doi:10.1017/S143192761700126X
Microsc. Microanal. 23 (Suppl 1), 2017

Data Clustering and Scanning Precession Electron Diffraction for Microanalysis

Benjamin H. Martineau[1], Duncan N. Johnstone[1], Joshua F. Einsle[1], Paul A. Midgley[1] and Alexander S. Eggeman[1]

[1.] Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, UK

Modern transmission electron microscopes are capable of recording and storing unprecedented amounts of data, which conventional analysis methods are ill-equipped to deal with. However, recent advances in computing power and the birth of 'big data' have given rise to a wealth of new statistical techniques to fill the gap. One family of these techniques is data clustering, which we apply to scanning precession electron diffraction (SPED) data.

SPED involves scanning a double-rocking electron beam across the sample and recording a PED pattern at each position. Previous work on the resultant four-dimensional datasets has focused on the use of virtual dark-field imaging [1] or pattern-matching approaches [2], enabling microscale analysis of local crystallography, such as strain, misorientation, or atomic ordering. However, such methods underutilise the available data and suffer in cases of where distinct microstructural components overlap. More recent work has extended such efforts, exploiting redundancy in the data by employing non-negative matrix factorisation (NMF) [3], successfully examining the three-dimensional characteristics of a volume of material [4]. However, NMF and similar statistical decompositions prove to be limited when there is overlap in the diffraction signal from two structural elements, such as at a low-angle tilt boundaries. It is also often unclear how many decomposition components to use.

Data points from parts of the sample which produce similar diffraction patterns lie close together, when represented as points in high-dimensional space, and lie further from dissimilar parts of the sample. Algorithmically identifying these groups of points via data clustering produces a segmentation of structurally similar parts of the sample with little prior knowledge, and the geometrical centre of the groups can be understood as a prototypical signal, or basis diffraction pattern. A signal may be composed combinations of basis patterns, so the unambiguous assignment of a given pattern to a cluster centre, known as hard clustering, is inappropriate here – instead fuzzy clustering, in which each data point has a 'membership' to each centre, is used. This approach must typically be performed in a reduced-dimensionality space to avoid the so-called 'curse of dimensionality'.

Figure 1 presents a SPED simulation of a gallium arsenide Σ3 twin boundary produced using a multislice method. The boundary was rotated 30° about a <100> zone axis so that the twins overlapped in projection. As the NMF algorithm attempts to model features in terms of reducing significance it produces 'pseudo-subtractive' artefacts in the representative features to efficiently describe the data, and thus the localisations are somewhat misleading. The method presented instead involves reducing the dimensionality of the data using singular value decomposition (SVD) which preserves data structure, and then clustering using a probabilistic approach. The data clustering focuses on the end members in the data, rather than the variation, and so the derived patterns and localisations match the originals more closely.

Figure 2 shows a cluster analysis of an experimental SPED dataset collected from the 'cloudy zone' of an iron-nickel meteorite [5]. NMF identifies a unique matrix phase, but the learnt component pattern is not physical, as there is too much signal overlap with the tetrataenite precipitates. The cluster analysis

https://doi.org/10.1017/S143192761700126X Published online by Cambridge University Press

identifies the same matrix component, but the superlattice peaks in the matrix phase are now much clearer. This result corresponds well with predictions for the hitherto-unobserved Fe_7Ni structure.

Cluster analysis is well-suited to SPED data and it has proved to be particularly useful in cases of where structural similarity leads to common diffracting vectors, such as for coherent phases. Used in conjunction with other methods, it provides a powerful, quick, and reliable extension to a researcher's toolbox [6].

References:

[1] EF Rauch and M Véron, Eur. Phys. J. Appl. Phys. **66** (2014), p. 10701.
[2] J Portillo *et al.*, Mater. Sci. Forum **644** (2010), p. 1.
[3] DD Lee and HS Seung, Nature **401** (1999), p. 788.
[4] AS Eggeman, R Krakow and PA Midgley, Nat. Commun. **6** (2015), p. 7267.
[5] JFJ Bryson *et al.*, Earth Planet. Sci. Lett. **388** (2014), p. 237.
[6] The authors acknowledge funding from the Royal Society, the EU's Seventh Framework Programme, grant numbers 320750 and 291522-3DIMAGE, and the EC Seventh Framework Programme, grant number 312483 (ESTEEM2)

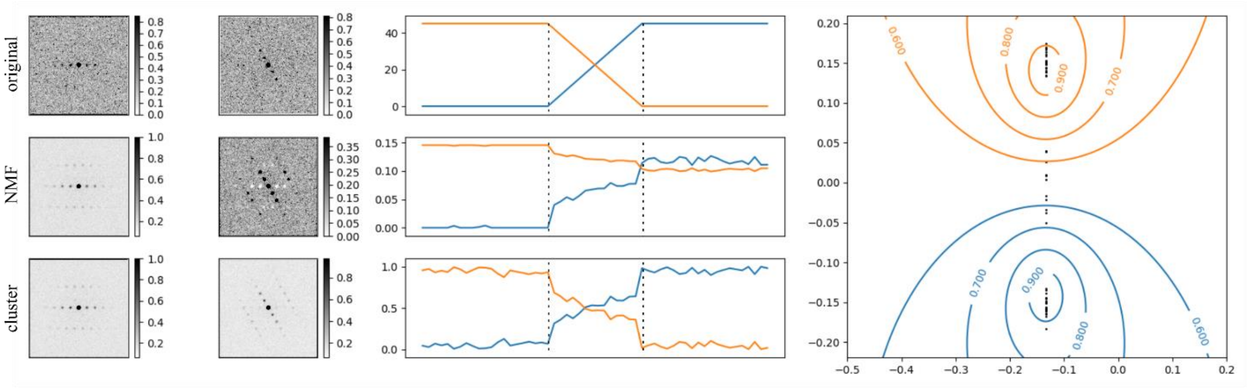

Figure 1. NMF and cluster analysis of a simulated dataset. Left: representative features. Centre: localisations. Right: the clustered patterns in a low-dimensional projection.

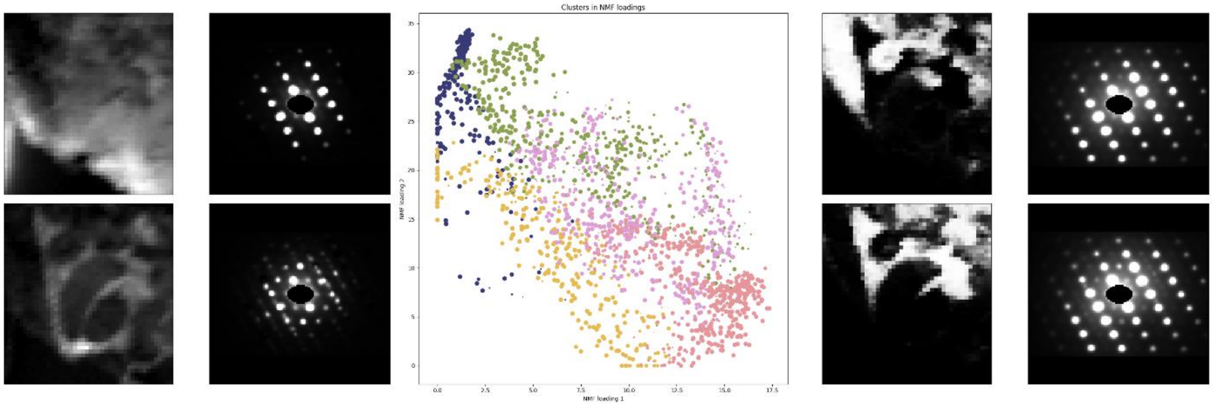

Figure 2. NMF and cluster analysis of an ordered matrix phase in tetrataenite. Left: NMF localisations and representative features. The bottom pair of images represent an ordered matrix phase. Centre: clusters in a low-dimensional projection. Right: cluster localisations and representative features. The bottom pair are the same ordered phase.

https://doi.org/10.1017/S143192761700126X Published online by Cambridge University Press