1423 Microsc. Microanal. 21 (Suppl 3), 2015
doi:10.1017/S1431927615007898 Paper No. 0710

Revisiting Noise Scaling for Multivariate Statistical Analysis

P.G. Kotula and M.H. Van Benthem

Sandia National Laboratories, PO Box 5800, MS 0886, Albuquerque, NM 87185-0886, USA

It has been demonstrated that scaling of spectral image data, such as those containing X-ray spectra or other count data, for non-uniform noise is an essential part of multivariate statistical analysis. [1-3] Multivariate statistical analysis algorithms, such principal components analysis (PCA), rank the respective components by variance. Variance can arise due to chemical signals (e.g., peaks) and also due to noise. The goal of the scaling is to weight the data such that variance due to noise in high signal regions of the spectral image is down weighted. The method described by Keenan and Kotula for optimal scaling [2,3] scales the spectral image data in **D** using the column and row space means

$$\widetilde{\mathbf{D}} = \mathbf{GDH} \tag{1}$$

where **G** is a diagonal matrix whose diagonal elements are the inverse of the square roots of the row means (mean image) and **H** is a diagonal matrix whose diagonal elements are the inverse of the square roots of the column means (mean spectrum). While this has been an extremely powerful normalization approach, for very sparse data sets this might not be optimal.

Today with the advent of large-area silicon-drift detectors (SDD) in analytical electron microscopes (AEM), there is a temptation to increase the number of pixels in a spectral image beyond what the nominal signal levels can support. Alternatively specimen damage can limit acquisition time. This, in principle, poses no significant problems as we can post process the data via geometrical binning for example, to improve the signal levels. Nanoparticles on a thin support are likely to generate few X-ray counts under imaging conditions which do not damage the specimen quickly. For the subsequent example, an FEI Company Titan G2 80-200, operated at 200kV and equipped with a high-brightness Schottky emitter, spherical aberration corrector on the probe forming optics, and 4 SDDs with a combined solid angle of collection of 0.7sr. The probe had a measured diameter of ~0.15nm with 200pA and a convergence angle of 18 mrad. Higher currents could have been used but the specimen damaged too quickly. A spectral image was acquired with 0.5nm pixels, 200 pixels by 200 pixels (100 nm by 100 nm field of view) by 2048 channels (10 eV/channel) in 1900 seconds. The specimen consisted of yttria particles decorated with smaller Pd particles on a thin carbon support. Figure 1 shows the results of a multivariate statistical analysis (multivariate curve resolution (MCR) [1]) of the data as acquired. This results in one reasonable looking factor with a number of non-physical factors seemingly comprised of random noise. The spectral image data were 99.8% sparse containing a total of 150k counts in almost 80 million data elements. Given this overall sparsity and the fact that the sample is one central yttria particle decorated with Pd particles, it is also very spatially sparse. Figure 2 shows the results of an analysis of the same spectral image where we have binned spatially 4x4, spectrally by a factor of two and modified the normalization method in Eq. 1 so that we are not performing spatial weighting. The equation thus becomes

$$\widetilde{\mathbf{D}} = \mathbf{DH} \tag{2}$$

indicating that we have weighted only in the spectral domain. The same MCR calculation as before now results in only three factors all with physical significance. As sparse data may be more the rule than exception in the AEM, this combined approach of binning of the data and a modified noise normalization approach may make sense for many analyses.

[1] Kotula, P.G., Keenan, M.R. & Michael, J.R. (2003). *Microsc Microanal* **9**, 1–17.
[2] M.R. Keenan & P.G. Kotula, (2004) *Appl. Surf. Sci.*, **231/232** 240-244.
[3] Keenan, M.R. & Kotula, P.G. (2004) *Surf Int Anal* **36**, 203–212.
[4] Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The authors would like to thank Prof. Barry Carter at the University of Connecticut for provision of the specimen.

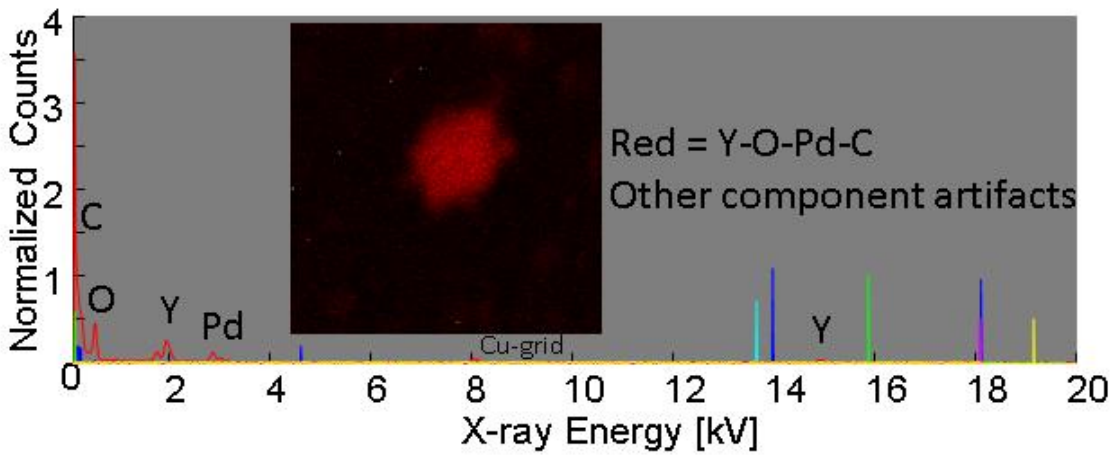

Figure 1. MCR of the raw spectral image showing the first six factors of which only one is physically reasonable.

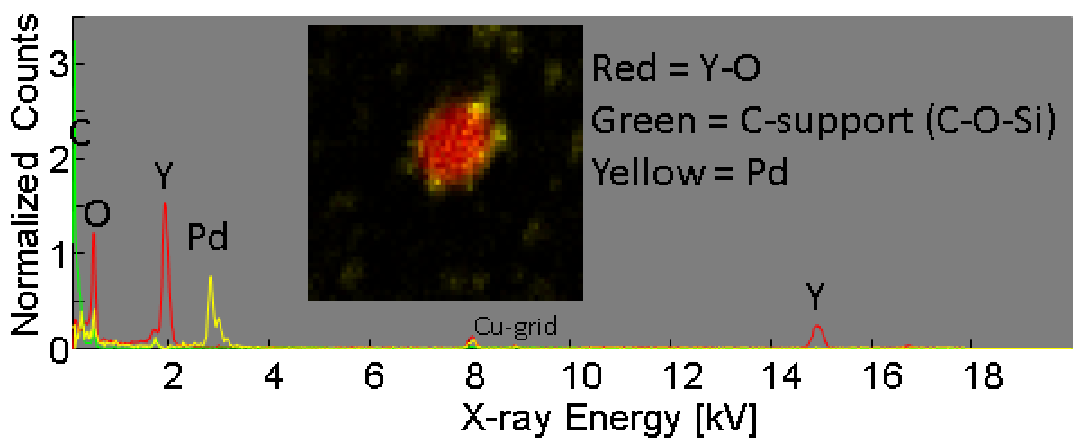

Figure 2. MCR analysis of the binned spectral image normalized for noise in only the spectral domain showing three physically reasonable factors.

https://doi.org/10.1017/S1431927615007898 Published online by Cambridge University Press