

An Open Evaluation of Hyperspectral Unmixing Strategies for EDS Analysis

Joshua A. Taillon¹

¹ Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA

Oftentimes in an X-ray energy dispersive spectroscopy (EDS) analysis, the desired result is a quick understanding of what phases are present within a dataset rather than an exact quantitative elemental characterization. A clear picture of what elements are spatially correlated and where they are located is often enough to answer the research question at hand. Commercial EDS vendors have realized this, and almost every available software package includes some option of “phase mapping” (*e.g.* Oxford – *AutoPhaseMap*, EDAX – *Smart Phase Mapping*, Bruker – *AutoPhase*, etc.). Most operate as some proprietary form of principal component analysis (PCA), and are often “black box” in nature, with no indication of what algorithm was used or how the results were obtained. This presents a significant challenge for comparisons between different vendors’ systems and hinders efforts towards more open and reproducible science.

Thankfully, EDS mapping is not unique in terms of the need to identify “pure phases” within hyperspectral data; the fields of remote sensing and chemometrics have been developing factor analysis tools for this problem since the 1970s [1], and many of the methods used in those disciplines can be directly applied to EDS data with only minor changes to their implementation. In fact, the first examples of EDS factor analysis were reported over 20 years ago [2], but the methods have seen only limited use since that time, due in part to “lock-in” from vendor software and the lack of easy to use tools. Considering this, the recent development of open-source software tools for hyperspectral data analysis that can read the data output by proprietary vendor software packages (*e.g.* HyperSpy [3]) has made it relatively simple to perform advanced EDS analysis beyond the scope of the vendor software packages.

In an EDS spectrum image, each spatial position contains a spectrum that is a mixture of signals from one or more “prototype” spectra or “pure phases.” The goal of hyperspectral unmixing is to accurately determine those spectral prototypes (called components or endmembers) and the strength or weight of each component at each pixel (called a loading or score map). This is an unsupervised machine learning problem of factor analysis, where the computer code should identify and describe the interesting components, without input from the user. There are many algorithms available to perform this task, each with associated benefits and drawbacks depending on the assumptions made by the algorithm.

In this work, numerous methods of spectral unmixing have been performed on multiple simulated and experimental EDS spectrum images to compare their performances. Among the algorithms examined are: PCA + independent component analysis (ICA) [4], non-negative matrix factorization (NMF) [5] [6], multivariate curve resolution (MCR) [1], vertex component analysis (VCA) [7], Bayesian linear unmixing (BLU) [8], and simplex identification via split augmented Lagrangian (SISAL) [9].

An example of one of these algorithms is displayed in Figure 1. The demonstration data is a two-dimensional EDS spectrum image of a jadeite/omphacite geological sample acquired with an electron beam accelerating voltage of 15 kV. The proper number of components to include (4) was determined via inspection of a scree plot, and an NMF analysis was run. As expected, this algorithm produced 4

easily interpretable components with positive spectral signatures. Each component corresponds to a different material phase within the dataset, allowing for easy determination of the sample structure. In general, NMF has been found to be one of the most reliable unmixing procedures for EDS data in comparison to the other investigated algorithms. [10]

References:

- [1] WH Lawton and EA Sylvestre, *Technometrics*. **13** (1971), p. 617.
- [2] JM Titchmarsh and S Dumbill, *Journal of Microscopy*, **184** (1996), p. 195.
- [3] F de la Peña *et al.*, *HyperSpy 1.3*, (2017).
- [4] A Hyvarinen and E Oja, *Neural Networks* **13** (2000), p. 411.
- [5] DD Lee and HS Seung, *Nature* **401** (1999), p. 788.
- [6] O Nicoletti, *et al.*, *Nature* **502** (2013), p. 80.
- [7] JMP Nascimento and JM Bioucas-Dias, *IEEE Transactions on Geoscience and Remote Sensing* **43** (2005), p. 898.
- [8] N Dobigeon *et al.*, *IEEE Transactions on Signal Processing* **57** (2009), p. 4355.
- [9] JM Bioucas-Dias, *First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (2009) p. 1.
- [10] Commercial materials and products identified in this article are not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is the intention of this article to imply that the materials identified are necessarily the best available for the stated purpose.

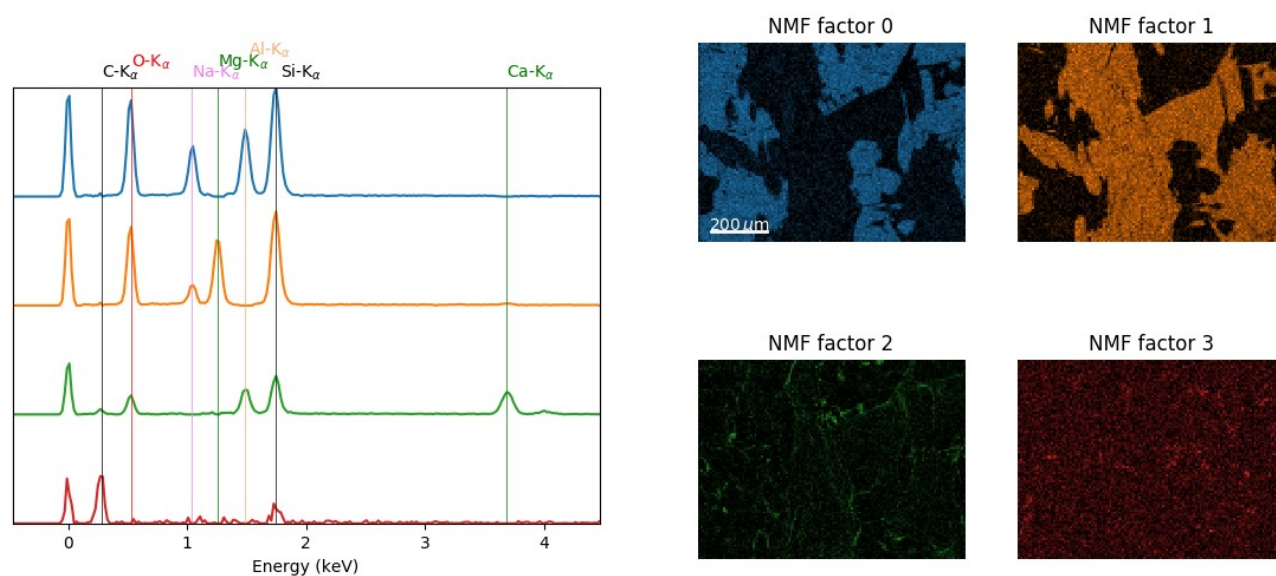


Figure 1. An example of hyperspectral unmixing on a jadeite/omphacite sample using a non-negative matrix factorization (NMF) technique. Such a method is particularly well-suited to unmixing EDS spectral data due to the positivity constraint enforced during the matrix decomposition. In this example, the two primary phases (labeled 0 and 1) are successfully identified, together with two additional phases (2 and 3) that were difficult to observe in the raw data, corresponding to Ca-rich deposits at grain boundaries (factor 2), and a buildup of organic matter at certain locations on the sample (factor 3).