All Mixed Up: Using Machine Learning to Address Heterogeneity in (Natural) Materials

J.F. Einsle [1, 2], Ben Martineau[2], Iris Buisman[1], Zoja Vukmanovic[1], Duncan Johnstone[2], Alex Eggeman[4] and Paul A. Midgley[2], Richard J. Harrison[1]

[1.] Department of Earth Sciences, University of Cambridge, Cambridge, UK
[2.] Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, UK
[3.] CEA-LETI, Grenoble, France
[4.] School of Materials, University of Manchester, Manchester, UK

Modern electron microscopes offer a variety of techniques for mapping chemical and crystallographic information in samples from micron to nanometer scales in both two and three dimensions. A typical data set comprises hundreds of thousands of spectra acquired pixel by pixel as the electron beam scans over the specimen. Routine analysis procedures include background subtraction, peak identification and quantification. However, while this approach may work on point spectra and on systems where the nanostructures are not buried, it is a physically impossible task when working with an elemental map of millions of spectra. Application of data decomposition techniques (similar to Principal Component Analysis (PCA)) to multidimensional data sets reveals key and often overlooked features. The use of machine learning in a wide range of systems has been demonstrated to allow for the unmixing of buried chemical structures and identification of phases [1, 2]. This paper will demonstrate how the use of machine learning on Energy Dispersive Spectroscopic (EDS) data sets allows for the unique identification of distinct chemical phases and present insights into the limitations of this approach.

SEM-EDS (Scanning Electron Microscopy) maps of a naturally formed eutectoid texture from an igneous rock (Fig. 1) demonstrates both the power of machine learning to pull out material phases, but also the limitations of EDS as a technique for phase identification classification. Using the decomposition method of Non-negative Matrix Factorization (NMF) produces the four mineral phases identified in Fig. 2. However, the spectrum labeled biotite (Fig. 2c) is missing several elements that would identify the phase as biotite. Importantly in this spectrum, silicon is completely absent. Due to the compositional similarity of many natural materials (in this case all four phases are Si and O rich), EDS characterization does not fully describe the material system. Since spectrographic data does not include any crystallographic information, the boundary conditions of a fully described data set are not met. Crystal structure is the other key bit of information to uniquely identify a material phase. This highlights the need for combining composition information with crystallographic data. I will present a method for addressing this through the use of a statistical spectral model of the system, combined with NMF based phase masks (Fig. 3). This allows us to identify the correct material components and perform simple standardless quantification verification. Further, I will present the potential advantages of analyzing the EDS and Electron Back-Scatter Diffraction (EBSD) measurements concurrently. Crucially this paper demonstrates that a data driven approach to analysis better equips researchers to make better decisions in how to proceed with sample characterization as opposed to presenting a definitive answer [3].

References:

[1] Z. Saghi *et al*, Nano Letters **11** (2011), p. 4666.
[2] A. S. Eggeman, R. Krakow, and P. A. Midgley, Nature Communications **6** (2015), p. 1.
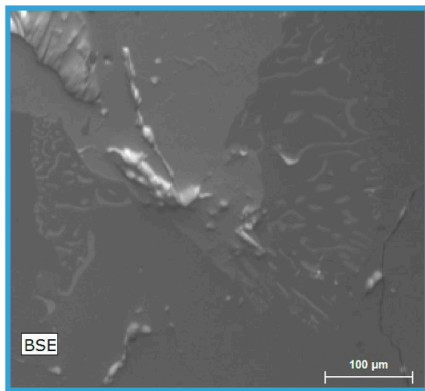
Figure 1. Back Scattered electron micrograph of eutectic texture in natural igneous sample.
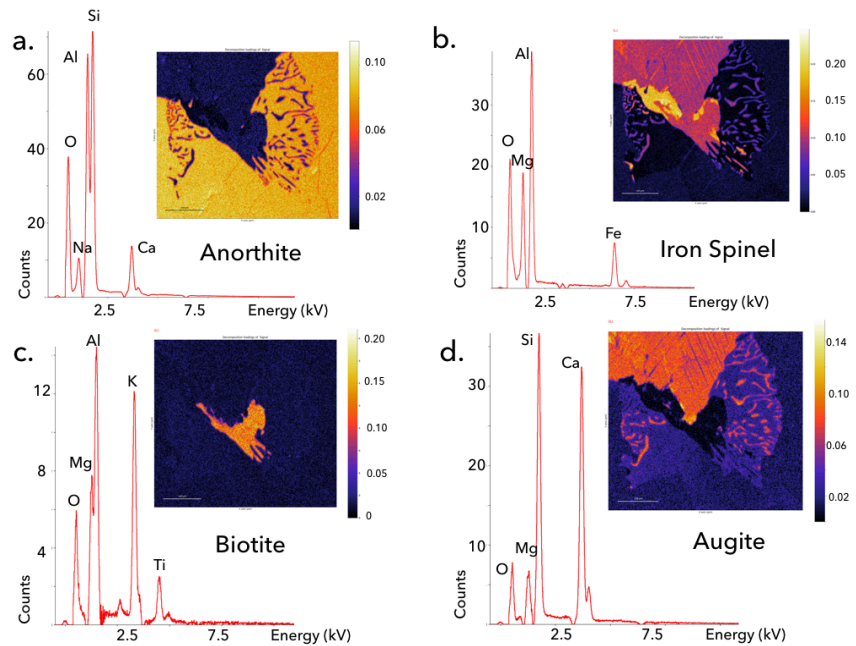

Figure 2. Non-negative matrix factorization factors and loading maps show the presence of four material phases. Note that the raw factors are not complete phase compositions due impact factorization of the phases.
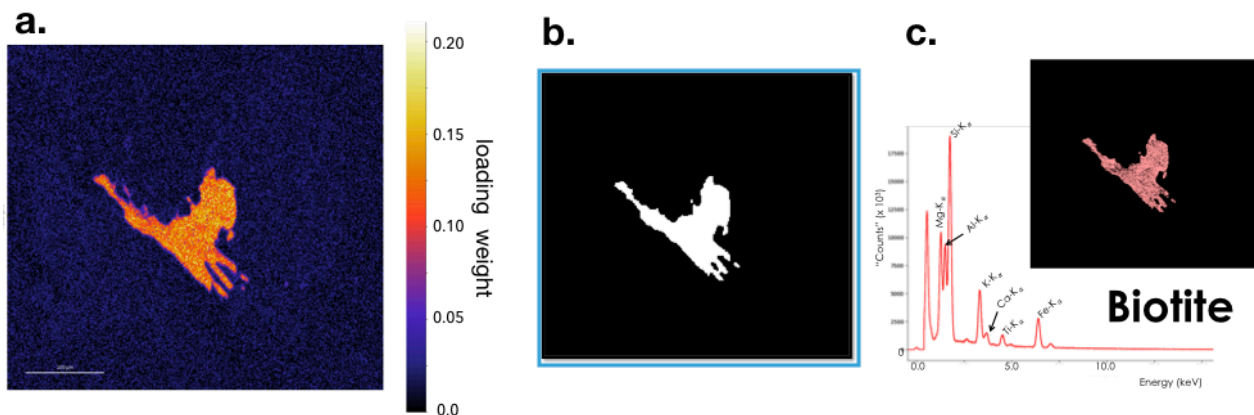

Figure 3. transformation of NMF loading map (a) into a phase mask (b) and application to statistical model of data (c). This recovers to the complete phase composition.