Applications of Multivariate Statistical Methods to Analysis of Electron Backscatter Diffraction and Transmission Kikuchi Diffraction Datasets

Angus J Wilkinson[1], Yevhen Zayachuk[1], David M Collins[1], and Rajesh Korla[1,2]

[1] Department of Materials, University of Oxford, Oxford, UK.
[2] Department of Materials Science, Indian Institute of Technology Hyderabad, India

Multivariate statistical analysis (MSA) is widely used across science disciplines including microscopy, however, to date its application to electron backscattered diffraction (EBSD) data has been surprisingly limited. Brewer, Kotula, and Michael [1] demonstrated that MSA could be applied to EBSD data but this appears to be the only previous publication in the area. MSA are typically used for data reduction, identifying inter-related variables, and clusters in datasets. In applying MSA to EBSD and microstructural analysis the key benefits to pursue are its use in (i) segmenting the spatial domain into distinct similar regions (i.e. grains or sub-grains), and (ii) producing representative lower noise patterns associated with these domains to aid pattern indexing. MSA is unlikely to ever replace Hough-based analysis of EBSD patterns but may augment it by improving analysis of low quality data sets, or differentiating finer details not routinely detected though the Hough/Radon transform. We have used factor analysis within Matlab to implement MSA and present some illustrative examples below.

The first step in the analysis is to re-write each EBSD pattern as a vector containing intensities at each effective pixel after binning on the camera. The set of patterns are then arranged into a data array with columns containing intensities from an individual pattern, and rows containing the intensity variation across the spatial map for a particular detector pixel. The data array can be large and so significant binning of the detector is advised. We then seek to decompose this data array into a set of basis vectors describing underlying intensity distributions on the detector (i.e. basis patterns) and the variation of their strength from one observation point to the next (i.e. a spatial map). Principal component analysis is perhaps the most widely used MSA approach, and seeks a first basis pattern that extracts the largest possible variance in the spatial data, followed by succeeding components that must be orthogonal to the preceding components but are chosen to again have the highest variance possible. Limiting the data to the first few of these component patterns is an effective means of data reduction. However, an issue with applying PCA to EBSD data is that it does not guarantee a single spatial point in the map is dominated by a single basis pattern. Instead the VARIMAX solution is more suitable as this generates basis patterns such that each observation (ie EBSD map point) tends to be dominated by a single basis pattern. Formally the VARIMAX solution maximizes the sum of squared correlations between and basis patterns and their spatial variation.

Figure 1 gives a simple example in which MSA was used to distinguish grains in a small map from a ferritic steel sample with good quality EBSD pattern intensities recorded at 100x72 pixels. Four selected basis patterns are shown to each side along with grayscale maps showing where the signal for this pattern was strong. For the two examples on the right uniform strong (white) signal is shown within the grain with high contrast from the very low (black) signal elsewhere in the map. For the cases on the left the two basis patterns come from the same grain and are visually quite similar (though the corresponding vectors are still orthogonal). This grain contains some lattice curvature which results in a gradual variation in signal strength across the grain between these two basis patterns which represent a small change in lattice orientation.

A transmission Kikuchi diffraction (TKD) dataset from a nanoscale SiC sample presents a more challenging example. The top row of figure 2 shows an orientation map based on Hough-based analysis along with two examples of the weak patterns obtained which lead to the poor indexing. The lower row shows the successful segmentation of the microstructure into grains achieved using the VARIMAX MSA approach. Two of the basis patterns are also given to illustrate the improved pattern quality achieved by combining information from within grains.

To generate orientation information it only necessary to index the basis patterns rather than each individual pattern in the map which provides a major reduction in the number of analyses to be completed. This is of huge operational benefit in reducing computational resource issues associated with library or dictionary based approaches to pattern indexing [2]. Figure 3 shows measurements from a Zr sample generated by first applying the VARIMAX MSA analysis and then finding the best match between the 250 basis patterns and a library of 300,000 simulated patterns. The library was populated by interpolation of a dynamical diffraction simulation on a stereographic projection produced using Bruker Dynamics software.

References:
[1] LN Brewer, PG Kotula and J Michael, Ultramicroscopy **108** (2008), p. 567-578.
[2] YH Chen, SU Park, D Wei et al, Microsc & Microanal. **21** (2015), p. 739-752
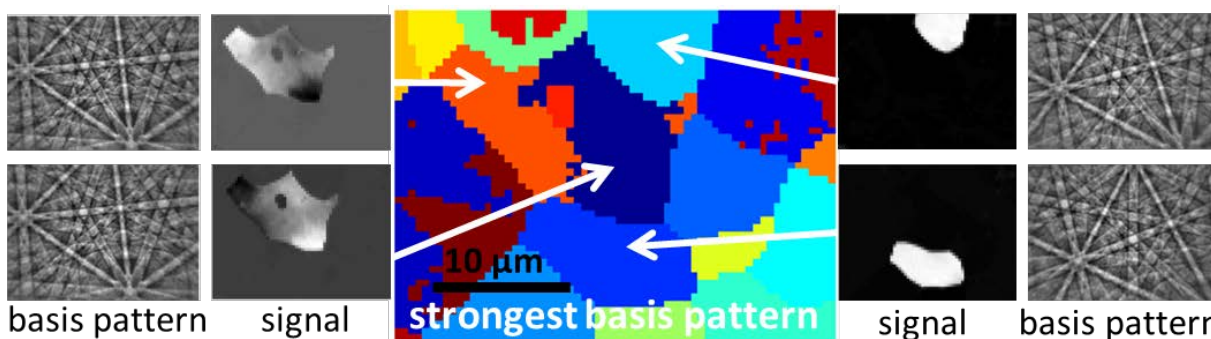[3] funding from EPSRC under grants EP/K034332/1, EP/K032518/1 and EP/N017110/1 is gratefully acknowledged

Figure 1. Central panel shows grains in a ferritic steel segmented by strongest basis pattern signal, with 4 examples of individual basis patterns and their signal variations across the map.
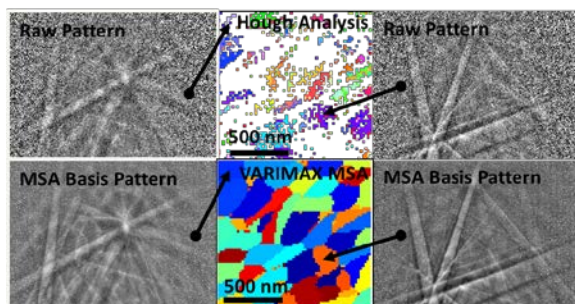

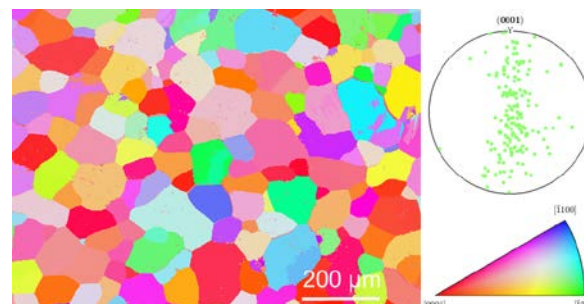**Figure 2.** TKD map from a nanoscale SiC sample comparing Hough and MSA – based analyses.

Figure 3. IPF map for pure Zr bar and (0001) pole figure, generated by a library based analysis of 250 basis patterns from MSA analyses.