Hyperspectral Cathodoluminescence Examination of Defects in a Carbonado Diamond

Nicholas C. Wilson,[1,*] Colin M. MacRae,[1] Aaron Torpy,[1] Cameron J. Davidson,[1] and Edward P. Vicenzi[2]

[1]*CSIRO Process Science and Engineering, Clayton, VIC 3168, Australia*
[2]*Museum Conservation institute, Smithsonian Institution, Suitland, MD 20746, USA*

Abstract: Hyperspectral cathodoluminescence mapping is used to examine a carbonado diamond. The hyperspectral dataset is examined using a data clustering algorithm to interpret the range of spectral shapes present within the dataset, which are related to defects within the structure of the diamond. The cathodoluminescence response from this particular carbonado diamond can be attributed to a small number of defect types: N-V^0, N$_2$V, N$_3$V, a 3.188 eV line, which is attributed to radiation damage, and two broad luminescence bands. Both the N$_2$V and 3.188 eV defects require high-temperature annealing, which has implications for interpreting the thermal history of the diamond. In addition, bright halos observed within the diamond cathodoluminescence, from alpha decay radiation damage, can be attributed to the decay of ^{238}U.

Key words: cathodoluminescence, carbonado, data clustering, hyperspectral

INTRODUCTION

Carbonado diamonds are a variety of polycrystalline diamond found in alluvial deposits in the Central African Republic and Brazil (Trueb & Christiaan De Wys, 1969). They are typically black and have high porosity, finding use in industrial applications that require super-hard materials. The polycrystalline nature of carbonado makes the propagation of cracks difficult giving them superior properties over single crystal diamonds in applications such as cutting and polishing hard materials. From a geological point of view, carbonado diamonds are interesting because their genesis is not fully understood (Heaney et al., 2005; McCall, 2009). They differ from kimberlite diamonds in that they have higher porosities than microcrystalline diamond aggregates that are found in kimberlites, and mineral inclusions found in the pores of carbonado diamonds are typical of the Earth's crust, not of the upper mantle. This has led to a number of theories (Smith & Dawson, 1985; Garai et al., 2006; McCall, 2009; Cartigny, 2010; Demeny et al., 2011; Sautter et al., 2011) on the formation of carbonado diamonds including formation in the upper mantle from subducted crustal organic matter, radiogenic fission of U and Th in a carbon-rich matrix, and conversion of organic matter through meteorite impact.

Within the diamond structure the most common impurity element is nitrogen, as it is easily accommodated owing to its atomic radius being almost equal to that of carbon. Nitrogen can be present in concentrations up to 0.25 wt% (Collins, 1992). It is difficult to quantify nitrogen in diamonds by using electron-beam microanalysis to measure the nitrogen Kα X-ray emission, as very few N X-rays escape the carbon matrix due to the very high (25,500) mass absorption coefficient of nitrogen in carbon. The characterization of nitrogen is important, as the aggregation of nitrogen is mainly controlled by the residence time of the diamond in the mantle and the temperature of the mantle (Taylor et al., 1996). To identify the defects in diamonds, cathodoluminescence (CL) is commonly used because there is a strong signal from nitrogen incorporation and because the cathodoluminescence is sensitive to a number of different nitrogen defect types. For this reason, luminescence has been used as a means to distinguish between natural and synthetic diamonds. Cathodoluminescence is also advantageous as it offers a higher spatial resolution than Fourier transform infrared spectroscopy, a common technique for studying nitrogen aggregation (Garai et al., 2006). It should be noted that cathodoluminescence cannot directly determine the defects that give rise to particular CL emissions. Instead, the identity of defects may be inferred from observed CL peaks by comparison with known electronic structures of defects that have been characterized by other techniques, such as micro-Raman and electron spin resonance (ESR) (Baranov et al., 2009; Shames et al., 2012).

Cathodoluminescence techniques have previously been used to study carbonado diamonds by using multispectral (i.e., three color) flood gun images, supplemented with a small number of spectral CL point analyses (Harte et al., 1999; Vicenzi et al., 2006; Kagi et al., 2007; Rondeau et al., 2008). These previous spectral analyses show that within the analyzed volume a mixture of defects can occur. While these spectra can be deconvoluted by fitting Gaussian or other functions into lines from the various defects, the more commonly used multispectral flood-gun images do not contain sufficient spectral information to extract the spatial distribution of all defect types. To do this, a "hyperspectral" dataset is needed; this is obtained by mapping an area and for each pixel of the map an entire CL spectrum is recorded.

With such a dataset, a set of peaks corresponding to a particular defect type can be deconvoluted from the spectrum at each pixel and projected across the mapped area.

Hyperspectral datasets are now a routinely collected in mapping microanalysis, with the great advantage of allowing *post hoc* data examination, which can lead to the discovery of unexpected phases (in X-ray data) and defects or centers (CL data) within a sample (MacRae et al., 2005). An important advantage for CL analysis, in which the signal can be sensitive to beam damage, is that the spectra are collected in a single mapping pass avoiding artifacts such as spectra shape changes that may be present in multipass CL band mapping.

Within a single hyperspectral dataset, a variety of spectral shapes can exist across the mapped area, and each may require quite different sets of Gaussians to obtain a good fit. Attempting to create a "superset" of Gaussians to fit the whole map inevitably leads to poor convergence or overfitting of the data. However, by using a data clustering technique, we can partition the data into different groups with similar spectral shapes, which can then be separately examined and fitted using appropriate Gaussian functions. The aim of this study is to investigate what additional information hyperspectral cathodoluminescence can yield regarding the defect distribution within a carbonado diamond.

A common problem encountered when trying to manage and process large hyperspectral datasets is how to process the data so as to maximize the information. One common approach for analyzing hyperspectral data is the use of principal component analysis, which reduces the dimensionality of the dataset (Kotula et al., 2003). Another approach is the use of data clustering, which is finding application with hyperspectral microanalysis datasets (Wilson & MacRae, 2005; Wilson et al., 2008; Stork & Keenan, 2010). There are two main classes of clustering algorithms: hierarchical and nonhierarchical (Jain et al., 1999). In hierarchical clustering algorithms, a linkage of clusters is made that can be pictorially represented in a dendrogram; a tree diagram showing the connection of the clusters. This linkage of clusters can be achieved by agglomeration, where each measurement starts in a single cluster, and then pairs of clusters are successively merged, with each merger recorded, or by division, where all measurements begin in one cluster and then are successively split into smaller clusters. The record of these mergers or divisions gives a hierarchical dendrogram, a powerful tool to explore the relationships between pixels within a map, with the most significant spectral differences in the top branch working down to the least significant chemical differences at the bottom. The drawback of hierarchical methods is that they scale poorly with an increasing number of data points, with the computation cost of agglomerative clustering rising with the cube of the number of data points. Consequently, it is impractical to apply this approach to cluster every pixel in a typical map of one million pixels or more.

Nonhierarchical algorithms are better suited to large datasets; for example, they are often used in applications such as remote sensing where spectral maps are acquired. These algorithms partition the dataset into a set of clusters but do not produce any linkage or association between the clusters such as a dendrogram.

The most commonly used nonhierarchical clustering method is the k-means algorithm (MacQueen, 1967), which is an iterative method that searches for the best set of k cluster centroids. The k-means algorithm requires an arbitrary choice of k, the number of clusters to be found. The centroids are given an initial position in the n-dimensional dataset, and then each pixel in the map is assigned to the closest centroid. The centroids are then moved in n-dimensional space to the mean of the data points assigned to them and then, with these new centroid positions, the data points are reassigned to their closest centroid. This process is repeated until some convergence criterion is met, such as no further movement of the centroids.

There are various implementations of the k-means algorithm, using different definitions of distances between pixel and centroid and different choices for the initial positioning of clusters. To produce a satisfactory phase map, the most important parameter is k, which sets an upper bound on the number of phases that can be found. The k-means algorithm moves the centroids to phases that contain the largest number of points, thus choosing a small number for k can lead to phases that occupy only a small number of pixels in the map to be missed in complex samples. To address this, the value of k can be increased, but this leads to an overwhelming number of clusters, which makes a spatial projection onto a phase map difficult to interpret.

In this study, we have taken an approach (Ross, 1968) that uses both hierarchical and nonhierarchical clustering techniques (Wilson & MacRae, 2005; Wilson et al., 2008). The first step is to use a k-means-like algorithm to partition the dataset into a large number of clusters, typically many hundreds to a few thousand. Having reduced our original dataset from the order of one million pixels to a few thousand cluster centroids, it is then possible to process the centroids using a hierarchical clustering method. The result of the hierarchical clustering is a dendrogram, with the most similar clusters joined at the outer branches of the tree and the most dissimilar clusters joined at the base of the tree. This tree can then be used to interactively control the displayed phase map. Working down a branch of the tree, the broad top level groupings can be split to reveal subtler classifications. Thus the user can easily visualize the subtly different phases, which are important to them, while ignoring other subtle variations down other tree branches that are unimportant to the problem being investigated.

Method

Experimental

The sample examined was a Brazilian carbonado diamond, sample L-5 (Magee, 2001). Hyperspectral CL data were collected at room temperature on a JEOL (JEOL Ltd.,

Tokyo, Japan) field emission gun electron microprobe analyzer (JXA 8500F) customized with quartz optics, and equipped with an Ocean Optics (Dunedin, FL, USA) QE65000 grating charge-coupled device (CCD) spectrometer. This spectrometer employs a Hamamatsu (Hamamatsu City, Japan) S7031-1006 back-thinned CCD operating at a temperature of $-15°C$ via Peltier cooling. A 100 μm aperture optical fiber was used to connect the spectrometer to the collection optics of the microprobe with the aperture of the fiber forming the entrance slit of the spectrometer. This resulted in a wavelength resolution of 3.4 nm (MacRae et al., 2005). Spectra were collected with 1,024 equal wavelength windows over the full spectrometer range of 199 to 993 nm. A measurement of the CCD dark-noise signal was made prior to the mapping and subtracted from each pixel in the dataset. This dark-noise measurement was made by averaging together a number of measurements taken with the beam blanked, with the measurements being performed at the same sampling time as used during the mapping, as CCD dark noise varies subtly with sampling time.

In preparation for mapping, the sample was sectioned to expose a fresh surface by grinding with a series of diamond-impregnated polishing pads to ensure flatness, then polished with diamond paste down to 1 μm and finishing with 20 nm colloidal silica, and finally coated with 15 nm of amorphous carbon. The map was collected at 15 kV and 50 nA with a step size of 500 nm, in a stage scan mode, with 4,000 steps in X and 3,000 steps in the Y direction, giving a 2×1.5 mm scan. The advantage of the stage scan mode is that large distances may be mapped without affecting the collection efficiency of the wavelength dispersive X-ray spectrometers or CL data. A dwell time of 25 ms per pixel was used, and at each pixel a full 1,024 channel CL spectrum was collected in parallel with the collection of spectral energy dispersive spectrometer (EDS) data, wavelength dispersive X-ray data, and the backscatter electron signal, resulting in 24 GB of CL data and 98 GB of X-ray data.

Clustering

A two-step clustering approach was applied to the collected dataset. In the first pass, a nonhierarchical partitioning of the data was performed to group the pixels into a set of clusters, where clusters were constrained so that pixels were no more than 3σ (where $\sigma = \sqrt{n}$) from the cluster centroid. The 3σ cutoff was derived from our work on EDS spectral analysis where the counting statistics obey Poisson distribution. It should be noted that in the case of cathodoluminescence, we are not measuring the number of photons directly but are measuring the voltage of the CCD readout. While the photons arriving at the CCD will obey a Poisson-type distribution, the measured signal will also contain a small component of random noise from the voltage conversion and the CCD dark noise.

To generate the initial set of seed centroids and to speed convergence, an analysis of the uniformity of the dataset was made. Seed centroid positions were taken from

pixels in uniform areas (i.e., those pixels close in value to their surrounding pixels), thereby increasing the likelihood that initial positions reflect true chemistry rather than statistical outliers. Then, in a similar approach to that of the k-means algorithm, pixels were associated with their closest centroid, measured in terms of the Euclidian distance metric $D(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, where x_i and y_i are the i'th components of two points x and y. Centroid positions were then updated to the average of mean value of all associated pixels. Where the algorithm differs from the k-means approach is that after the centroid positions were updated, extra centroids were then introduced to the pixels that fell outside an n-dimensional 3σ hypersphere (for an n-channel dataset) around the centroid. In this way, the algorithm did not impose an arbitrary limit on the number of clusters but rather limited the size of the clusters. The process was then iterated until a convergence criterion was reached.

The second pass of the clustering procedure was used to perform an agglomerative hierarchical clustering on the cluster centroids generated from the first pass. For this hierarchical clustering, we used Ward's Method (Ward, 1963), in which clusters were successively merged in order of the pair-wise combination that gave the smallest increase in the variance. Initial variances for the clusters were given by

$$c_k = \sum_{j=1}^{j=m_k} \sum_{i=1}^{i=n} (x_{ijk} - \bar{x}_{ik})$$

and then the clusters with the minimum

$$W = \sum_{i=1}^{i=g} \sum_{j=1}^{j=n_i} (x_{ij} - \bar{x}_i)^2$$

were successively merged.

Deconvolution of CL Spectra

Deconvolution of spectra was achieved by fitting, in energy space, a set of Gaussian functions of the form $G(x) = ae((x - b)^2/2c^2)$, where a gives the height, b position, and c the full-width at half-maximum. The fitting was performed by applying a nonlinear least-squares minimization algorithm.

RESULTS AND DISCUSSION

Calculations were performed on a standard 3.0 GHz Intel Core™2 Duo PC with 2GB of RAM. Clustering of the 24 GB of CL data took approximately 20 h, and subsequent peak fitting to extract peak intensity maps took a further 15 h.

The luminescence of diamonds is a well-studied area, with over 100 centers reported, giving rise to over 1,000 spectral lines (Walker, 1979; Zaitsev, 2001). In terms of this study, only a small subset of these were observed. The N-V^0 defect (Fig. 1a) consists of a nearest-neighbor pair of nitrogen substituting for carbon and a neutral vacancy that gives rise to a zero phonon line (ZPL) at 2.15 eV. A vacancy trapped at a double nitrogen site (Fig. 1b), the N$_2$V defect (sometimes referred to as a H3 center) produces a ZPL at
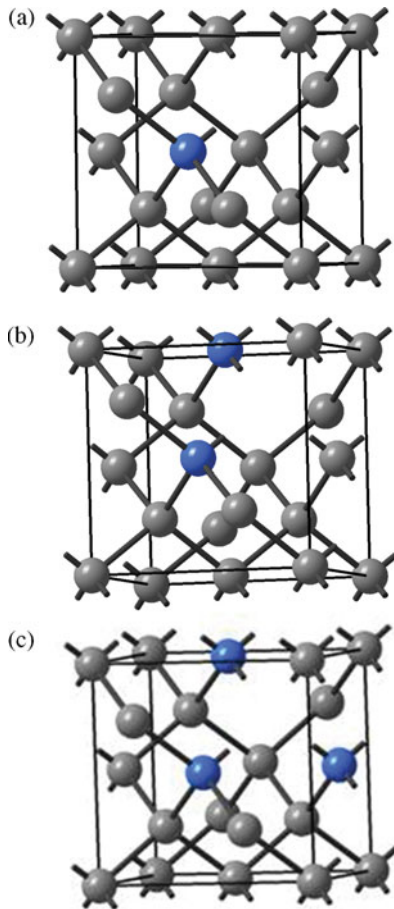
Figure 1. Structural diagrams of diamond defects. (**a**) N-V^0, (**b**) N$_2$V, and (**c**) N$_3$V. Carbon atoms in gray and nitrogen in blue.

2.46 eV, and three nitrogen atoms in a plane with a common vacancy (Fig. 1c), the N$_3$V defect (also referred to as a N3 center) gives rise to a ZPL at 2.99 eV. In addition to these nitrogen defects, we observed a ZPL at 3.188 eV, which is attributed to beta decay radiation damage in natural diamonds and has been experimentally induced in synthetic diamonds post radiation damage by 2 MeV electrons followed by annealing (Collins & Lawson, 1989). Finally, we observe broad bands centered around 2.3 and 2.88 eV.

In addition to the purely electronic transition of the ZPLs of these defects, we also observed the phonon replicas for these lines. Vibronic emission fine structure arises from highly localized defects, generating a set of equally spaced phonon replicas relating to fundamental lattice vibration frequencies ($m\hbar\omega$, $m = 1,2,3\ldots$). In the case of emission, these exist on the low energy side of the ZPL, broadening as they decrease in energy, forming a side band.

Before applying any data cluster analysis, a simple process to apply to the spectral dataset is to integrate all CL channels at each pixel, forming a pan-chromatic image (Fig. 2). One feature observable in this image is a number of alpha particle damage halos from radionuclide decay. These are present as a series of concentric rings at regularly spaced intervals around a central bright source. Twenty-five halos were identified within the mapped area, with the radii then
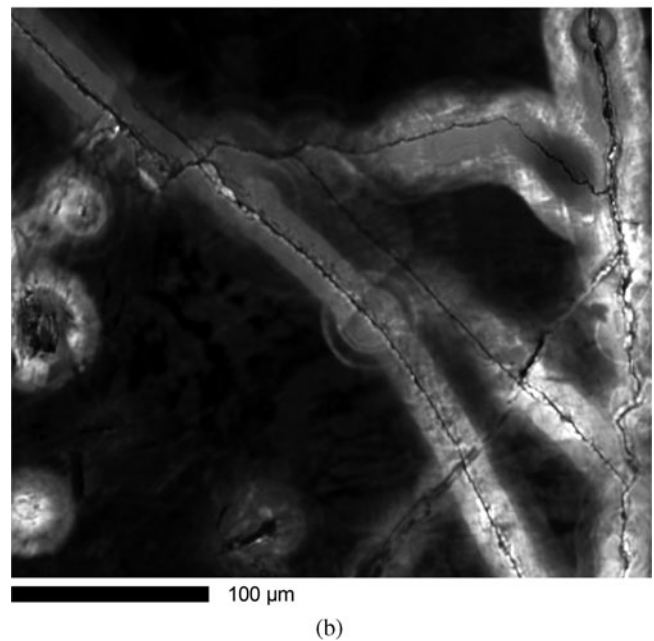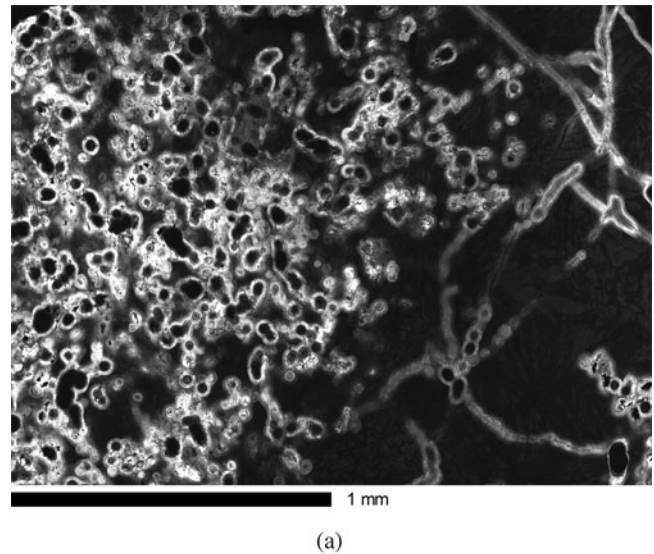

Figure 2. (**a**) Net intensity map of the sample generated from the spectral dataset. (**b**) Close up view of radiation halos.

being measured by manually selecting several control points around each ring (typically 10–20 points/ring), and then using a "gradient descent" algorithm to find the ring centroid by minimizing the square of the differences between the radius to each point. The results of this fitting are given in Table 1, along with stopping distances within diamond calculated using the Bragg-Kleeman rule (Bragg & Kleeman, 1905). The two largest measured rings had mean radii of 23.2 μm and 15.8 μm, respectively, which are in good agreement with the calculated stopping distance of the two most energetic transitions in the ^{238}U decay chain, ^{214}Po and ^{218}Po decays (at 23.1 and 15.6 μm, respectively). The third measured ring at 12.09 μm does not directly match any of the other decays in the ^{238}U chain; however, the third ring is broad and is a good match for the overlap of the ^{222}Rn (13.2 μm), ^{210}Po (12.5 μm), ^{226}Ra (10.7 μm), and

Table 1. Alpha Particle Decay Energies for Various Isotopes, Their Range in Air and Calculated Range in Diamond, and Radii Measurements of Halos from the Net Intensity Map (Fig. 2).*

Decay	E (MeV)	Range Air (cm)	Range Diamond (μm)	Measurement
		^{238}U Decay Chain		
^{214}Po	7.68	6.91	23.1	R1 23.22
^{218}Po	6.00	4.66	15.6	R2 15.79
^{222}Rn	5.48	4.05	13.5	R3
^{210}Po	5.30	3.84	12.8	R3 12.09
^{226}Ra	4.78	3.28	10.9	R3
^{234}U	4.77	3.21	10.7	R3
^{230}Th	4.68	3.11	10.4	
^{238}U	4.20	2.65	8.8	
		^{232}Th Decay Chain		
^{212}Po	8.79	8.57	27.9	
^{216}Po	6.77	5.64	18.3	
^{220}Rn	6.28	5.00	16.3	

*Decay energies taken from Owen (1988).

^{234}U (10.7 μm) decays. The other alpha particles in the ^{238}U decay chain (namely those from ^{230}Th and ^{238}U) have calculated ranges in diamond of 10.4 μm and 8.8 μm but could not be resolved from the CL dataset. It may also be seen from Figure 2 that the area within ~10 μm from the halo center is quite uniform in CL intensity, and no distinct rings can be resolved within this area. The measured rings do not match well with the calculated distances for the ^{232}Th decay chain, indicating that the phase that produced the halo was thorium poor.

An RGB image was extracted from the hyperspectral dataset to approximate the luminescence that would be observed by the human eye or an RGB camera on a flood-gun CL system (Fig. 3). The component color intensities in this image correspond to energy ranges of 2.0–2.3 eV (red), 2.3–2.5 eV (green), and 2.5–3.1 eV (blue). Such three color images may be useful for qualitative investigation of luminescence, but three color maps may convolute the luminescence of different emission lines and conceal the distribution of individual defects. To study the component defects the spectral dataset must be analyzed in full. The top level branches of the dendrogram and spectral shapes calculated via the clustering algorithm outlined in the Method section are shown in Figure 4. The root node of the dendrogram is in the center, with branches emanating above and below. To understand the map, an examination of the spectral shapes in the dendrogram is required. Figure 5 shows a detailed view of the spectral shape labeled (i) in Figure 4. It is generated from the sum of all pixels that fall into subclusters of this branch of the dendrogram and can be explained by decomposing into components from three different defect types. The most intense spectral feature arises from the N-V^0 defect, with the ZPL at 2.15 eV and a set of lower energy phonon replicas. The next most intense feature in
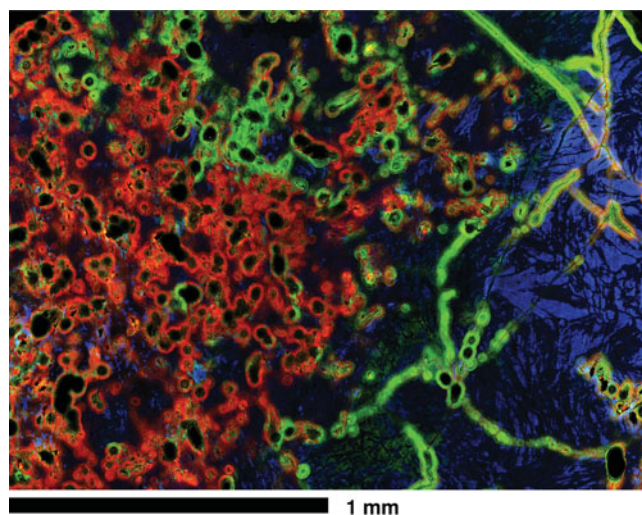

Figure 3. A three color RGB map taken from the sample, using ROI bands of 2.0–2.3 eV for red, 2.3–2.5 eV for green, and 2.5 to 3.1 eV for blue.
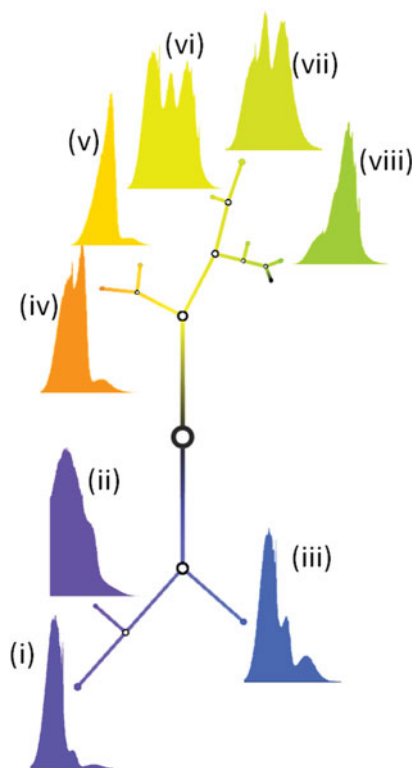

Figure 4. Dendrogram output from the hierarchal clustering algorithm. Representative spectra are plotted on an energy scale ranging from 1.3 to 3.5 eV.

Figure 5 is generated by a N$_2$V defect and the associated set of phonon replicas, and third feature in the spectrum can be fitted with a 3.188 eV ZPL and set of phonon replicas. In addition to the phonon replicas, a broad peak is observed that is not seen in room temperature measurements on synthetic diamonds with the 3.188 eV defect induced (Zaitsev, 2001). This broad band is most likely that known in the literature as the "Blue Band."
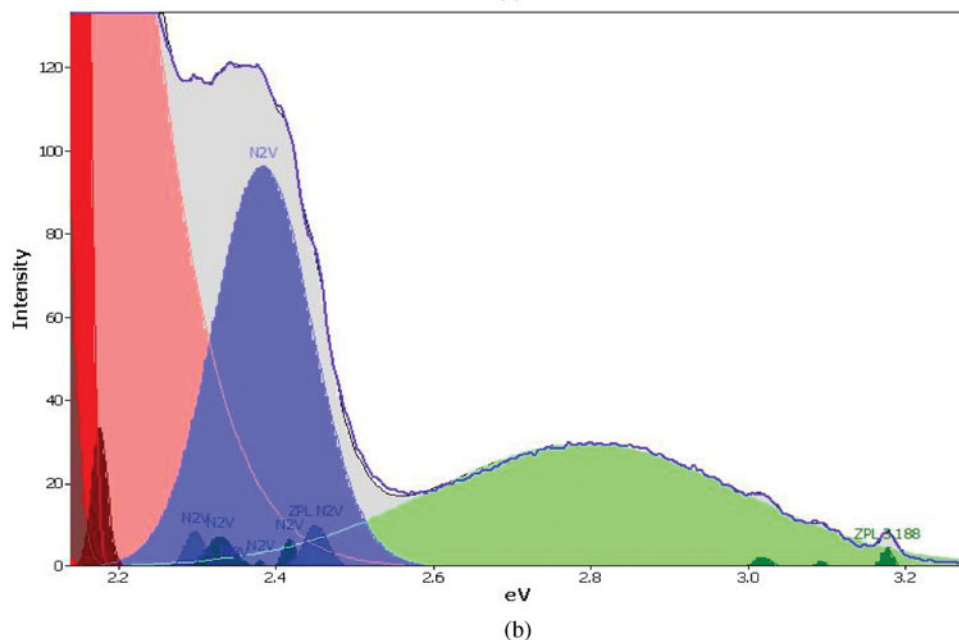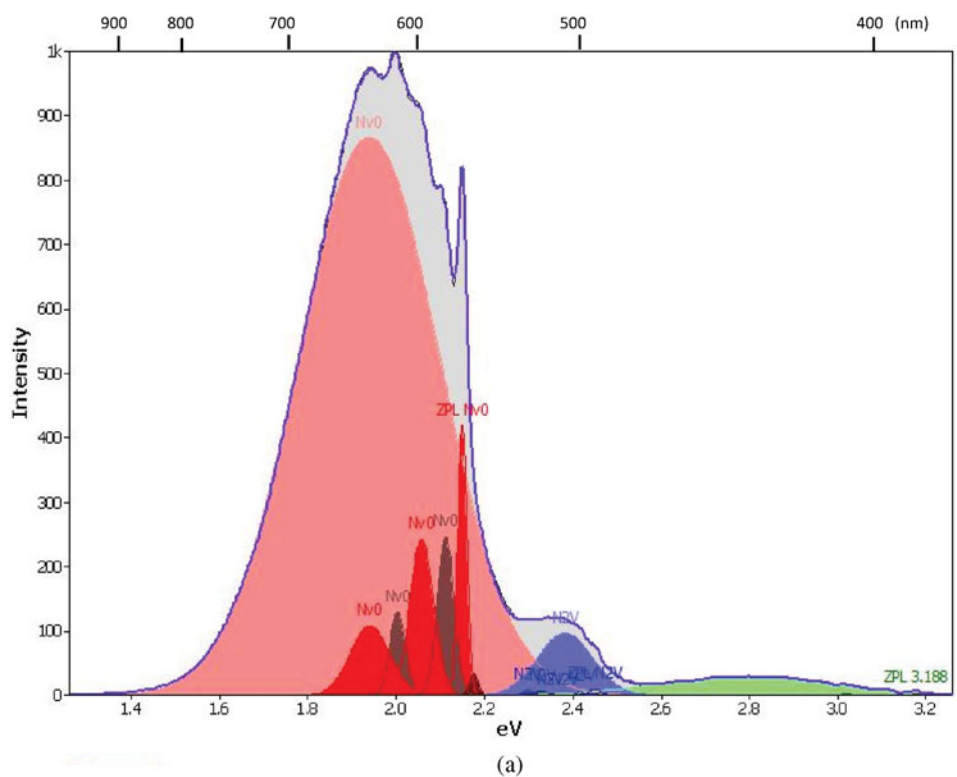
Figure 5. Detailed view showing (**a**) fitted peaks to spectrum (i) given in Figure 4, and (**b**) zoomed region to show detail in the 2.1 to 3.3 eV region.

At room temperature the phonon replicas are more difficult to resolve than when measured at liquid nitrogen (LN) temperature; nevertheless, it was still possible to fit the phonon replicas and calculate $\hbar\omega$ for the various defects. For the N-V^0 it is approximately 50 meV, which compares well to LN measured values of 45 meV (Iancu et al., 2008) and 50 meV (Rondeau et al., 2008). For the N$_2$V defect we measure $\hbar\omega = 35$ meV [LN 40 meV (Davies, 1981; Rondeau et al., 2008) and 39 meV (Iancu et al., 2008)]. This allows us to establish that the defect is the N$_2$V defect, and not a 3H center, which has a ZPL at a similar energy to the N$_2$V defect, but a different energy for the phonon assisted peak

($\hbar\omega = 70$ meV). Identifying the 2.46 eV line with the N$_2$V defect, as opposed to the 3H center, gives important information about the thermal history of the diamond. The 3H center is destroyed by heating at 400–500°C for 1 h (Kagi et al., 1994), while high temperatures form the N$_2$V defect. Care should be taken in interpreting the absence of the 3H center, as it has been shown the sample preparation can anneal this defect (Kagi et al., 2007). For the 3.188 eV ZPL we measure $\hbar\omega = 80$ meV.

The spectral shape labeled (ii) in Figure 4 has a sharp cutoff on the low-energy side, which is attributed to the infrared cutoff of the spectrometer. This cluster is com-
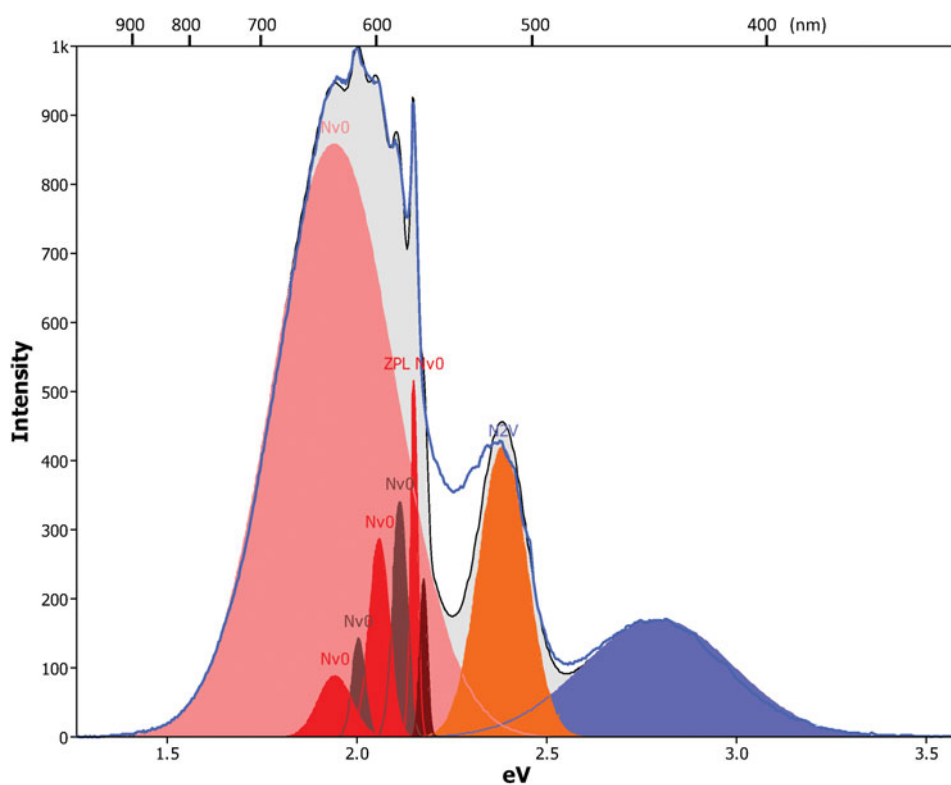
(a)
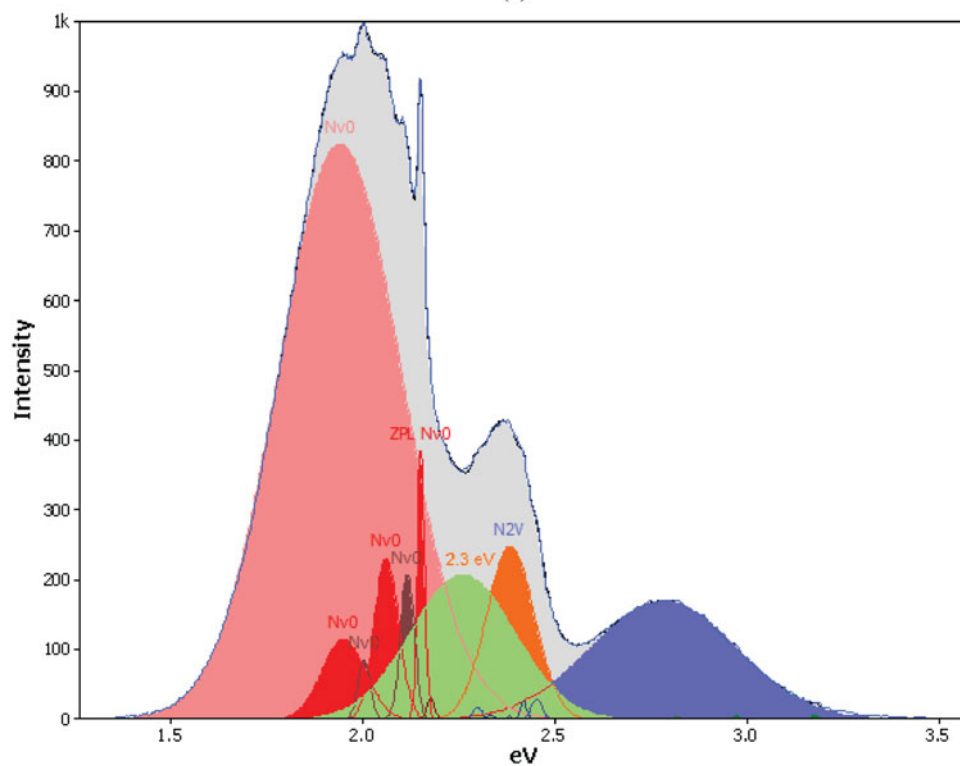
(b)

Figure 6. (**a**) Spectrum (iii) from Figure 4 fitted using peak positions and widths used for spectrum (i) in Figure 5, showing that these peaks cannot reproduce the spectrum. (**b**) Fit of spectrum (iii) with the addition of another Gaussian centered at 2.3 eV.

prised of 60 pixels (0.0005% of the mapped pixels), having a broad peak centered around 1.8 eV and is from a small particle trapped within an iron silicate mineral inclusion and can most likely be attributed to a piece of embedded polishing material. This spectrum from these 60 pixels is quite different from the rest of the collected pixels and is therefore placed toward the top of the dendrogram, as the greatest differences are the last to be merged. The remaining spectral shape on the bottom half of the dendrogram [Fig. 4(iii)] is shown in detail in Figure 6. Figure 6a shows
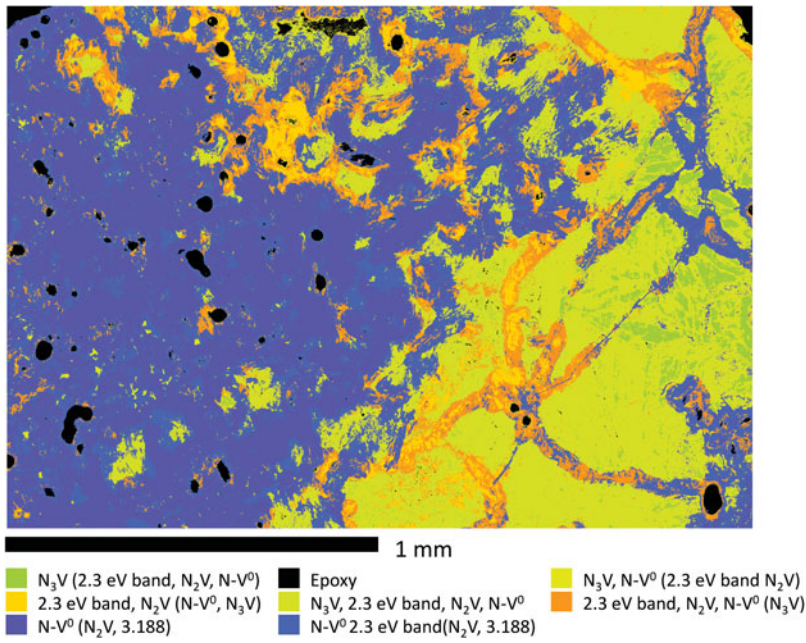
N₃V (2.3 eV band, N₂V, N-V⁰)   Epoxy   N₃V, N-V⁰ (2.3 eV band N₂V)
2.3 eV band, N₂V (N-V⁰, N₃V)   N₃V, 2.3 eV band, N₂V, N-V⁰   2.3 eV band, N₂V, N-V⁰ (N₃V)
N-V⁰ (N₂V, 3.188)   N-V⁰ 2.3 eV band(N₂V, 3.188)

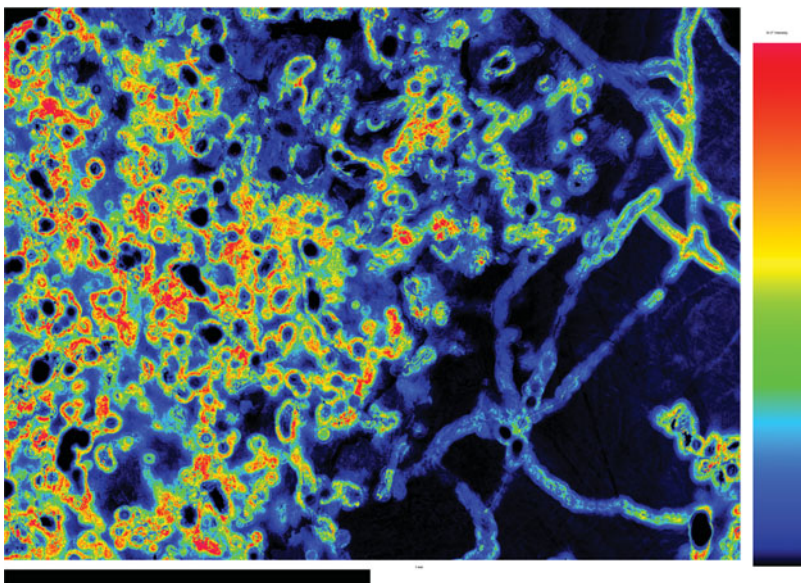**Figure 7.** Pixel classification map for the hyperspectral dataset.

Figure 8. N-V⁰ defect CL intensity map.

this spectrum fitted using Gaussians of the same width and position used to fit the spectrum in Figure 5. However, these peaks cannot completely describe the peak shape, and an additional broad peak at 2.3 eV is required (Fig. 6b).

In the top half of the dendrogram, spectrum (iv) is composed of major peaks of N_2V, $N\text{-}V^0$, and the band at 2.3 eV, with a minor contribution from a N_3V defect. The difference between spectrum (v) and spectrum (iv) is that the $N\text{-}V^0$ component is reduced. Spectra (vi)–(viii) are all differentiated by containing a significant contribution from a N_3V defect. All three of these peaks contain smaller contributions from the N_2V defect and the 2.3 eV band, and a $N\text{-}V^0$ peak that decreases in intensity from spectrum (vi) to (viii). For the N_3V defect, $\hbar\omega$ is measured to be 80 meV compared with other measurements of 78 meV (Rondeau et al., 2008) and 80 meV (Zaitsev, 2001).

From this examination of the spectral shapes, qualitative names can be given to the clusters in terms of the contributing defects and these are given in the classification map (Fig. 7). It shows a number of black regions within the diamond, which are mineral inclusions filling voids within the carbonado diamond. The left-hand side of the map is dominated by spectra whose major contributor is from the $N\text{-}V^0$ defect, interspersed with patches where the 2.3 eV band becomes a contributor, with some patches surrounding pore-filling inclusions. The right-hand side of the image is dominated by phases where the N_3V defect is the major contributor to the spectra. The 2.3 eV band shows a correlation with radiation damage in this sample. However, the source of the band could not be uniquely determined by this CL study and may require further investigation by a complementary technique such as ESR.
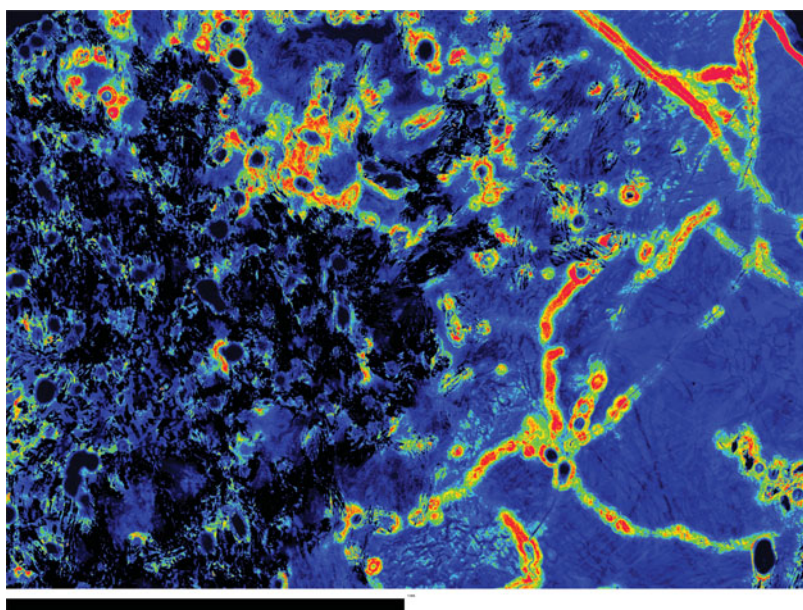
Figure 9. 2.3 eV band CL intensity map.

While the sample may contain up to five defect types, the spectral shapes of all but the 2.3 eV band are complex, containing not just a ZPL but a set of phonon replicas, requiring a number of Gaussian functions to reproduce. This leads to a large number of Gaussian functions being needed to fit all possible defect shapes. The partitioning of the pixels using data clustering has eased this task reducing problems with overfitting. Deconvolution of the spectra then allows the projections of the CL intensity of individual defects. Figure 8 shows the projection of the CL intensity of the N-V^0 defect and Figure 9 that of the 2.3 eV band. Within this sample, the 2.3 eV band signal is always over-lapped with peaks from other defects; this illustrates one of the great advantages of the hyperspectral data collection, as maps like these require that spectral deconvolution be performed.

CONCLUSIONS

Hyperspectral CL mapping was performed on a carbonado diamond, and the acquired dataset was subsequently investigated using cluster analysis. The CL response from this particular carbonado diamond can be attributed to a small number of defect types: N-V^0, N$_2$V, N$_3$V, a 3.188 eV line that is seen in diamonds after radiation damage by 2 MeV electrons followed by annealing, and a broad luminescence band. Both the N$_2$V and 3.188 eV defects require high-temperature annealing, which has implications regarding the thermal history of the diamond. Examination of the size of the radiation damage halos correlates well with those calculated for the ^{238}U chain decay series.

The collection of a hyperspectral dataset makes possible the extraction of individual defect maps because there is a spectrum collected for each pixel that can be deconvoluted. However, the spectral structure for each defect type is not simple. Along with the purely electronic ZPL, a set of phonon replicas makes for a complex spectral structure for each of the defects, and therefore the use of data clustering makes this deconvolution task manageable.

REFERENCES

BARANOV, P.G., IL'IN, I.V., SOLTAMOVA, A.A., VUL'L, A.Y., KIDALOV, S.V., SHAKHOV, F.M., MAMIN, G.V., ORLINSKI, S.B. & SOLA-KHOV, M.K. (2009). Electron spin resonance detection and identification of nitrogen centers in nanodiamonds. *JETP Lett* **89**, 409–413.

BRAGG, W.H. & KLEEMAN, R. (1905). Alpha particles or radium, and their loss of range passing through various atoms and molecules. *Philos Mag* **10**, 318–334.

CARTIGNY, P. (2010). Mantle-related carbonados? Geochemical insights from diamonds from the Dachine komatiite (French Guiana). *Earth Planet Sci Lett* **296**(3–4), 329–339.

COLLINS, A.T. (1992). The characterisation of point defects in diamond by luminescence spectroscopy. *Diam Relat Mater* **1**, 457–469.

COLLINS, A.T. & LAWSON, S.C. (1989). Cathodoluminescence studies of isotope shifts associated with localised vibrational modes in synthetic diamond. *J Phys-Condens Mat* **1**, 6929–6937.

DAVIES, G. (1981). The Jahn-Teller effect and vibronic coupling at deep levels in diamond. *Rep Prog Phys* **44**, 787–830.

DEMENY, A., NAGY, G., BAJNOCZI, B., NEMETH, T., GARAI, J., DROZD, V. & HEGNER, E. (2011). Hydrogen isotope compositions in carbonado diamond: Constrains on terrestrial formation. *Cent Eur Geol* **54**, 51–74.

GARAI, J., HAGGERTY, S., REKHI, S. & CHANCE, M. (2006). Infrared absorption investigations confirm the extraterrestrial origin of Carbonado-diamonds. *Astrophys J* **653**, L153–L156.

HARTE, B., FITZSIMONS, C.W., HARRIS, J.W. & OTTER, M.L. (1999). Carbon isotope ratios and nitrogen abundances in relation to cathodoluminescence characteristics for some diamonds from the Kaapvaal Province, S. Africa. *Mineral Mag* **63**, 829–856.

Heaney, P.J., Vicenzi, E.P. & De, S. (2005). Strange diamonds: The mysterious origins of carbonado and framesite. *Elements* **1**, 85–89.

Iancu, O.G., Cossio, R., Korsakov, A.V., Compagnoni, R. & Popa, C. (2008). Cathodoluminescence spectra of diamonds in UHP rocks from the Kokchetav Massif, Kazakhstan. *J Lumin* **128**, 1684–1688.

Jain, A.K., Murty, M.N. & Flynn, P.J. (1999). Data clustering: A review. *ACM Comput Surv* **31**, 264–318.

Kagi, H., Sato, S., Akagi, T. & Kanda, H. (2007). Generation history of carbonado inferred from photoluminescence spectra, cathodoluminescence imaging, and carbon-isotopic composition. *Am Mineral* **92**, 217–224.

Kagi, H., Takahashi, K., Hidaka, H. & Masuda, A. (1994). Chemical properties of Central African carbonadao and it genetic implications. *Geochim Cosmochim Acta* **58**, 2629–2638.

Kotula, P.G., Keenan, M.R. & Michael, J.R. (2003). Automated analysis of SEM X-ray spectral images: A powerful new microanalysis tool. *Microsc Microanal* **9**, 1–17.

MacQueen, J.B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

MacRae, C.M., Wilson, N.C., Johnson, S.A., Phillips, P.L. & Otsuki, M. (2005). Hyperspectral mapping—Combining cathodoluminescence and X-ray collection in an electron microprobe. *Microsc Res Techniq* **67**, 271–277.

Magee, C.W. (2001). Geologic, microstructural, and spectroscopic constraints on the origin and history of carbonado diamond, p. 247. PhD Thesis. Canberra, Australia: Research School of Earth Sciences, Australian National University.

McCall, G.J.H. (2009). The carbonado diamond conundrum. *Earth Sci Rev* **93**, 85–91.

Owen, M.R. (1988). Radiation-damage halos in quartz. *Geology* **16**, 529–532.

Rondeau, B., Sautter, V. & Barjon, J. (2008). New columnar texture of carbonado: Cathodoluminescence study. *Diam Relat Mater* **17**, 1897–1901.

Ross, G.J.S. (1968). Classification techniques for large sets of data. In *Numerical Taxonomy*, Cole, A.J., (Ed.). New York: Academic Press, Inc.

Sautter, V., Lorand, J.-P., Cordier, P., Rondeau, B., Leroux, H., Ferraris, C. & Pont, S. (2011). Petrogenesis of mineral micro-inclusions in an uncommon carbonado. *Eur J Mineral* **23**, 721–729.

Shames, A.I., Osipov, V.Y., von Bardeleben, H.J. & Vul', A.Y. (2012). Spin S=1 centers: A universal type of paramagnetic defects in nanodiamonds of dynamic synthesis. *J Phys-Condens Mat* **24**, 225302-1–8.

Smith, J.V. & Dawson, J.B. (1985). Carbonado: Diamond aggregates from early impacts. *Geology* **13**, 342–343.

Stork, C.L. & Keenan, M.R. (2010). Advantages of clustering in the phase classification of hyperspectral materials images. *Microsc Microanal* **16**, 810–820.

Taylor, W.R., Canil, D. & Milledge, H.J. (1996). Kinetics of Ib to IaA nitrogen aggregation in diamond. *Geochim Cosmochim Acta* **60**, 4725–4733.

Trueb, L.F. & Christiaan De Wys, E. (1969). Carbonado: Natural polycrystalline diamond. *Science* **165**, 799–802.

Vicenzi, E.P., Rose, T., Fries, M., Steel, A. & Magee, C. (2006). A cathodoluminescence (and Raman) imaging and spectroscopic study of ancient polycrystalline diamond. *Microsc Microanal* **12**(Suppl 2), 1518–1519.

Walker, J. (1979). Optical absorption and luminescence in diamond. *Rep Prog Phys* **42**, 1605–1659.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58**, 236–244.

Wilson, N.C. & MacRae, C.M. (2005). An automated hybrid clustering technique applied to spectral data sets. *Microsc Microanal* **11**(Suppl 2), 434–435CD.

Wilson, N.C., MacRae, C.M. & Torpy, A. (2008). Analysis of combined multi-signal hyperspectral datasets using a clustering algorithm and visualisation tools. *Microsc Microanal* **14**(Suppl 2), 764–765.

Zaitsev, A.M. (2001). *Optical Properties of Diamond*. Berlin: Springer.