Data Exploration, End-member Identification, and Spectral Unmixing with the Python Hyperspectral Analysis Tool (PyHAT).
I. Aneece[1] and R. Anderson[2], [1]Western Geographic Science Center, USGS, Flagstaff, AZ (ianeece@usgs.gov). [2]Astrogeology Science Center, USGS, Flagstaff, AZ.

Introduction: Planetary science is transitioning from a data-poor to a data-rich regime, and this includes larger, higher dimensionality spectroscopic data sets. Although these data are scientifically valuable, they can be challenging to analyze. "Big data" analysis tools exist in terrestrial sciences and machine learning but implementation can be challenging and time-consuming, particularly for scientists who are not expert programmers. We have begun an effort to build upon the existing free and open source Python Hyperspectral Analysis Tool (PyHAT) library [1] to make a variety of well-known, cutting-edge spectral analysis algorithms available and easy to use for the planetary science community.

The proposed algorithms fall into two main groups: dimensionality reduction, and algorithms for end-member identification and spectral unmixing. Both are of high interest within the planetary science community. They can be applied to planetary (e.g. CRISM, M^3, ChemCam, TES, Cassini VIMS, Ralph on New Horizons, upcoming MISE) and terrestrial spectroscopic data to support planetary analog work.

A single centralized repository for the wide variety of spectral processing and analysis algorithms relevant to planetary science will be a useful resource for the entire community. Furthermore, implementing many of the algorithms in the existing GUI increases access to these methods.

PyHAT: The Python Hyperspectral Analysis Tool (PyHAT) library is a free, open source Python library for the analysis of planetary spectral data. It has been developed by the USGS Astrogeology Science Center under two different PDART grants, one focused on orbital hyperspectral data (e.g. CRISM, M^3, etc.), the other on point spectra (e.g. ChemCam, SuperCam).

The point spectra work has focused on preprocessing and multivariate regression methods, with a particular emphasis on Laser Induced Breakdown Spectroscopy (LIBS) instruments such as ChemCam and SuperCam. Once the data have been preprocessed, they can be clustered or used to generate a regression model to relate spectral signals to continuous physical quantities such as chemical composition [2].

The PyHAT library can be imported in Python and used in developing custom scripts. It can also be accessed through a PyQT5-based GUI [1] that allows users to call PyHAT functions in any order in a custom "workflow" that can be saved and restored. PyHAT is intended to serve as a centralized resource for planetary spectroscopists, enabling them to work with various data types and methods to analyze spectra.

Task Summary: We have received new funding to continue the development of PyHAT. That development is divided into three tasks described below (Fig. 1). The first two involve implementing new algorithms within the existing PyHAT Python library and testing to ensure compatibility so that they can be used seamlessly in a planetary data analysis pipeline. The third task will then incorporate those new capabilities into the GUI for use by non-programmers.

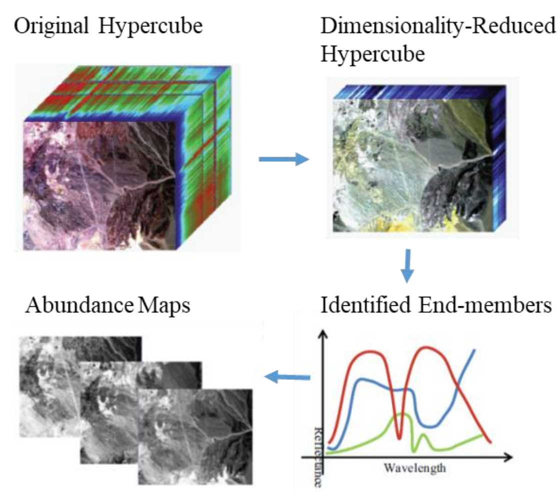

Figure 1: We propose to add well-known and state-of-the-art dimensionality reduction and end-member identification/unmixing algorithms into PyHAT, making them accessible in one package and available through a GUI to facilitate planetary spectroscopic data analysis. Figure modified from [3].

Task 1: Spectroscopic data sets typically suffer from data redundancy and the "curse of high data dimensionality" [4]. Dimensionality reduction can help avoid these issues, decrease computational and storage costs, increase accuracy, and aid spectral unmixing, end-member identification, target detection, and visualization [4,5]. We will be adding linear discriminant analysis (LDA, Fig. 2), the local fisher discriminant analysis (LFDA), Non-Negative Matrix Factorization (NNMF), and Minimum Noise Fraction (MNF) to PyHAT.

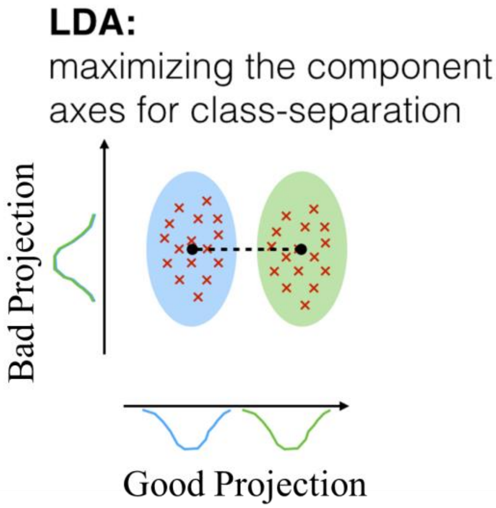

Figure 2: LDA projects the data to maximize distance between classes. Figure modified from [6].

Task 2: Dimensionality reduction can aid end-member identification and spectral unmixing [4]. In end-member identification, data are projected so that the most spectrally extreme and thus spectrally pure pixels are detected and identified as end-members. We will be adding the pixel purity index (PPI, Fig. 3), N-Finder (N-FINDR), automatic target generation process (ATGP), fast iterative pixel purity index (FIPPI), and sequential maximum angle convex cone (SMACC) for end-member identification in PyHAT.

The end-members can then be used in unmixing classification methods that estimate how much of a particular pixel corresponds to each end-member. Linear least squares (LLS), extended linear mixing model (ELMM), multiple end-member spectral mixture analysis (MESMA), general bilinear model (GBM), and multilinear mixing model (MLMM) will be added to PyHAT for spectral unmixing.

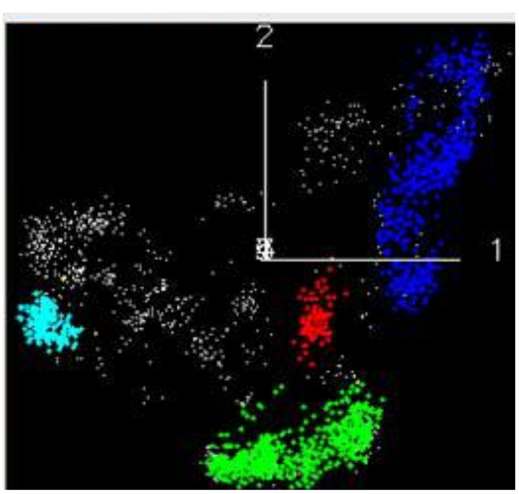

Figure 3: Example of PPI, modified from [7], showing end-members identified at the extremities of the cloud of data points.

Task 3: We will also add access to the algorithms from Tasks 1 and 2 to the PyHAT GUI (Fig. 4), which is designed around the concept of "workflows": series of discrete data processing steps each of which is referred to as a "module." The order of the modules is defined by the user. Modules can be inserted or deleted from within a workflow to efficiently arrive at an optimal workflow, which can then be saved to and restored from human-readable json formatted files.

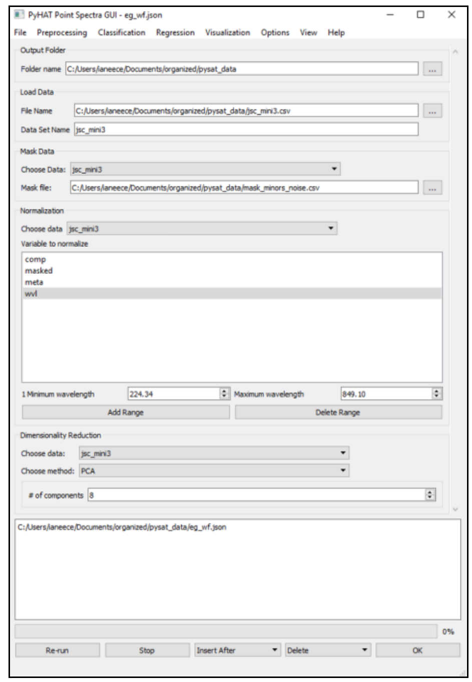

Figure 4: Example PyHAT GUI workflow.

This task includes development of documentation, tutorials, and examples to help users understand how to use the GUI. We will publish an overview of PyHAT and its improvements in the open-access journal SoftwareX, and final code will be made available on NASA's GitHub repository for PyHAT, as well as the current USGS GitHub repositories.

References: [1] https://github.com/USGS-Astrogeology/PyHAT_Point_Spectra_GUI. [2] Anderson, R.B., et al. 2017. Spectrochimica Acta Part B: Atomic Spectroscopy, 129, 1, 49-57. [3] Bioucas-Dias, J., et al. 2012. IEEE J Sel Topics in Appl Earth Obs and Rem Sens, 5(2): 354-379. [4] Ghamisi, P., et al. 2017. IEEE Geosci and Rem Sens Mag, DOI: 10.1109/MGRS.2017.2762087. [5] Ye, Z., et al. 2017. Europe J Rem Sens, 50(1): 166-178. [6] Raschka, S. https://sebastianraschka.com/. [7] Mukhopadhaya, S. 2016. JBAER, 3, 10, 831-837.