2978 Microsc. Microanal. 28 (Suppl 1), 2022
doi:10.1017/S1431927622011163

Physics-Guided Machine Learning for the Analysis of Low SNR STEM-EDXS Data

Adrien Teurtrie[1,2], Nathanaël Perraudin[3], Thomas Holvoet[3,4], Hui Chen[1], Duncan T. L. Alexander[1], Guillaume Obozinski[3], Cécile Hébert[1,*]

[1.] Electron Spectrometry and Microscopy Laboratory, Institute of Physics, EPFL, Lausanne
[2.] Current address : Unité Matériaux et Transformations, UMET, UMR-CNRS 8207, Université de Lille, France.
[3.] Swiss Data Science Center, EPFL and ETH Zürich, Lausanne and Zürich
[4.] Ghent University, Ghent.
* Corresponding author: cecile.hebert@epfl.ch

The combination of Scanning Transmission Electron Microscopy (STEM) and Energy-Dispersive X-ray (EDX) spectroscopy is a very useful tool for the analysis of samples containing several phases. However, EDX is a secondary process of electron ionization, and the data are often noisy, with a poor signal to noise ratio (SNR). This is particularly so when the sample can only tolerate a limited electron dose. Additionally, the microstructure of the sample can be complex, thus rendering the recovery of the spectra of the constituent phases challenging. To tackle these problems, the community has been using machine learning tools such as Non-negative Matrix Factorization (NMF) [1]. These algorithms perform a statistical decomposition of the data. However, they do not take full advantage of it because they ignore prior knowledge on the physics of the problem. Further, in some cases, a chemical quantification cannot be directly performed on the solutions. To overcome these issues, here we combine knowledge in the fields of EDX spectroscopy and machine learning to design a physics-guided decomposition algorithm based on a structured variant of a Poisson likelihood NMF, that we call “esmpy”.

The algorithm incorporates a physics-based modelling of the EDX spectrum with a custom-designed factorization model. The data Y are modelled with the product GWH (see Figure 1). G is a constant matrix, whose columns are the EDX spectra of individual elements and the bremsstrahlung model. The rows of W represent the learned weights of the different elements and the parameters of the bremsstrahlung for each phase. The columns of H correspond to the learned abundances of each phase at a given pixel of the datacube. The elemental EDX spectra are simulated based on the ionization cross-sections of Bote et al. [2] and on radiative rates of Schoonjans et al. [3]. A new bremsstrahlung model is proposed with a non-negative and linear parametrization inspired from Lifshin [4]. The model is learned by optimizing a Poisson likelihood using multiplicative updates. Moreover, constraints are added to the optimization. In addition to the non-negativity constraint, a sum-to-one constraint is applied on H, given that abundances are proportions that should sum to one. This way, the columns of the GW product correspond to EDX spectra and thus, the rows of W are proportional to chemical concentrations.

To assess the performance of the algorithm, we apply it to simulated data, based on artificial geological samples where two minor phases Ferropericlase (Fp) and Ca-Perovskite (Ca-Pv) are embedded in a Bridgmanite (Brg) main phase [5]. In Figure 2, we test its limits on noisy data with very poor SNR. First trying the standard Poisson likelihood NMF of scikit-learn [6], this gives results where the different phases cannot be identified. In contrast, using our esmpy algorithm, the distributions of all 3 phases are seen in their abundance maps. Both algorithms fail to correctly assign some of the peaks. However, the shape of some the peaks in the output of NMF are heavily crippled by noise, which can make their intensities difficult to extract, in constrast to the spectra obtained using esmpy. These results show that, by including physics-based models, esmpy identifies important features of a dataset even when the data are heavily crippled by noise.

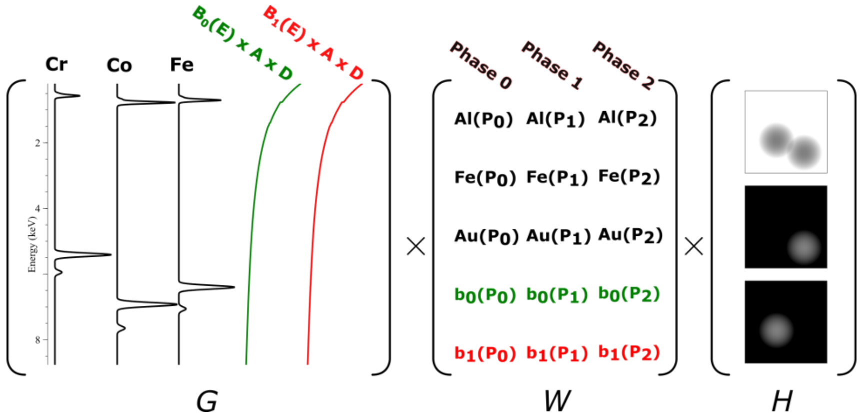

Figure 1. Example of the different elements composing the EDX modelling matrix *G*, learned weights of *G* columns *W* and the learned abundance *H*

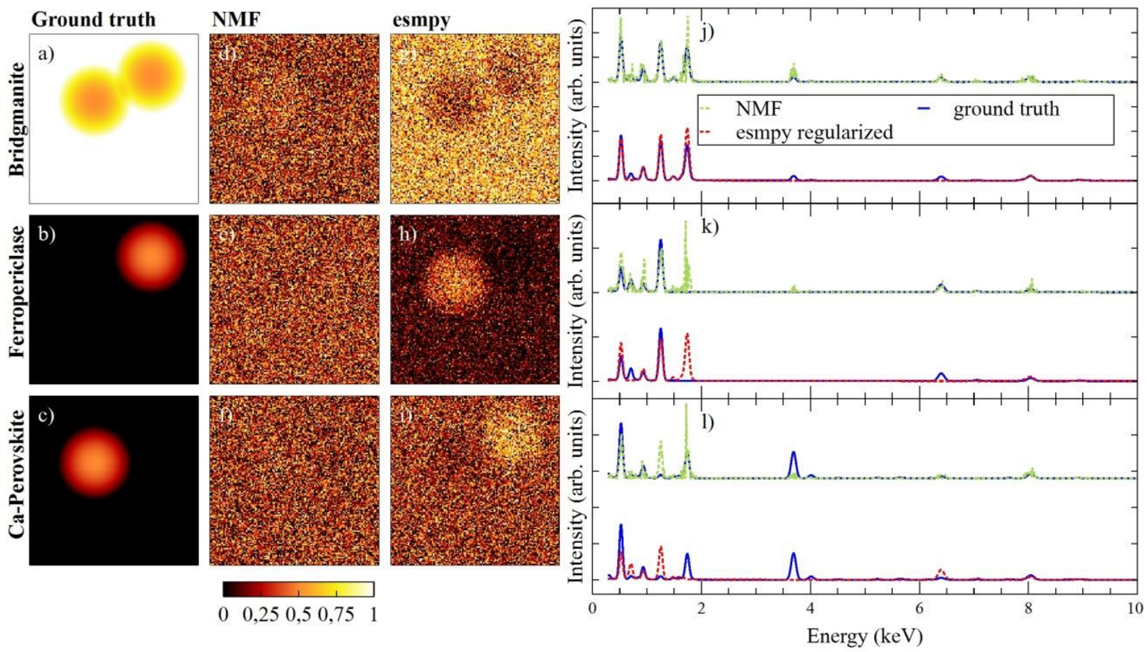

Figure 2. a, b and c) Ground truth abundance maps of the 3 phases in simulated dataset. d, e and f) Abundance maps as obtained with NMF. g, h and i) Abundance maps as obtained with esmpy. j, k and l) Comparisons between the spectra of the ground truth and the results of NMF and esmpy for Brg, Fp and CaPv, respectively.

References:

[1] D. D. Lee, H. S. Seung, Nature 401 (6755) (1999)
[2] D. Bote et al., Atomic Data and Nuclear Data Tables 95 (6) (2009) 871–909
[3] T. Schoonjans et al., Spectrochimica Acta Part B 66 (2011) 776–784.
[4] E. Lifshin, Proc. 9th Annual Conf. Microbeam Analysis Society, Vol. 53, 1974
[5] F. Nabiei et al., Geophysical Research Letters 48 (12) (2021).
[6] F. Pedregosa et al., Journal of Machine Learning Research 12 (2011) 2825–2830.

https://doi.org/10.1017/S1431927622011163 Published online by Cambridge University Press