly, the use of PCA and weighted PCA is not appropriate here. The fact that the reconstructed maps visually improved as the noise was reduced which makes this artefact even more dangerous in practical cases.

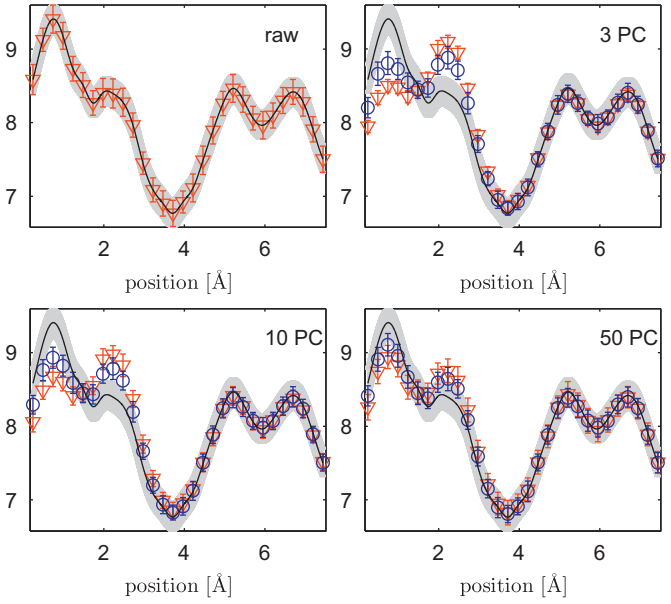

Fig. 4. The average value and standard deviation for the N coefficient was obtained from 100 reconstructed data sets using least squares fitting. The estimates for the raw data (left panel, red diamonds) are unbiased and their standard deviation attains the CRLB (grey band). The estimates for the data reconstructed with either PCA (red diamonds) or weighted PCA (blue circles) by three, ten or fifty components (3 PC, 10 PC, 50 PC), show a significant bias with respect to the true values (black line). The variance is in all case lower than the CRLB (grey band) and increases with increasing number of reconstructing components. Note that increasing the number of components only gradually reduces the bias at the expense of increased standard deviation. Even 50 components are not enough to faithfully reconstruct the noisy data set which intrinsically has only three components. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

on the data, we do a nonlinear pretreatment of the data which can introduce bias. The averaged results for PCA and wPCA retaining 10 components is shown in rows 3 and 4 of Fig. 3. For the low noise case we get the correct result showing an extra N atom in position 1. For the high noise case however we note a significant bias. For the PCA treated data, extra N signal appears on position 2 rather than position 1 while the wPCA results in extra N signal on both position 1 and 2. This example demonstrates how significant artefacts are introduced when applying PCA and derived algorithms exactly in the area of interest. It is important to note that both PCA and weighted PCA reduce the noise in the resulting map making the map to appear visually better, but the above example shows that the treatment can entirely change the information content in the data. This is clearly a situation that should be avoided at all cost when applying PCA on an experimental data set.

In order to get a deeper understanding of this bias effect we select a line trace through the columns of interest indicated by a line in Fig. 3 for the high noise case. This allows us to also show the standard deviation of each point in the elemental map over 100 noise realizations as an indicator of the noise level depending on the amount of components retained after (weighted) PCA reconstruction in Fig. 4. The standard deviation of the untreated data (Fig. 4, left panel) equals the best possible standard deviation for this noise level given by the Cramér–Rao lower bound (CRLB) shown as a grey band obtained from Eq. (A.5). This confirms the most precise property of the least squares fitter as an unbiased estimator for normally distributed noise.[1]

First, three principal components were used to reconstruct the data set, as the underlying model consists of three different components: the background, the boron and nitrogen K-edge . The estimates (Fig. 4, second panel) for this reconstructed data set show a significant bias with respect to the true values for both PCA and weighted PCA. In all pixels the variance is lower than the CRLB which is impossible for an unbiased estimator. This shows that the noise is reduced at the expense of an introduced bias.

Retaining more components than the three suggested from the scree plot (Fig. 6) does not significantly improve the situation as is shown in the third and fourth panel of Fig. 4 for 10 and 50 retained components. We observe that the noise reduction benefit over the untreated data reduces while the bias only gradually decreases. This is a crucial observation as it shows that we cannot correct the biasing problem by taking a few more components as is often assumed in practice.

In order to further elucidate the essence of this behaviour we will focus on PCA treatment only as it requires no prescaling of the data and the Gaussian noise allows to make the connection with previous theoretical results. Nevertheless, the principle of the introduced bias is the same for weighted PCA even though it may accept slightly more noise before breaking down in the way we will describe next.

As a first step towards understanding the bias behaviour we plot the loadings V and scores T for the first four components in both the low and high noise situation in Fig. 5. We can understand

[1] More information about the estimation model and the CRLB can be found in Appendix A.

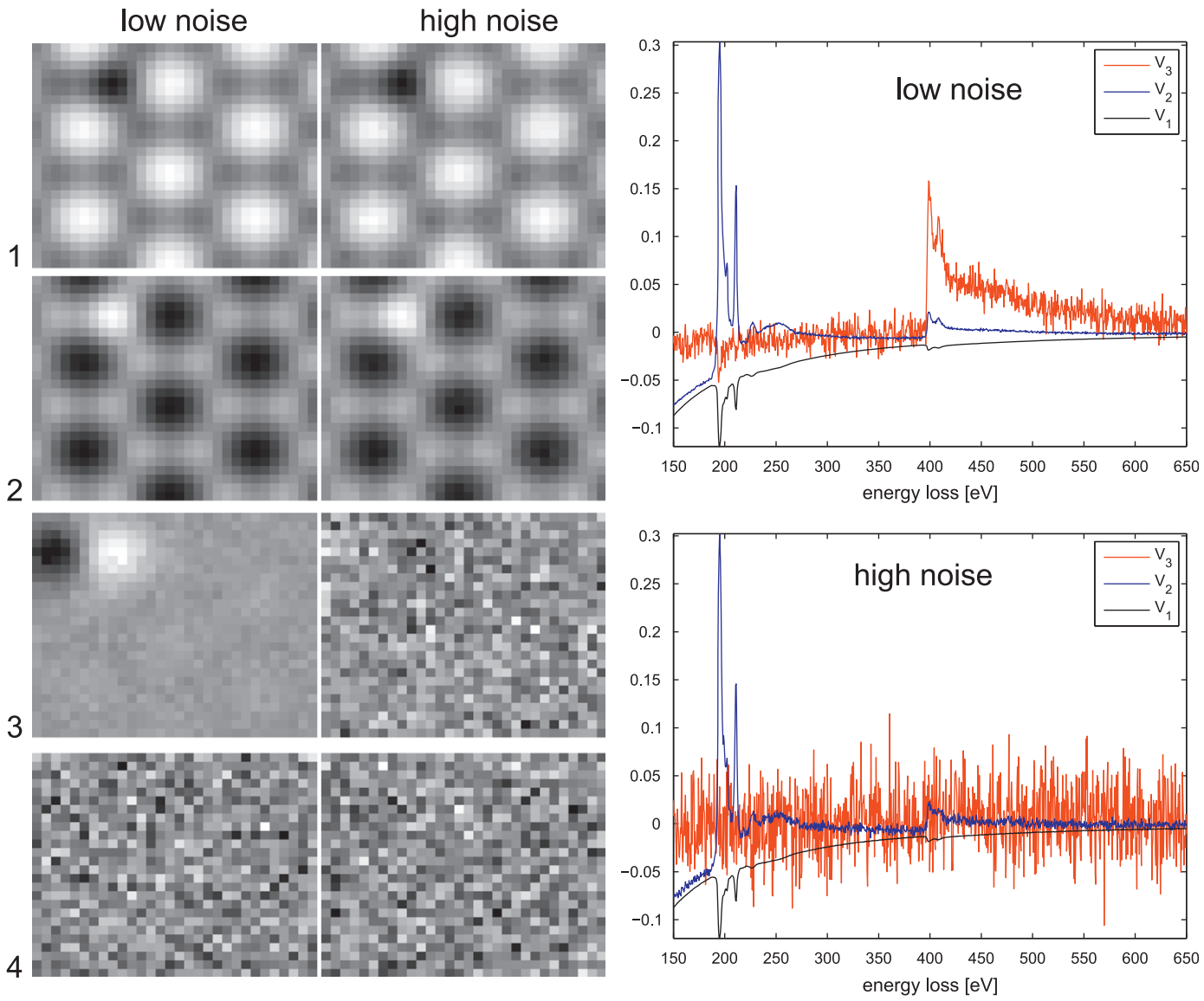

Fig. 5. Scores (images left) and loadings (spectra right) for the first four components in the low and high noise situation. This image reveals how the three components present in the image are correctly retrieved in the low noise case, while the third component gets corrupted in the high noise case as the information of the third component is spread out over all remaining components.

the score images for the low noise case qualitatively as follows. The first score image indicates columns with an average BN spectrum consisting of two B and two N atoms and a suppression of intensity in position 2. The second score image corrects position 2 to have three B and three N atoms and the third score image further corrects position 1 to have an extra N and position 2 to reduce the number of N atoms by one. Component 4 consists of noise as expected from the fact that the phantom data only contains three components. In the high noise case however, we observe similar score images and loadings except for component 3 which contains only noise. The missing of the information from the third component is the direct reason for the introduction of bias as a very important part of the information is suddenly missing. Indeed, missing this component would encode for two N and two B atoms in all columns except position 2 where there would be three B and three N as encoded in the second component. This would wrongly lead to the information that position 2 has an extra N atom as seen in the average output for the N elemental map in Fig. 3 for high noise and PCA. It is also clear from the score images that this missing information in component 3 is not simply redistributed to the fourth component as this component is also dominated by noise. This explains why slightly increasing the number of components does not take away the bias effect. The above observations seem to indicate that there is a noise level beyond which suddenly the third component is no longer correctly retrieved. This threshold behaviour can be understood better by taking a look at the singular values as shown in the scree plot (Fig. 6) which indeed would indicate only two relevant components for the high noise case and three for the low noise case. This means that the reconstruction can only retrieve these two components correctly even though the data consisted of three components and therefore a significant part of information is missing, leading to bias when the noise level gets higher.

2.2. Scree plot and singular values

Singular values are mainly used as a cutoff determination when denoising EELS data. Using properties of the SVD, they give more information on the underlying variation present in the data set. The scree plot is often given as a justification for the amount of components retained in the reconstructed data set. The distinction is made on the observation of a plateau of singular

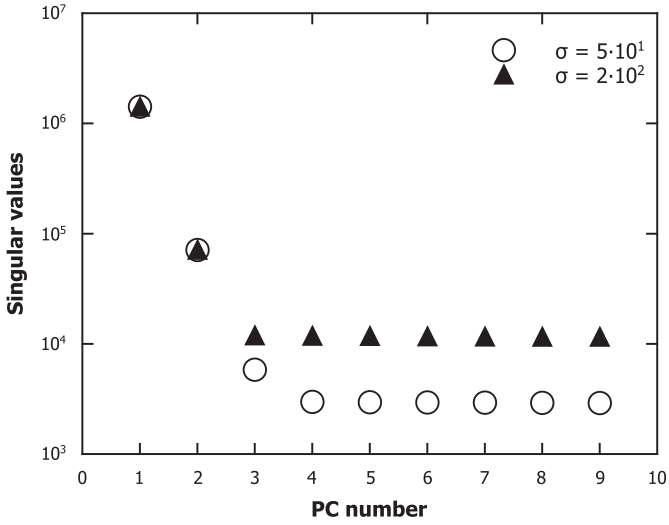

Fig. 6. The scree plot for both noise levels is given. The singular values of components referred to as noise components will lie on a straight line on the logarithmic plot. The low noise data (open circles) displays three relevant number of components, whereas for the high noise case (triangles), only two are separated from the line of noise components.

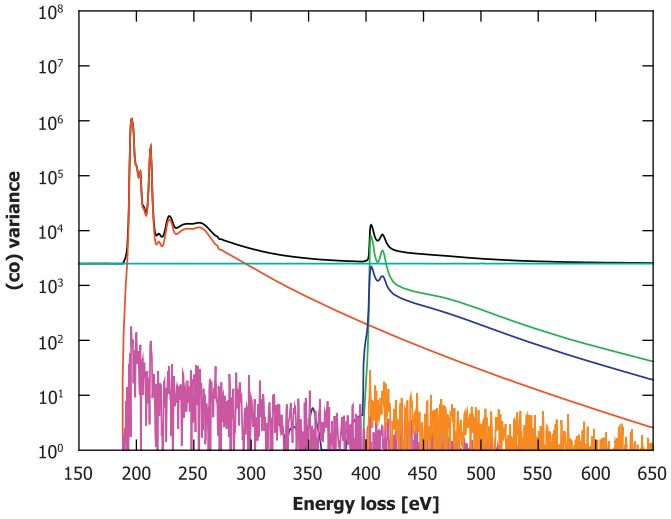

Fig. 7. Total variance of every EELS slice of the data set and the different contributions. The noise variance is a constant value under the given normally distributed noise (horizontal line, cyan). After every edge onset, the largest part of the total variance (black) is due to the variance of that element (boron=red, nitrogen=green). This implies that the images at those energy losses are closely related to the corresponding elemental map. (Other lines shown are covariances between B and N (blue), noise and B (magenta) and noise and N (orange).) (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this article.)

values indicating noise components and a limited number of singular values clearly above this line which are labelled as relevant components. The latter are then used for reconstruction of the original data set, removing a large part of the noise [27]. Using the orthonormality properties of the U and V matrices of the SVD, we can rewrite the projection of the data set D onto the orthonormal basis set given by the principal scores [28]:

$$DV = (U\Sigma V^T)V = U\Sigma = T \tag{4}$$

The projections are given by the principal loadings and represent the weight of a principal loading in every spectrum of the data set. We are now interested in the variance matrix T^TT of these projections

$$T^TT = (U\Sigma)^T(U\Sigma) = \Sigma^T\Sigma \tag{5}$$

The variance of the ith principal loading is given by the square of the ith singular value. As the singular value matrix Σ is diagonal, all covariances between different principal loadings are zero. Using this knowledge we can make another observation about the scree plot (Fig. 6). We note a difference in singular values between the low and high noise data set from the third component onward. This means that the variance of the associated principal loading is different. All singular values starting from the fourth component are biased with respect to the singular values obtained from a noise free data set. In the noise free case, the full data set is contained within the first three components, leading to only three singular values being non-zero. Adding noise however, changes this situation dramatically as two effects occur. First of all the noise will make all singular values non-zero. But on top of this, a bias in the singular values results in signal information being spread out over more than three components. Note that the presence of a bias in the variance of the principal loadings does not necessarily introduce bias in the estimates for the elemental contributions. In the low noise simulation, e.g. the singular values are biased but the associated elemental map represents the true spatial distribution of nitrogen (Fig. 3). For the high noise case this is no longer true and an observable bias effect is introduced also in the spatial distribution of the elements.

In order to demonstrate this point more clearly we calculate the contribution to the variance in the principal loading image from both the random noise and the signal given by the (co) variances of the spatial variation of the different elements. As the underlying model is known for the phantom data, we are able to separate both contributions. We start by calculating the variance contributions of the original EELS data cube (Fig. 7). The variance definition is written in the appendix and represents the amount of variation in a single slice of the EELS data cube at energy loss E by noise or by the spatial variation of a certain chemical element. The noise variance is constant as expected for uncorrelated Gaussian noise. The signal variance peaks at energies following an edge onset as the variance is dominated by the variance of that distinct element.

Viewing PCA as a transformation of the data set to a new basis we expect that also the variance is redistributed upon this transformation. Indeed, the new basis set is given by the principal loadings V while the projections are given by the principal scores T in Eq. (4). When the principal loadings V_{NF} are obtained from a noise free data set, they consist of a linear combination of the model spectra for the background and boron and nitrogen K-edge . We can see that applying these principal loadings on the low noise data leads to principal scores which are a superposition of the signal contained in the first three components and a constant contribution from the uncorrelated noise (Fig. 8, left panel). The total variance of the noise contribution to each component equals the total noise variance. In case we obtain the principal loadings from the noisy data we obtain an estimate V of the noise-free principal loadings V_{NF}. Using these estimated principal loadings for both low and high noise phantom data is represented in the middle and right panel (Fig. 8). We clearly see that the use of V instead of V_{NF} leads to a different contribution of noise and signal variance in each component and that the signal contribution is no longer bound to only the first three components but rather spread out over all components. This fact is the essence why noise filtering using PCA introduces bias as together with the removal of noise, also signal bearing components are discarded. This happens even for data sets that have an intrinsically limited number of components as the data set used here and becomes worse with increasing noise level.

This knowledge may act as a warning on the use of PCA in denoising EELS data sets as the data can be significantly altered at high noise levels. From the example we showed that the bias in the estimates does not vanish when using more components in the reconstruction process as the signal gets distributed over all

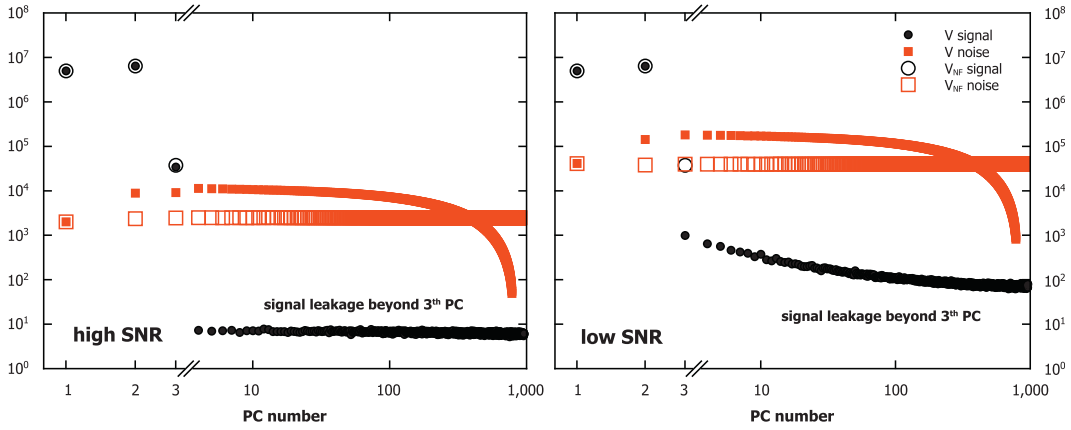

Fig. 8. The variance of every score image is split into a pure signal (dots) and noise (squares) contribution. The noise corrupted data is projected on the loadings obtained from the noisy data (V, closed symbols) and from the noise free data (V_{NF}, open symbols). In the high SNR case (left panel) all first three signal contributions (black dots) are very similar indicating good reconstruction. They show however a significant difference in the low SNR situation for the third PC (right panel). In both figures, the signal contribution originating from projection on the realistic loadings is non-zero in the so-called noise components (line of black dots from fourth PC).

components due to the inevitable use of an estimated principal loading. The appearance of bias on the singular values due to random noise has already been described by Faber et al. [29]. They derived an expression for the bias $b_{\Sigma_{a,a}}$ and the variance $\sigma_{\Sigma_{a,a}}$ of the singular values. The bias on the singular values is expressed as [29, Eq. (13)]

$$b_{\Sigma_{a,a}} = E[\Sigma_{a,a}] - \Sigma_{a,a,NF} \tag{6}$$

With E the expectation value and $\Sigma_{a,a,NF}$ the singular values of the noise free data. The variance on the singular values is equal to the variance on the original data as the random noise variance is equally distributed over the transformed data set [29, Eq. (4)]:

$$\sigma_{\Sigma_{a,a}} = \sigma \tag{7}$$

The expression for bias then reads according to Faber et al. [29, Eq. (14)]:

$$b_{\Sigma_{a,a}} \approx \frac{1}{2}(p+N_E-A-1)\frac{\sigma^2}{\Sigma_{a,a,NF}} \approx \frac{p+N_E}{2}\frac{\sigma^2}{\Sigma_{a,a,NF}} \tag{8}$$

with A the number of independent signal components. The last approximation for the bias was made under the realistic assumption for EELS experiments that the number of pixels p and/or the number of energy intervals N_E is much larger than A. In Table 1 we calculate these values for the data set with high and low noise for the first principal components and compare them with the simulated results. We calculate the bias either making use of the noise free $\Sigma_{a,a,NF}$ (which is unavailable in experiments) or replace it by the estimated $\Sigma_{a,a}$ obtained from the noisy data sets. The simulated bias values agree well with the calculated bias using the true values, whereas using estimated values underestimates the bias in the singular values. In real experiments, unfortunately, only these estimated singular values are accessible.

Using the expressions by Faber et al. we can provide two evaluation criteria for the scree plot.

1. The bias of every singular value is lower than α times its variance

$$b_{\Sigma_{a,a}} < \alpha\sigma_{\Sigma_{a,a}} \Rightarrow \Sigma_{a,a} > \frac{p+N_E}{2\alpha}\sigma \tag{9}$$

2. The bias of every singular value is lower than β times its value

$$b_{\Sigma_{a,a}} < \beta\Sigma_{a,a} \Rightarrow \Sigma_{a,a} > \sqrt{\frac{p+N_E}{2\beta}}\sigma \tag{10}$$

Table 1
Bias in the first few singular values of the simulated boron nitride structure with two noise levels $\sigma = 2\times10^2$ and $\sigma = 5\times10^1$, $p=780$, $N_E=1024$, $A=3$. The low noise simulation is each time shown as the top row.

Bias calculation method	PC1	PC2	PC3
Bias calculated from [29] with the true values (%)	0.00 0.00	0.04 0.71	7.67 122.77
Bias calculated from [29] with the estimated values (%)	0.00 0.00	0.04 0.70	6.62 25.24
Bias calculated from simulation (%)	0.00 0.00	0.04 0.70	7.62 120.54

With α and β parameters to be chosen depending on the confidence one requires. If the number of components satisfying these criteria is lower than the amount of elements for which a different spatial variation is present, bias in these estimates can be expected. Note that we interpret an *element* here in a wider sense than a chemical element. Every atomic contribution with a different ELNES fingerprint (due to valence state, chemical environment) is considered as a distinct element.

2.3. Variance maximization and SNR of elemental maps

The improvement in the SNR of individual EELS slices after PCA application can be explained by the variance maximization of the SVD. The result of the background subtraction method for two energy windows is shown in Fig. 9 for both the raw data set and for the reconstruction using three principal components. A value of 5×10^1 was chosen for the standard deviation of the Gaussian noise. The nitrogen elemental maps are calculated using the background subtraction method. To correctly estimate the background under high noise conditions, a power law was fitted using EELSMODEL [30]. This method ensures that no prior knowledge about the power-law coefficients was used even though we know the values for A and r in our phantom data set. An elemental map generated using a narrow energy window shows a significant improvement after the use of PCA noise filtering with three components (top row). The mechanism behind the improvement in the SNR of the elemental map is the variance maximization property of the SVD. The variance contributions provide estimates for the SNR of a single EELS slice or

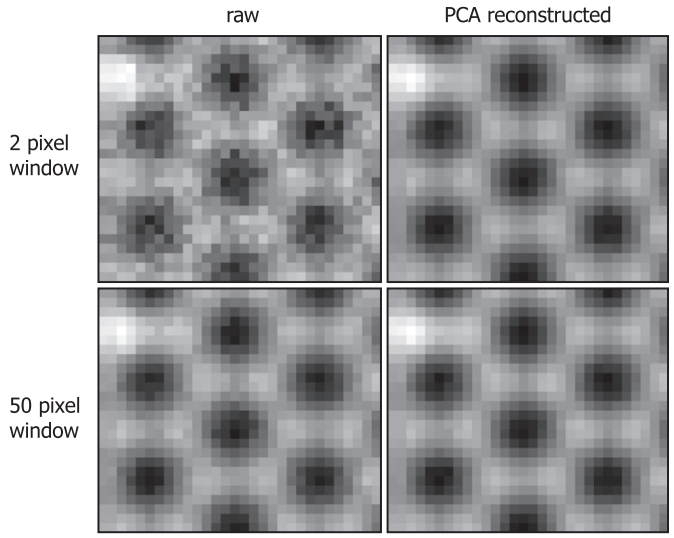

Fig. 9. Comparison of elemental maps before (left column) and after the application of PCA (right column). The data was reconstructed using three principal components. The nitrogen maps were generated by power law background subtraction fitted using EELSMODEL in the energy interval [349.5–399.0 eV] and integration in a small energy window of 1 eV [404.0–404.5 eV] (top row) or a large energy window of 25 eV [402.0–426.5 eV] (bottom row). We clearly see that the SNR of the elemental map at the small energy window is largely improved by principal component analysis. As the quality of the map increases significantly with the original data, this does not occur with the PCA reconstructed data.

principal score image as both the variance for the signal (spatial variation of chemical elements) and noise is known for the phantom data set. The PCA loadings (Fig. 8) show a SNR which is significantly higher than any slice in the original EELS data cube (Fig. 7). As every reconstructed slice is nothing more than a linear combination of the chosen principal loadings, the SNR of a single slice will be greatly enhanced. Variance maximization results in the image enhancement of individual slices in the EELS data set.

Using a broad integration energy window leads to similar SNR in the elemental maps for both raw and PCA treated data. The bottom row of Fig. 9 shows elemental maps generated with a 50 pixel integration window. The improvement after the use of PCA is not visible as opposed to the small window case. This is consistent with the observations by Dudeck et al. who did not find an improvement of the global SNR of elemental maps using optimized choice for integration and background windows with PCA application [22]. As expected the SNR of the elemental map using the raw data increases when a larger integration window is used. As noise in consecutive EELS slices is uncorrelated, the SNR will improve approximately with a factor of $\sqrt{n}$ when n images are summed.[2] After PCA noise filtering, the data is no longer uncorrelated as it is constructed from linear combinations of only a limited number of principal scores. Therefore, the noise will be similar for consecutive slices in the EELS data cube. The noise variance will scale linearly with the number of integrated images as will the (co) variances of underlying spectral variation. In general there is no improvement of these elemental maps with respect to non-treated data and bias will be introduced as discussed above.

3. Discussion

From our simulations, we can draw some conclusions on the application of PCA as a noise filter on real EELS data sets. The first conclusion is that treating data using PCA does not result in estimates with better precision and accuracy than raw data using maximum likelihood estimation. This was also mentioned by Lozano-Perez et al., by comparing MLLS fitting results [31]. Under high noise conditions, the variances of the estimated parameters can be lower than those expected by maximum likelihood estimation theory, but a bias will occur invalidating the results. This is consistent with the maximum likelihood estimator being an efficient estimator. Combining PCA noise filtering and maximum likelihood fitting deteriorates the situation to at best a consistent estimator on increasing number of components and decreasing noise level.[3] Using additional components in the reconstruction process both increases the variance and decreases the bias of the estimated parameter but the latter does not vanish for any choice of the number of retained components. Simulations based on Raman spectra previously showed the inability of PCA to correctly estimate the number of components with a high amount of added Gaussian noise although there was no direct link to the misinterpretation of the chemical content [32].

The reason for the bias in the estimates is the inaccurate principal loading retrieval when calculated from noisy data sets. For higher noise levels, a significant part of the signal is distributed over all components and therefore the reduction of components inevitably introduces bias as part of the signal is discarded. On top of this, signal components can have a variance lower than the variance of the noise in that component making the retrieval of that component impossible. In such case, the number of components will be erroneously estimated from the scree plot. Even if this is not the case noise will inevitably lead to bias of estimated relevant components when applying PCA denoising. Depending on the noise level this bias might be negligible as can be estimated by the criterion proposed here.

Elemental maps obtained with the conventional method can be visually improved using PCA in the case of a narrow integration window (both for signal as background region). However a bias in the estimated chemical concentration will be introduced and can lead to false conclusions about the data.

4. Conclusions

In this paper we investigated the statistical consequence of applying PCA noise filtering on a phantom data set mimicking a realistic experimental situation. We came to the conclusion that even though the application of PCA as a noise filter on EELS spectrum images can greatly reduce the variance in elemental maps even beyond the theoretical limit imposed by parameter estimation theory; a significant bias is introduced in the process. This bias can strongly influence the interpretation of the data as demonstrated by the retrieval of signal in the wrong atomic column of our phantom hexagonal BN test case. As there is no way to test the occurrence of such bias in real experimental conditions and the consequences can be severe, we used a criterion to estimate whether this bias will be significant for real experimental conditions. The criterion is based on a previously published expression for bias on the singular values based on the

[2] This statement assumes a constant value for the EELS spectrum at all n energy losses. As the selecting window is traditionally positioned around the maximum edge height, the improvement will be in general less than that predicted by the simple assumption of additive Gaussian noise.

[3] An *efficient* estimator has a mean value equal to the true value (unbiased) and a variance asymptotically attaining the CRLB (most precise). The estimated value of a *consistent* estimator converges to the true value with increasing number of measurements.

amount of noise, the number of independent components and the number of spatial and energy points.

Acknowledgments

S.L. acknowledges financial support of the Fund for Scientific Research-Flanders (FWO). J.V. acknowledges funding from the European Research Council under the 7th Framework Program (FP7), ERC Grant 246791 COUNTATOMS and ERC Starting Grant 278510 VORTEX. The authors acknowledge financial support from the European Union under the Seventh Framework Program under a contract for an Integrated Infrastructure Initiative. Reference no. 312483-ESTEEM2.

Appendix A. Parameter estimation theory

In this section we describe the methodology of the simulated data sets and the statistical analysis used to test the behaviour of PCA filtered data with respect to raw data. We will treat the observed intensities w_k^i at energy level k $(1,\ldots,N_E)$ and spatial position i $(1,\ldots,p)$ as stochastic variables ω_k^i. The joint probability distribution function $p(\omega^i)$ gives the probability for these observations to occur given an expression for the expectation value for w_k^i and the noise model. In this case we write the expectation value $f_k^i(\boldsymbol{\theta})=\mathbb{E}[w_k^i]$ by a linear combination of model spectra x_k^j $(j=1,\ldots J)$ with J the number of model spectra

$$f_k^i(\boldsymbol{\theta})=\sum_{j=1}^{J} A_j^i\, x_k^j \tag{A.1}$$

The parameters of the model are thus given by

$$\boldsymbol{\theta}=(A_1^i,\ldots,A_J^i)^T \tag{A.2}$$

The distribution of the observations w_k^i is Gaussian, resulting in following form for the joint probability distribution:

$$p(\omega^i)=\prod_{k=1}^{N_E}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(\omega_k^i-f_k^i(\boldsymbol{\theta}))^2\right) \tag{A.3}$$

Estimates $\hat{\boldsymbol{\theta}}$ for the parameters $\boldsymbol{\theta}$ are calculated by maximizing the likelihood function $p(\omega^i;\mathbf{t})$. In the case of an unbiased estimator, we can determine the minimum variance of every estimated parameter by the Fisher information matrix $F_{\boldsymbol{\theta}}$ which can be calculated under the given model for every spectrum [33]:

$$F_{\boldsymbol{\theta}}=-\mathbb{E}\left[\frac{\partial^2 \ln p(\omega^i)}{\partial}\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathbf{T}}\right]\rightarrow F_{jj'}=\frac{1}{\sigma^2}\sum_{k=1}^{N_E}x_k^j x_k^{j'} \tag{A.4}$$

The minimum variances for the parameters A_j^i in case of an unbiased estimation method are then given by the diagonal elements of the inverse of the Fisher matrix.

$$\mathrm{var}(A_j^i)\geq[F^{-1}]_{jj} \tag{A.5}$$

We clearly see from Eq. (A.4) that the minimum variance will be determined by the noise variance and the similarity between the model spectra. In a linear model as described above the estimates do not explicitly enter the expression of the Fisher information matrix. More details about the estimation method can be found in [30].

Appendix B. Variance decomposition

Given a certain model for the observed counts in an EELS spectral data set, we can calculate the different contributions to the variance of a single slice in the EELS data cube. At every energy loss E, the variance of the corresponding image will not only be determined by the elemental distribution of the data set, but also by the noise. We describe the EELS data set w_k^i using a normal distribution with the average equal to the expression for the expectation value and a variance σ^2

$$w_k^i=\mathcal{N}(\mathbb{E}[w_k^i],\sigma)=\mathcal{N}\left(\sum_{j=1}^{J}A_j^i\,x_k^j,\sigma\right) \tag{B.1}$$

We now define $\overline{w}_k=(w_k^1,\ldots,w_k^p)$ the vector with all pixel values at energy loss E_k and similarly $\overline{A}_j=(A_j^1,\ldots,A_j^p)$, the relative weight of model spectrum j at every pixel. $\overline{w}_k$ represents a single EELS slice of the total data cube which variance under this linear model is now given by

$$\mathrm{var}(\overline{w}_k)=\sum_{j=1}^{J}\mathrm{var}\left(\overline{A}_j\right)\left(x_k^j\right)^2+\sigma^2+2\sum_{j=1}^{J}\sum_{j'=1,j'>j}^{J}\mathrm{cov}\left(\overline{A}_j,\overline{A}_{j'}\right)x_k^j x_k^{j'} \tag{B.2}$$

References

[1] P.E. Batson, N. Dellby, O.L. Krivanek, Nature 418 (2002) 617–620.
[2] E. Okunishi, H. Sawada, Y. Kondo, M. Kersker, Microscopy and Microanalysis 12 (S2) (2006) 1150–1151.
[3] G.A. Botton, S. Lazar, C. Dwyer, Ultramicroscopy 110 (8) (2010) 926–934.
[4] K. Suenaga, M. Koshino, Nature 468 (2010) 1088–1090.
[5] M. Varela, M. Oxley, W. Luo, J. Tao, M. Watanabe, A. Lupini, S. Pantelides, S. Pennycook, Physical Review B 79 (2009) 085117.
[6] A.B. Shah, Q.M. Ramasse, X. Zhai, J.G. Wen, S.J. May, I. Petrov, A. Bhattacharya, P. Abbamonte, J.N. Eckstein, J.-M. Zuo, Advanced Materials 22 (2010) 1156–1160.
[7] H. Tan, S. Turner, E. Yücelen, J. Verbeeck, G. Van Tendeloo, Physical Review Letters 107 (2011) 107602.
[8] S. Turner, S. Lazar, B. Freitag, R. Egoavil, J. Verbeeck, S. Put, Y. Strauven, G. Van Tendeloo, Nanoscale (2011).
[9] D.A. Muller, L. Fitting Kourkoutis, M. Murfitt, J.H. Song, H.Y. Hwang, J. Silcox, N. Dellby, O.L. Krivanek, Science 319 (5866) (2008) 1073–1076.
[10] D.A. Muller, Nature Materials 8 (4) (2009) 263–270.
[11] M. Varela, S. Findlay, a. Lupini, H. Christen, A. Borisevich, N. Dellby, O. Krivanek, P. Nellist, M. Oxley, L. Allen, S. Pennycook, Physical Review Letters 92 (2004) 095502.
[12] C. Colliex, Journal of Electron Microscopy 60 Suppl 1 (2011) S161–S171 .
[13] G.A. Botton, MRS Bulletin 37 (2012) 21–28.
[14] R. Egerton, Electron Energy-Loss Spectroscopy in the Electron Microscope, third edition, Springer, New York, 2011.
[15] P. Trebbia, N. Bonnet, Ultramicroscopy 34 (1990) 165–178.
[16] N. Bonnet, Ultramicroscopy 77 (3–4) (1999) 97–112.
[17] M.R. Keenan, P.G. Kotula, Surface and Interface Analysis 36 (2004) 203–212.
[18] M.G. Burke, M. Watanabe, D.B. Williams, J.M. Hyde, Journal of Materials Science 41 (2006) 4512–4522.
[19] M. Bosman, V. Keast, J. García-Muñoz, A. D'Alfonso, S. Findlay, L. Allen, Physical Review Letters 99 (2007) 086102.
[20] J. García-Barriocanal, F.Y. Bruno, A. Rivera-Calzada, Z. Sefrioui, N.M. Nemes, M. Garcia-Hernández, J. Rubio-Zuazo, G.R. Castro, M. Varela, S.J. Pennycook, C. Leon, J. Santamaria, Advanced Materials 22 (2010) 627–632.
[21] S. Trasobares, M. López-Haro, M. Kociak, K. March, F. de La Peña, J.a. Perez-Omil, J.J. Calvino, N.R. Lugg, A.J. D'Alfonso, L.J. Allen, C. Colliex, Angewandte Chemie 123 (2011) 898–902.
[22] K.J. Dudeck, M. Couillard, S. Lazar, C. Dwyer, G.a. Botton, Micron 43 (2012) 57–67.
[23] T.J. Pennycook, M.P. Oxley, J. Garcia-Barriocanal, F.Y. Bruno, C. Leon, J. Santamaria, S.T. Pantelides, M. Varela, S.J. Pennycook, European Physical Journal—Applied Physics 54 (2011) 33507.
[24] P. Cueva, R. Hovden, J.A. Mundy, H.L. Xin, D.A. Muller, Microscopy and Microanalysis 18 (2012) 667–675.
[25] O.L. Krivanek, M.F. Chisholm, V. Nicolosi, T.J. Pennycook, G.J. Corbin, N. Dellby, M.F. Murfitt, C.S. Own, Z.S. Szilagyi, M.P. Oxley, S.T. Pantelides, S.J. Pennycook, Nature 464 (2010) 571–574.
[26] J. Verbeeck, G. Bertoni, Ultramicroscopy 109 (2009) 1343–1352.
[27] E.R. Malinowski, Factor Analysis in Chemistry, 2nd edition, Wiley-Interscience, 1991.
[28] I.T. Jolliffe, Principal Component Analysis, second ed., Springer-Verlag, New York, 2002.
[29] N.M. Faber, M.J. Meinders, P. Geladi, M. Sjöström, L.M.C. Buydens, G. Kateman, Analytica Chimica Acta 304 (1995) 257–271.
[30] J. Verbeeck, S. Van Aert, Ultramicroscopy 101 (2004) 207–224.
[31] S. Lozano-Perez, V. de Castro Bernal, R.J. Nicholls, Ultramicroscopy 109 (2009) 1217–1228.
[32] E. Levina, A.S. Wagaman, A.F. Callender, G.S. Mandair, M.D. Morris, Journal of Chemometrics 21 (2007) 24–34.
[33] B.R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, 2004.