An Examination of Defects in Diamond Using Data Clustering

N. C. Wilson[*], C. M. MacRae[*], A. Torpy[*], J. Boland[**], and E. P. Vicenzi[***]

[*]CSIRO Process Science and Engineering, Clayton South, VIC 3169, Australia
[**]CSIRO Earth Science and Resource Engineering, Pullenvale, QLD 4069, Australia
[***]Department of Mineral Sciences, Smithsonian Institution, Washington, DC 20560

Collection of hyperspectral data sets is now a routine task in mapping microanalysis, and for large maps this can lead to data sets that are many gigabytes in size, requiring new tools for data processing. This work uses data clustering algorithms to examine such large data sets and classify defect types in maps collected on two different diamond samples. The first example is of a carbonado diamond, where spectral x-ray and cathodoluminescence (CL) data was collected on a map 4000 x 3000 pixels in size, resulting in a data set size of about 120 GB. The second example is of diamond composite material used in a superhard tool for cutting, with a 3000 x 2000 pixels map resulting in a 60 GB file.

Hyperspectral data sets offer the advantage of allowing post-hoc examination of the data, giving the ability to find unexpected phases (x-ray data) and defects or centres (CL data) within a sample [1]. We have investigated the use of clustering algorithms as a means to explore these large datasets, and to allow the user to sort and present the relevant information from a sample. This is achieved by an initial partitioning of the data in a k-means like approach. However, unlike the k-means approach, we do not stipulate the number of clusters into which the data is partitioned; rather we set a statistical size for the addition of a pixel to a cluster, and add more clusters as needed. This typically leads to a data set being split into several thousand clusters, which are then grouped together in a hierarchical dendogram. The dataset can then be explored using an interactive dendogram interface that controls the phases displayed on a phase patched image.

To relate cathodoluminescence spectra to fundamental physical processes, one approach is to deconvolute the spectrum by fitting Gaussian functions. However, within a single hyperspectral data set a variety of spectral shapes can exist, with different spectral shapes requiring a different set of Gaussians to obtain a good fit. Attempting to create a 'super-set' of Gaussians to fit the whole map inevitably leads to poor convergence in fitting. Using data clustering we can split the data into different spectral shapes, which can then be separately examined and fitted.

Carbonado diamonds have been found in deposits in Brazil and Central Africa, and are interesting as the origin of the diamonds is controversial [2]. Cathodoluminescence has been used to study carbonado diamonds previously [3], by using flood gun images and a small number of spectral point analyses. In this study we have performed combined hyperspectral CL and x-ray mapping on a carbonado diamond to determine defect structures and elemental inclusions. Clustering allows the spectral shapes to be visualised across the map in both intensity and shape projection. Figure 1 shows a clustered map from a carbonado diamond, along with some of the different peak shapes, relating to different defect types.

Diamond composite materials can be used to form coatings for super-hard tool bits that are used in industry, offering greater wear resistance [4]. However the wear resistance can be variable and hyperspectral CL has been used to examine the microstructure of the embedded diamonds within the coating to understand this problem. Clustering the hyperspectral CL, in this case, extracts the underlying spectral shapes which are related to the diamond defect types. These can be used to correlate with wear rates providing an insight into why certain tool pieces fail.
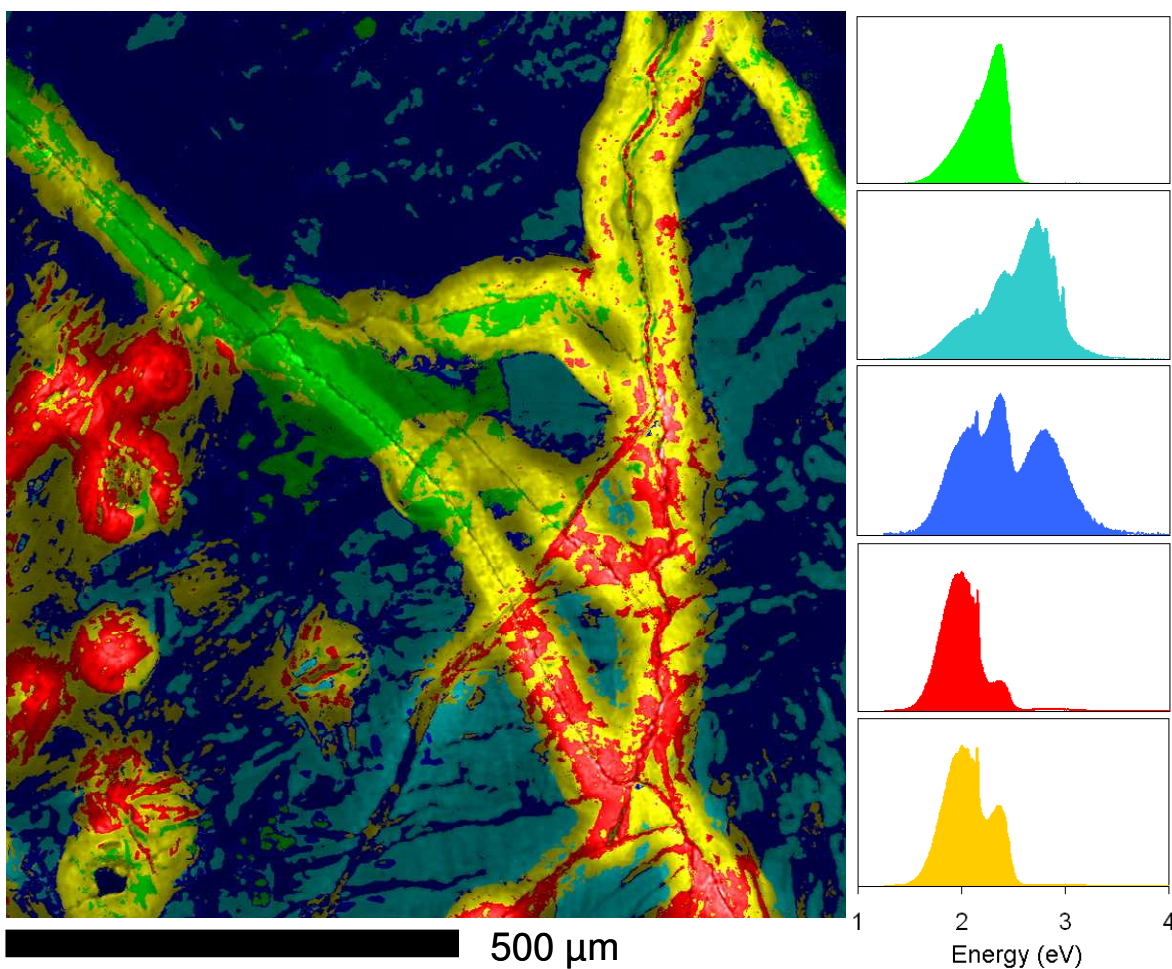

Figure 1. Clustered map from a section of a carbonado diamond, with color representing the different spectral type, and brightness representing the intensity. Collected at 15 kV and 50 nA with a step size of 500 nm with average spectra of clusters found within the sample shown on the right.

References

[1] N. C. Wilson et al, *Microscopy and Microanalysis* 14 (Suppl. 2) (2008) 764
[2] P. J. Heaney, et al, *Elements*, **1** (2005) 85
[3] B. Rondeau, et al., *Diamond & Related Materials*, **17** (2008) 1897
[4] J N Boland, et al., Euro PM2006, Belgium 2006