Ultramicroscopy 196 (2019) 88–98

Contents lists available at ScienceDirect

Ultramicroscopy

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

Applications of multivariate statistical methods and simulation libraries to analysis of electron backscatter diffraction and transmission Kikuchi diffraction datasets

Angus J Wilkinson[a,*], David M Collins[a,b], Yevhen Zayachuk[a], Rajesh Korla[a,c], Arantxa Vilalta-Clemente[a,d]

[a] Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK
[b] School of Metallurgy and Materials, University of Birmingham, Edgbaston B15 2TT, UK
[c] Department of Materials Science & Metallurgical Engineering, Indian Institute of Technology, Hyderabad, India
[d] Groupe de Physique des Matériaux, Université de Normandie Rouen, INSARouen, CNRS 6634, 76000 Rouen, France

ABSTRACT

Multivariate statistical methods are widely used throughout the sciences, including microscopy, however, their utilisation for analysis of electron backscatter diffraction (EBSD) data has not been adequately explored. The basic aim of most EBSD analysis is to segment the spatial domain to reveal and quantify the microstructure, and links this to knowledge of the crystallography (e.g. crystal phase, orientation) within each segmented region. Two analysis strategies have been explored; principal component analysis (PCA) and k-means clustering. The intensity at individual (binned) pixels on the detector were used as the variables defining the multidimensional space in which each pattern in the map generates a single discrete point. PCA analysis alone did not work well but rotating factors to the VARIMAX solution did. K-means clustering also successfully segmented the data but was computational more expensive. The characteristic patterns produced by either VARIMAX or k-means clustering enhance weak patterns, remove pattern overlap, and allow subtle effects from polarity to be distinguished. Combining multivariate statistical analysis (MSA) approaches with template matching to simulation libraries can significantly reduce computational demand as the number of patterns to be matched is drastically reduced. Both template matching and MSA approaches may augment existing analysis methods but will not replace them in the majority of applications.

1. Introduction

Multivariate statistical analysis (MSA) tools are widely used across science disciplines including microscopy and diffraction methods. MSA methods have been developed to aide analysis of datasets that have grown increasingly large and complex. MSA aims to help extract the most useful information within the data. This is achieved by revealing hidden ('latent') variables that better and more simply describe the data, or by identifying clusters within the data that act in similar ways. This often leads to a reduction in the overall quantity of data while improving the information content or quality of the data retained. MSA has been applied to data from a range of microscopical techniques initially within the biological community, but increasingly within the materials sciences. The area where most application has been found in electron microscopy of materials is analysis of spectroscopy data obtained from the transmission electron microscopy (TEM) [1–8], though there have been notable applications in the scanning electron microscope to both energy dispersive X-ray spectroscopy (EDX) [9–11] and cathodoluminescence (CL) [12].

Electron backscatter diffraction (EBSD) is a scanning electron microscope (SEM) based technique that allows crystallographic information to be obtained from small volumes of material. The technique has been reviewed several times previously [13–18]. The technique finds widespread application in materials and earth/planetary sciences. Its huge versatility in mapping orientation [13,19–23], crystal type [24–28], strain [29–33], dislocation content [34–41] and crystal perfection over a wide range of length scales makes it a powerful microstructural characterization tool.

To date, the application of MSA methods to EBSD data has been surprisingly limited. Indeed the work of Brewer, Kotula, and Michael [42] appears to be the only previous publication in the area and is now ~10 years old, though a patent application by Stork and Brewer [43] describing a general hierarchical clustering method mentions potential application to EBSD data. Brewer et al. demonstrated the potential for MSA in reducing the number of EBSD patterns to be indexed, and improving their quality (i.e. signal to noise ratio, and thus indexibility). They foresaw the utility of the approach in deconvolving pattern overlap (a pattern comprising information from more than one crystal)

* Corresponding author.
E-mail address: angus.wilkinson@materials.ox.ac.uk (A.J. Wilkinson).

https://doi.org/10.1016/j.ultramic.2018.09.011
Received 6 June 2018; Received in revised form 16 September 2018; Accepted 20 September 2018
Available online 21 September 2018
0304-3991/

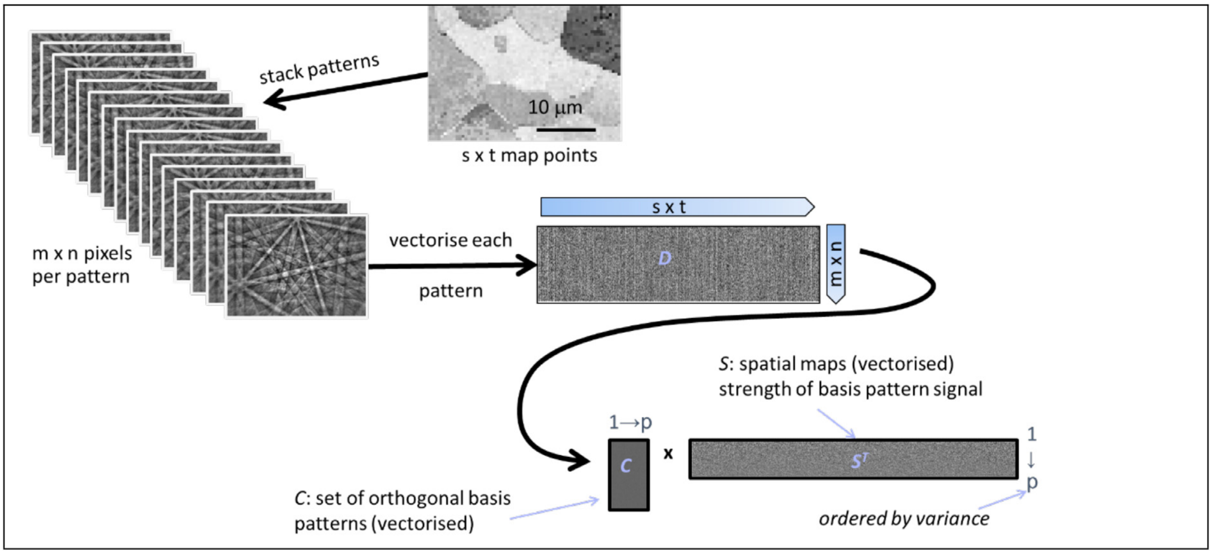

Fig. 1. Schematic diagram showing construction of a single data array ***D*** containing vectorised forms of each pattern in columns.

issues that occur in fine grained materials where MSA might extend the capabilities of EBSD.

In applying MSA to EBSD for microstructural analysis the potential benefits are in (i) segmenting the spatial domain into distinct similar regions (i.e. grains, sub-grains, or domains), and (ii) producing a smaller number of representative lower noise patterns associated with these regions. Implementing the first benefit should allow the morphology of the microstructure to be revealed and quantified without the requirement to index the patterns themselves. This would allow, for example, quantification of grain size and shape distributions to be undertaken. This may be advantageous in situations when the individual EBSD patterns themselves are too weak or blurred to be indexed with confidence. The possibility to assign each pixel to a single cluster with other pixels generating similar EBSD patterns directly addresses the pattern overlap issue and may provide sharper definition of grain boundaries. Most MSA schemes also provide measures of signal strength contributed from each of a set of neighbouring grains which offers a route to interpolate and better refine the boundary position independent of the EBSD grid points.

The second potential benefit of MSA analysis is the generation of a smaller number of better quality EBSD patterns. Brewer et al. [42] demonstrated that these representative patterns could then be indexed using established Hough-transform based methods. This has the added benefits that signal to noise is improved and time is saved from the reduced number of patterns to be indexed. The improved pattern quality may have real advantages when pattern quality is low due to sample preparation issues, high deformation states or beam sensitive samples, or when fine details in the patterns need to be used to resolve pseudo-symmetry issues. An important new possibility that has emerged since the work of Brewer et al. is indexing via template matching to simulated pattern libraries. Here, the number of test patterns to compare against is reduced, providing a major reduction in the computational time required. It seems unlikely that MSA will replace Hough-based analysis of EBSD patterns but it may augment it by improving analysis of low quality data sets, or differentiating finer details not routinely detected through the Hough/Radon transform.

The MSA methods will be described in Section 2, and the template matching analysis for pattern indexing in Section 3, with the main steps illustrated using a small data set from a ferritic steel sample. Further examples will then be given in Section 4 to demonstrate some benefits obtained by employing MSA in the analysis of EBSD and TKD (transmission Kikuchi diffraction) data.

2. Multivariate statistical analysis

2.1. Principal component analysis

EBSD patterns recorded in a map can be considered as a set of independent intensity measurements at each pixel (or group of binned pixels) on the detector. Perhaps the most commonly used of the MSA methods is principal component analysis (PCA) [44]. PCA aims to convert a set of observations (e.g. patterns in an EBSD map) of a number of variables (e.g. intensity at each detector pixel) into a set of values of linearly uncorrelated (latent) variables called principal components. An orthogonal transformation is used such that the first principal component has the largest possible variance from the mean i.e. it describes the greatest variability in the data. Succeeding components are then sequentially established so that they are orthogonal to the preceding components and have the highest variance possible. The resulting principal component vectors form an uncorrelated orthogonal basis set. PCA can be used to reduce the dimensionality of a problem by retaining only a subset of the most significant principal components.

To apply PCA to an EBSD dataset the first step is to re-write each EBSD pattern (m × n pixels) 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 ***D*** with columns containing intensities from an individual pattern (i.e. a single observation of the many m × n variables), and rows containing the intensity variation across the s × t points of the spatial map for a particular detector pixel (i.e. all s × t observations for a single variable). Construction of the data array is illustrated in Fig. 1. The data array can be large and so significant binning on the detector is advised. We then seek to decompose this data array into a set ***C*** of characteristic vectors describing the underlying intensity distributions on the detector (i.e. basis patterns) and ***S*** the variation of their strength from one observation point to the next (i.e. spatial maps, where the superscript T denotes a matrix transpose):

$$\boldsymbol{D}_{[m\times n, s\times t]} = \boldsymbol{C}_{[m\times n, p\leq m\times n]}\boldsymbol{S}^{T}_{[p\leq m\times n, s\times t]} \quad (1)$$

There are m × n characteristic vectors, which were determined using singular value decomposition within MatLab. Data reduction can be implemented by only retaining the first p ($p \leq m \times n$) of these eigen-vectors each of which can be reformed into a characteristic EBSD pattern.

To demonstrate the result of this decomposition we will use a small data set obtained from a ferritic steel sample using a Bruker e-Flash EBSD detector in Zeiss Merlin FEG-SEM system. The patterns were