

Unmixing Mineral Phases, Improving Quantification: Use Machine Learning to Understand Deep-Mantle with STEM-EDS Data

Hui Chen^{1*}, James Badro^{2,3}, Duncan T.L. Alexander¹ and Cécile Hébert¹

¹. Electron Spectrometry and Microscopy Laboratory (LSME), IPHYS, EPFL, Lausanne, Switzerland

². Earth and Planetary Science Laboratory (EPSL), IPHYS, EPFL, Lausanne, Switzerland

³. Institute de Physique du Globe de Paris, Sorbonne Paris Cité, Paris, France

* Corresponding author: hui.chen@epfl.ch

In geochemistry, (scanning) transmission electron microscopy ((S)TEM) has emerged as an advantageous technique for studying minerals and rocks for its excellent spatial resolution (down to sub-nm). Further, using STEM mode for energy-dispersive X-ray spectroscopy (EDS) mapping enables a robust and quick analysis of the composition of samples. However, several problems currently limit the analytical capability of STEM-EDS. First, the commonly used two-dimensional (2D) EDS scan merely captures an average of the 2D chemical information projected along the probe's direction, which can be problematic for phase quantification when phases are severely overlapped in the projection. While electron tomography can help retrieve the chemical composition of each phase in this overlapping scenario, it is not feasible for materials which are beam sensitive such as the Earth mantle minerals of interest here. Second, the fact that TEM samples are thin results in low counting statistics for the STEM-EDS signal, limiting its sensitivity and reliability when analyzing minor and trace elements. Considering data analysis in general, machine learning (ML) has begun to make an impressive impact in various fields [1]. Here we address the challenges of STEM-EDS quantification by exploiting several ML algorithms, and appropriately applying them to the data analysis. Ultimately, we can automatically identify, segment, and precisely quantify mineral phases in the heavily overlapping scenario. This improved phase quantification further enables us to gain a better understanding of the mineralogy of the deep Earth mantle and conjecture the processes which formed it in the mantle differentiation.

In this paper, the starting material is a synthetic pyrolite glass doped with Nd, Sm, Hf, Lu, and U (0.3 wt.% for each). Four samples were made by compressing the pyrolite across a range of pressures from 46 GPa to 88 GPa, using a diamond anvil cell. The samples were molten by double-sided laser heating and then slowly cooled down below the solidus temperature before quenching. Thin sections for STEM analysis were made by the focused ion beam lift-out technique from the samples recovered after decompression. The 71 GPa sample is used as an exemplar to demonstrate the effectiveness of ML algorithms in the un-mixing of phases and improving their quantification.

Figure 1(a)-(d) presents a high-angle annular dark-field (HAADF) image and EDS elemental maps of the 71 GPa sample. Three phases are identified: ferropericlasite (Fp), bridgmanite (Brg), and Ca-rich perovskite (CaPv). All three phases are partially overlapping, with Brg being the dominant matrix phase, as shown in the spectra of the selected region of interests (ROIs) (i.e. Si exists in ROI_2 and Mg exists in ROI_3 in Figure 1(f), (g)). Brg can be quantified directly, though with an inadequate signal-to-noise ratio (SNR) for revealing trace elements, such as U, that may be incorporated in its structure, as shown in the inset of Figure 1(e). As for Fp and CaPv, they can only be quantified indirectly. The ROI mixtures are first quantified; the mixing level of Fp/CaPv and Brg is then calculated; lastly, Brg constituent is subtracted from the mixtures to obtain Fp and CaPv compositions. Similar to Brg quantification, each ROI has a limited SNR, while error propagation further increases the uncertainty of quantification.

If all three phases can instead be identified and segmented spatially, the signal of each summed spectrum would be increased to the maximum extent, thus greatly increasing the SNR. Here, we use non-negative matrix factorization (NMF) [2] and penalized clustering [3] cooperatively to achieve this task. No matter how small the phase area is, or how complicated the phase morphology, the combined algorithms can easily and precisely manage the segmentation. From this, the pure spectra of Fp, CaPv, and Brg are obtained and presented in Figure 2(d)-(f), ready for a direct quantification. It is noted that U indeed exists in Brg under the synthesized condition, as displayed in the inset of Figure 2(d). Although not presented here, the SNR for the EDS peaks of the other trace elements, Nd, Sm, Lu, and Hf are also considerably increased. This allows us to analyze the partition behavior of the elements, including that of trace amounts, between phases in a fast and reliable approach, from simple 2D STEM-EDS data. Additionally, the abundance map of each phase (i.e. Figure 2(a)-(c)) is computed by a fully constrained least squares linear spectral mixture analysis [4]. We believe that, when applied appropriately, the proposed workflow of algorithms can further be extended to the quantification of other spectroscopic data, such as from electron energy-loss spectrometry.

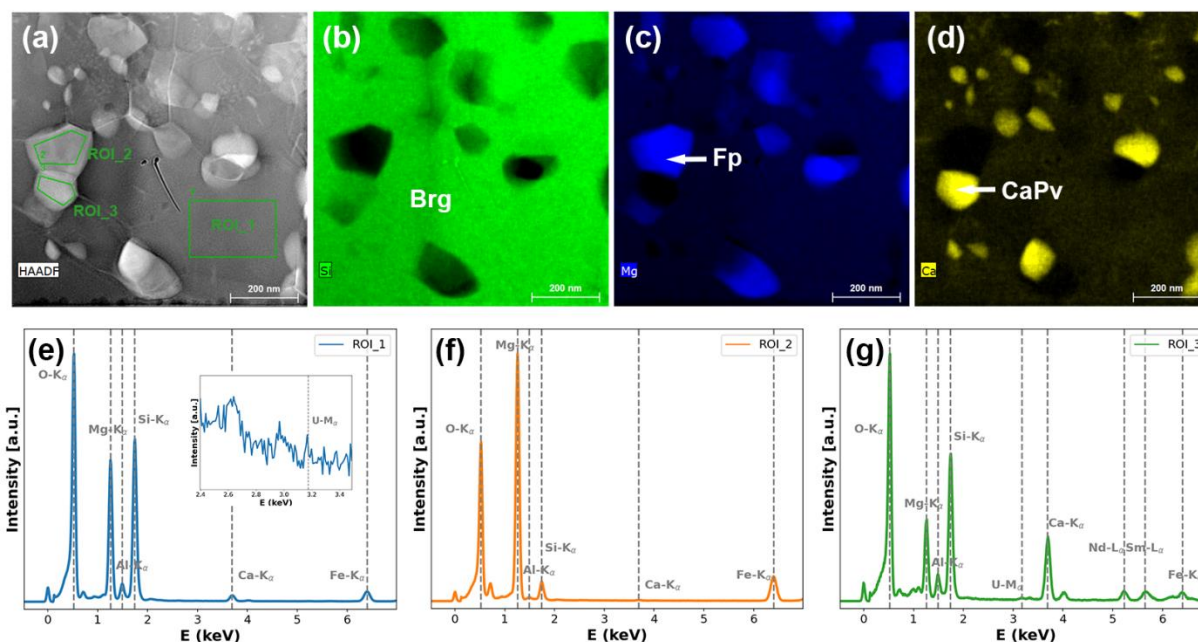


Figure 1. (a) HAADF image and (b)-(d) Si K α , Mg K α , and Ca K α elemental maps of the 71 GPa sample; (e)-(g) spectra of the ROI_1, ROI_2, and ROI_3, respectively.

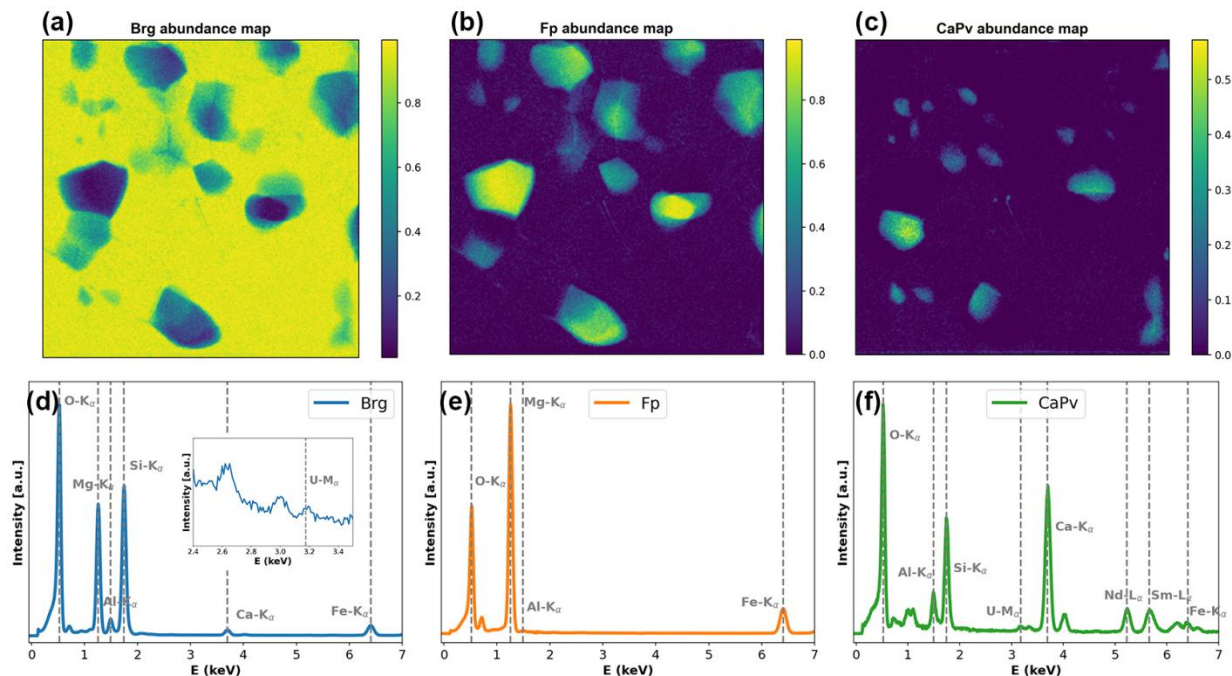


Figure 2. (a)-(c) phase abundance maps of Brg, Fp, and CaPv; (d)-(f) integrated spectra of Brg, Fp and CaPv.

References:

- [1] MI Jordan and TM Mitchell, *Science*, **349** (2015), p. 255. doi:10.1126/science.aaa8415
- [2] DD Lee and HS Seung, *Nature*, **401** (1999), p. 788. doi:10.1038/44565
- [3] Y Boykov, O Veksler and R Zabih, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23** (2001), p. 1222. doi:10.1109/34.969114
- [4] DC Heinz and Chein-I-Chang, *IEEE Transactions on Geoscience and Remote Sensing*, **39** (2001), p. 529. doi:10.1109/36.911111