Multivariate Statistical Analysis of Combined Hyperspectral Cathodoluminescence and X-Ray Datasets

P. R. Edwards,* R. W. Martin,* and M. R. Lee**

* Department of Physics, SUPA, University of Strathclyde, Glasgow, G4 0NG, U.K.
** Department of Geographical & Earth Sciences, University of Glasgow, Glasgow G12 8QQ, U.K

Cathodoluminescence (CL) is a well-established diagnostic tool in a number of fields, particularly within the geosciences and semiconductor research. The usefulness of the technique stems from its high sensitivity to a wide range of factors including subtle changes in material composition, although the complex nature of the data often necessitates its comparison with more readily quantifiable microanalysis tools such as wavelength- or energy-dispersive X-ray spectroscopy (WDX/EDX). A relatively recent advance in CL is the development of hyperspectral imaging (or *spectrum mapping*), in which a full emission spectrum is acquired at each position in a 2-dimensional scan [1]. This multidimensional dataset contains the full set of spatially- and spectrally- resolved luminescence information, combining the benefits of the CL microscopy and spectroscopy techniques.

A facility has been developed for acquiring such CL hyperspectral images *simultaneously* with WDX/EDX composition maps [2-3], directly combining the high sensitivity of CL with the quantitative nature of the X-ray data whilst eliminating any alignment error between the two measurements. However, the large size of the resultant CL/X-ray dataset presents new challenges in data analysis. Although the data is a function of a large number of dimensions, given by the sum of the number of spectral channels (typically ~1000) and number of elemental X-ray lines mapped (~10), the true number of significant, independent underlying variables will be considerably smaller.

In this work, different methods of reducing the number of variables have been employed in order to aid data interpretation. One such technique is principal component analysis (PCA), in which the data is re-described as a linear combination of spectral/composition components. The components (or *eigenspectra*) are ordered in terms of the magnitude of their contribution to the dataset, allowing all but the most significant components to be discarded and the dimensionality of data reduced.

We have investigated a number of luminescent systems using this technique, including zircons, calcite and opal. Fig. 1 shows a subset of a CL hyperspectral image of a calcite sample; the details of this sample and the significance of the observed luminescence is described elsewhere [3]. The results of PCA on the full combined CL/X-ray dataset are shown in Fig. 2. In this example, the first two principal components correspond closely to the spectral features identified and mapped manually in Fig. 1. In addition, this analysis yields the dependence of each dominant spectral feature on the composition. Positive and negative correlations between luminescence bands and the abundance of various elements are observed, and these have been shown to correspond to known physical effects.

Further statistical methods have been investigated which aim to maintain the physical meaningfulness of the data by imposing non-negativity constraints on the spectra, as have been employed in the analogous problem of EDX hyperspectral image analysis [4]. This has resulted in the successful application to our data of the non-negative matrix factorization algorithm of Lee and Seung [5], further simplifying the interpretation of an already powerful characterization technique.

References

[1] J. Christen et al., J. Vacuum Sci. Technol. B 9 (1991) 2358.
[2] P. R. Edwards et al., phys. stat. sol. (c) 0 (2003) 2474.
[3] M. R. Lee et al., J. Sedimentary Res. 75 (2005) 313.
[4] P. G. Kotula et al., Microsc. Microanal. 9 (2003) 1
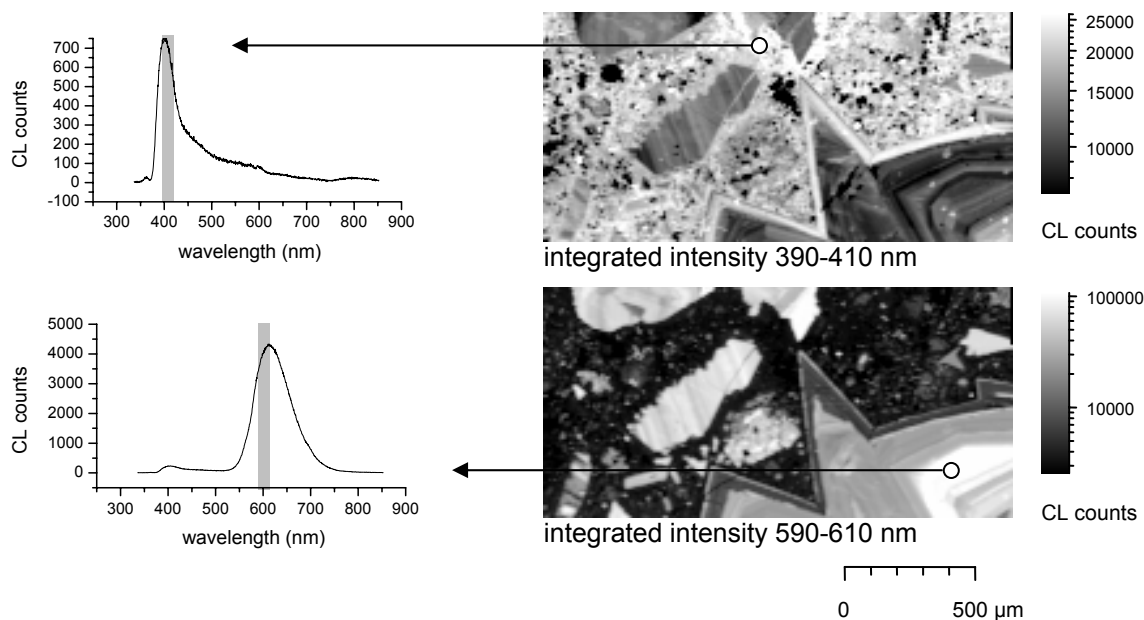[5] D. D. Lee and H. S. Seung, Nature 401 (1999) 788.

Fig.1. Results of manual analysis of a CL hyperspectral image of calcite, showing wavelength "slices" corresponding two dominant spectral features in the blue and red regions of the spectrum. A characteristic single-pixel spectrum from bright area of each image is also shown.
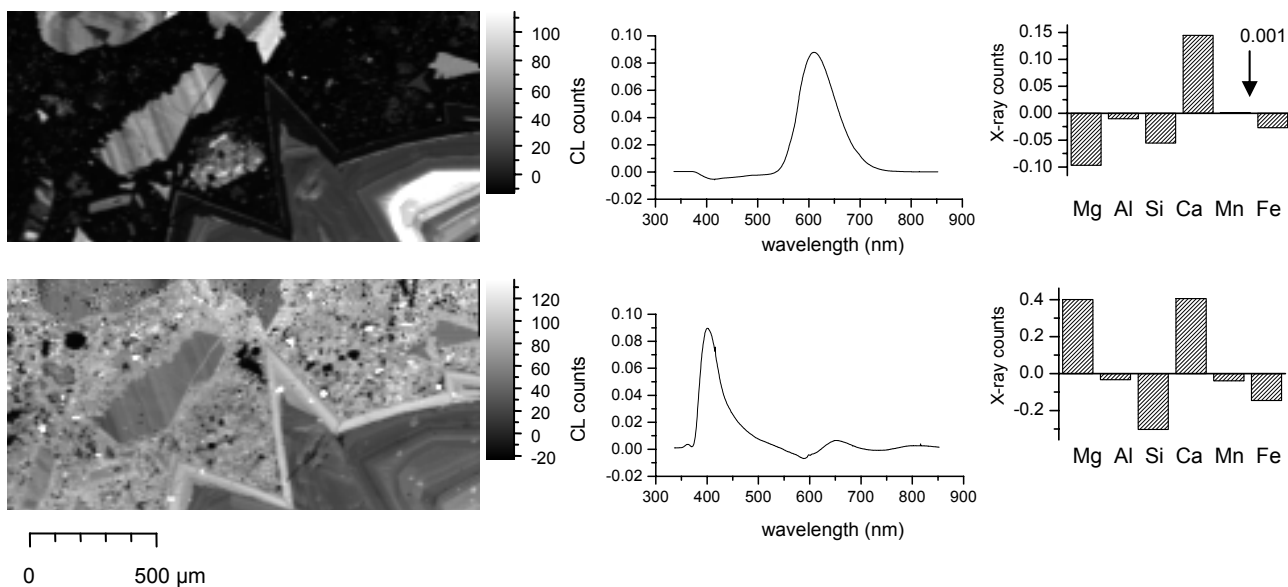

Fig. 2. First two principal components (each containing CL/composition data) of the same hyperspectral CL dataset as Fig. 1 combined with the simultaneously-acquired WDX/EDX data.