Supplemental Materials

Autonomous Identification of Regions of Interest (ROI) in the Sample

MapX-II, a lower resolution prototype that operates in air, was designed to evaluate various operating modes and data reduction algorithms. The identification of regions of interest (ROI) in analyzed samples can be achieved by producing false color images displaying correlations of different elements. This process requires access to the full hdf5 file produced by the instrument, a data packet that can be hundreds of megabytes in size. Figure S1a shows an illustration of a MapX-II hdf5 data cube from the ultramafic xenolith AMASE08 UI-3 in which X-ray energy is displayed on the Z axis and pixel position is shown on the X and Y axes. The full XRF signal from the data cube is rendered in colors representing X-ray energy along the z axis (e.g., Ca is shown in green, Cr is shown in blue, and Fe is shown in red). Each vertically integrated pixel in the array contains a full XRF spectrum for that X, Y position; however, in general, single pixels do not have sufficient counting statistics to be quantified (Fig. S1b). Summed spectra from ROI containing many pixels are required for full quantification (Fig. S1c) as is the case for ROI that constitute significant fractions of an imaged area.

For flight applications, ROI selection must be performed on the spacecraft to minimize the size of the downlinked data products. Identifying patterns in images is a standing challenge in data science, and multiple open-source packages are now available to address this issue. We employed the clustering algorithms in Scikit-Learn (https://scikit-learn.org/stable/) to find clusters of pixels with similar compositions in UI-3. UI-3 contains three major minerals (clinopyroxene, orthopyroxene and chrome spinel) plus basalt (Fig. S2). We used an agglomerative clustering algorithm to identify 5 clusters in the 3-dimensional data set comprised of Fe, Ca, and Cr abundance. Prior to clustering, we perform principal component analysis and then cluster the points within principal component space. Spatial location was not taken into consideration. Applying the resulting labels back to the two-dimensional dataset, the regions identified by eye in the false color image (Fig. S2-c) are shown to have been correctly identified by the algorithm (Fig. S2-g). For visualization purposes only three elements were employed here, but the algorithm can accept any number of elements to perform the clustering analysis. This method has proven robust for selecting ROI in a wide variety of samples. Up to 20 elements have been included in the clustering analysis with only minor increases in processing time. At the present time, as implemented the algorithm still requires human input. The number of clusters needs to be pre-selected by the user as well as which elements to include. Including too many elements with near zero abundance hinders the ability of the algorithm to correctly identify ROI, although this is partially mitigated by only including principal components comprising 95% of the variance. Work to automate the selection of these two parameters is ongoing.

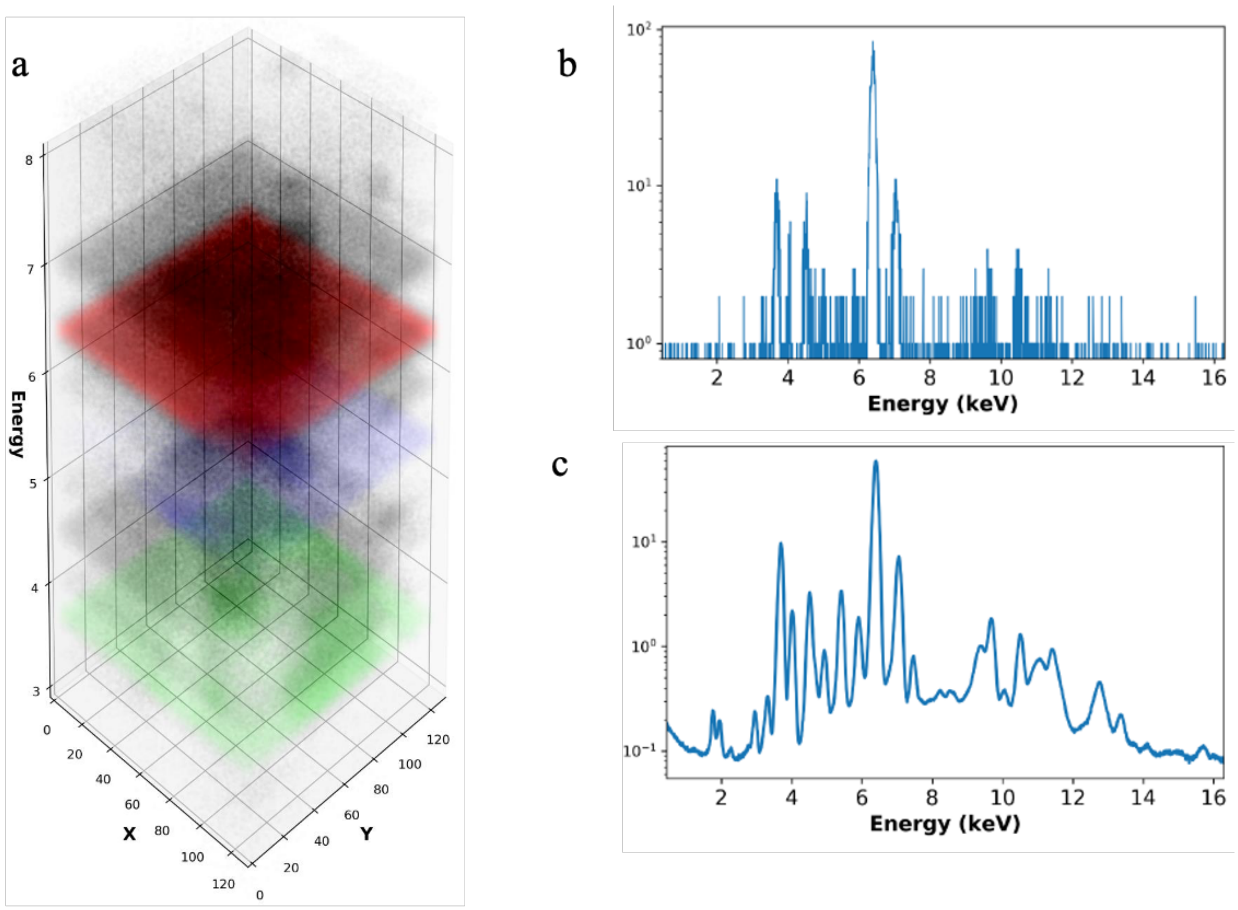

Figure S1: a). Illustration of an hdf5 data cube. X-ray photon energy is shown on the Z axis, X and Y pixels of the array shown on the X and Y axes. Selected elements shown in color (Ca @ 3.68 keV shown in green, Cr @ 5.37 keV shown in blue, Fe @ 6.4 keV shown in red). b). XRF energy histogram from a single X, Y pixel. Vertical axis shown in # of photons; counting statistics not suitable for quantification. c). Summed energy histogram for the entire image. Vertical axis now shown in relative counts, statistics sufficient for quantification (as is the case for ROI that constitute significant fractions of an imaged area).

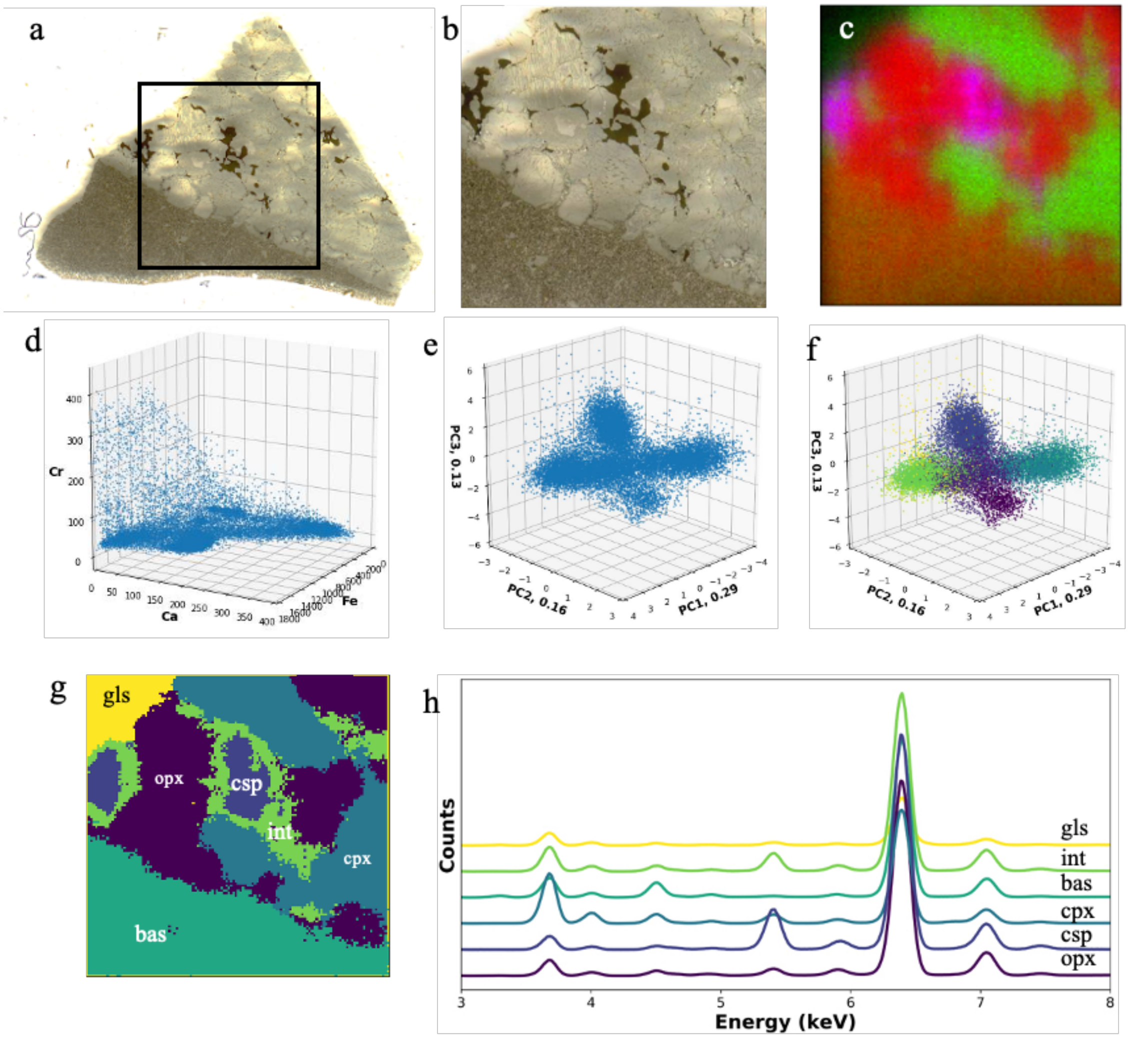

Figure S2: Illustration of Region of Interest (ROI) determination using unsupervised machine learning. a). Optical image of petrologic thin section from ultramafic xenolith AMASE08 UI-3. b). Closeup of area of thin section imaged by MapX-II. c). RGB element map from MapX-II, Red=Fe, Green=Ca, Blue=Cr (~250 µm resolution). d). Example 3D plot of Fe, Ca, and Cr pixel intensity. e). 3D plot of first three principal components. f). Clusters found by a hierarchical clustering algorithm. g). Element correlation clusters mapped into ROI on the sample, light blue ROI = clinopyroxene "cpx," $(CaFe)Si_2O_6$; dark blue ROI = orthopyroxene "opx," $(MgFe)Si_2O_6$; medium blue = chrome spinel "csp" $(FeCr_2O_4)$; medium green = basalt "bas," yellow = glass slide, light green areas are interfacial regions. h). XRF spectra of individual ROI (from g).