

Improvement of Principal Component Analysis for Quantification of Spectrum-Imaging Datasets

M. Watanabe¹, A. Ishizuka², and K. Ishizuka²

¹ Dept of Materials Science and Engineering, Lehigh University, Bethlehem, PA, USA.

² HREM Research Inc., Higashimatsuyama, Saitama, Japan.

Acquisition of big spectrum-imaging (SI) datasets has become routine by using modern scanning transmission electron microscopes (STEMs) with fast signal detection systems through electron energy-loss spectrometry (EELS) and X-ray energy dispersive spectrometry (XEDS). Although the capability to acquire such big SI datasets is desired, it is more challenging to handle (or even to visualize/analyse) big datasets. For examples, identification of unknown features and extraction of dominant trends, which are just primary steps of data analysis prior to more detailed analysis including quantification, are simply harder if datasets are very large. Multivariate statistical analysis (MSA) is one of efficient approaches to analyse the large-scale datasets by automatically extracting features from the datasets [1, 2]. Principal component analysis (PCA), which is the one of the most popular MSA approaches and used as the initial step of more sophisticated MSA procedures, has been applied to analyse SI datasets for feature extraction from datasets (data mining) and to separate them from random noise (noise reduction) [e.g. 3]. Despite that the PCA approach is very efficient and useful, it may create unexpected artefacts especially in higher noise conditions [4]. The one of the most serious artefacts is the net signal reduction after applying the PCA noise reduction. This signal reduction becomes noticeable especially when the signal variation is very weak, at the noise level or lower, which might cause misleading results. Therefore, quantification may not be applicable safely after PCA noise reduction.

In general, there are two conceptual approaches to improve the PCA sensitivity: (1) reduction of random noise and (2) enhancement of true variations. The former can be achieved by modifying experimental conditions such as a use of higher currents and/or longer signal acquisitions. Unfortunately, the higher beam currents can end up with severe irradiation damages and longer acquisition might degrade datasets due to any kind of instabilities such as spatial and/or energy drifts. In addition, the former approaches useless for existing datasets acquired previously. Then, how can the weak but true variations in datasets be enhanced? The weak variations can be enhanced by modifying densities of information features in datasets, i.e. by dividing a dataset into small segments. We have implemented this data segmentation scheme into PCA, and developed (a) spatial local PCA and (b) spectral local PCA techniques (Fig. 1). Both techniques are implemented as plug-ins in the Gatan DigitalMigrograph platform.

Figure 2 show a set of Mn maps extracted from various Cu-Mn XEDS SI datasets at different processing stages. The original SI dataset was simulated with 5 Cu-5wt%Mn spectra in 10x10 pixels (rest is Cu), then Poisson noise was applied to the SI dataset. So, the Mn map from the simulated dataset (clean) is modified by Poisson noise addition (noised). Noted that the 5wt%Mn $K\alpha$ peak variation is similar to the Poisson noise level, and hence some artefacts might be induced by PCA. Then, conventional weighted PCA and spatial local PCA were applied to the noised dataset for noise reduction prior to Mn map extraction. Even after applying conventional weighted PCA, the Mn map is still noisy. Conversely, the spatial local PCA reduces noise in the Mn map but most of the pixels with Mn are preserved. This feature of local PCA can be more pronounced when a bigger SI dataset is processed. Figure 3 compares a set of nitrogen (N) maps

extracted from atomic-resolution EELS SI datasets of (0001)-projected Si_3N_4 , associated with a spectrum extracted from the original dataset. Influence of noise is severer at the higher energy-loss N K edge. Noise in the N map from the original dataset is reduced by conventional weighted PCA. However, the noise component is more effectively removed by spectral local PCA and detailed features reveal in the N map. Since background intensities vary significantly in EELS, spectral local PCA is more effective than spatial local PCA. Detailed features and advantages of the local PCA approaches will be discussed in this study [5].

References:

- [1] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, New York, (2002).
- [2] E.R. Malinowski, *Factor Analysis in Chemistry*, 3rd ed., Wiley, New York, (2002).
- [3] M Watanabe *et al*, *Microscopy and Analysis*, **23** (2009), p. 5.
- [4] S. Lichtert and J. Verbeeck, *Ultramicrosc.* **125** (2013), p. 35.
- [5] The author (MW) wishes to acknowledge financial support from the NSF through grants DMR-0804528 and DMR-1040229.

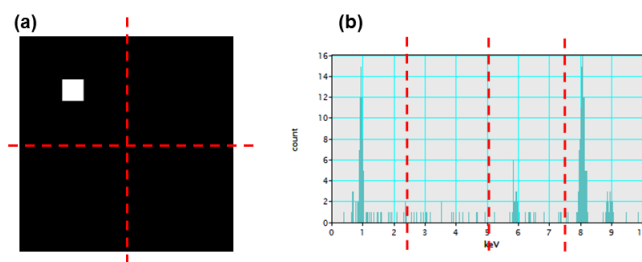


Figure 1. Schematic diagrams to show segmentation schemes for (a) spatial local PCA and (b) spectral local PCA methods.

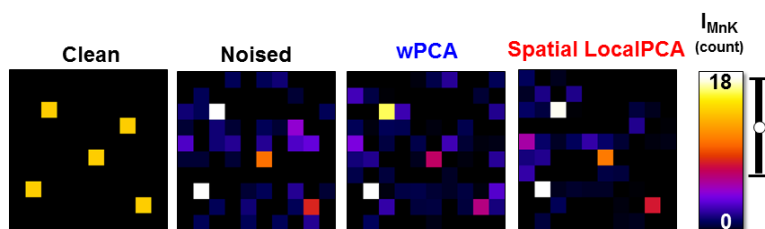


Figure 2. A Comparison of Mn maps from simulated, Poisson-noised, PCA processed and spatial local PCA-processed XEDS SI datasets.

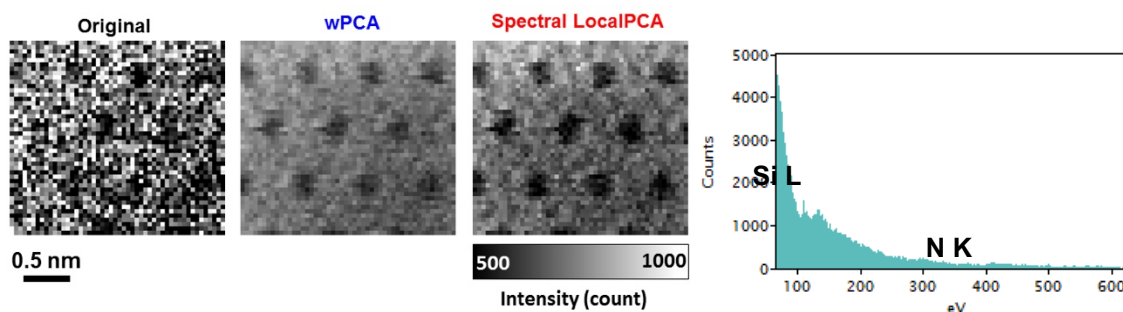


Figure 3. A Comparison of nitrogen maps from atomic-resolution EELS SI datasets of (0001)-projected Si_3N_4 (original, conventional weighted PCA and spectral local PCA processed).