PREDICTING A GYPSUM-DOLOMITE MIXTURE SPECTRUM WITH A SPECTRAL MIXTURE MODEL BASED ON PRINCIPLE COMPONENT ANALYSIS. Joseph S. Makarewicz and Heather D. Makarewicz, Olivet Nazarene University, Bourbonnais, IL.

Introduction: The OMEGA and CRISM visible/near-infared (VNIR) imagers continue to return hyperspectral images of the surface of Mars. Determining the abundance of minerals in the images helps planetary scientists to understand the geological processes and surface chemistry occurring on Mars. We are developing a method to use principle component analysis (PCA) to determine the abundances of alteration minerals including sulfates and carbonates. The current work is focused on laboratory spectra.

PCA-based Spectral Mixture Model: A spectral mixture model based on principle component analysis has been applied to several laboratory datasets. The algorithm was applied to a pyroxene mixture dataset and was used to predict precent composition and grain size [1]. It was applied to two phyllosilicate-containing mixture datasets to understand how percent composition affects spectral features [2]. Percent composition was predicted for tertiary and quaternary mixture datasets containing silicates [3]. It was also shown that the algorithm can be used to virtually mix endmembers within a model [4]. This study will generate a gypsum-dolomite mixture spectrum using a PCA-based spectral mixture model.

Methods: *Samples and Spectra.* A gypsum-dolomite dataset was selected for this study. The mixture samples were prepared by weight percent. The spectra were measured in another mixture study and reposed in the RELAB spectral database [5].

Pre-processing. The spectra were pre-processed in a similar manner as previously published studies [1-4]. The spectra were cropped to include 0.4 to 2.6 µm wavelengths and were resampled to 0.002 µm increments. Each spectrum was normalized such that the area under the curve was set to 500. The spectra mixture dataset after these pre-processing steps are shown in Fig. 1.

PCA procedure. Principle component analysis was performed using singular value decomposition (SVD). Before performing PCA, the mean normalized spectrum for the dataset was subtracted from each normalized spectrum. The mean normalized spectrum along with the most significant principle component vector is shown in Fig. 2. PCA also produced 7 principle component values corresponding to each mixture spectrum shown in Fig. 1. Each spectrum can be approximated as the mean normalized spectrum plus the first principle component vector multiplied by the first principle component value.

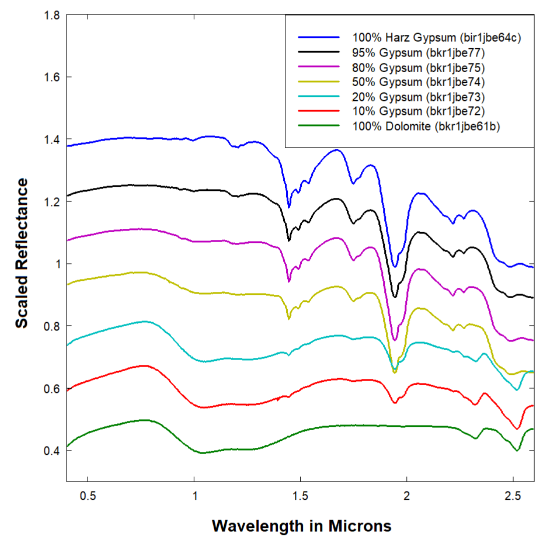

Fig. 1: Relab Harz Gypsum and Selasvann Dolomite mixture dataset after processing with offsets.

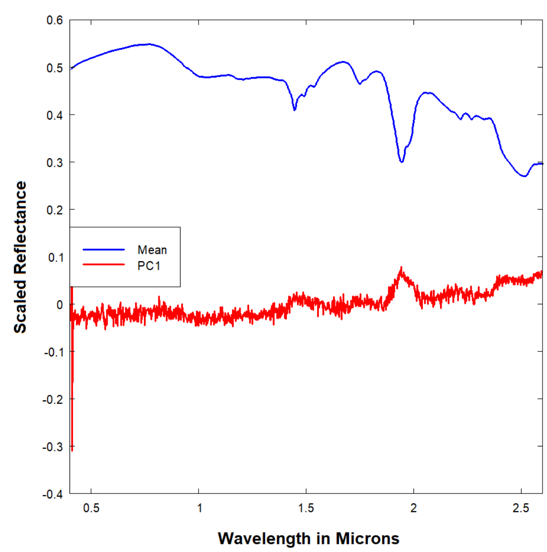

Fig. 2: Gypsum-Dolomite mean normalized spectrum and first principle component vector generated by PCA.

Principle Component Correlation. The first principle component values were correlated with percent gypsum. The relationship between percent gypsum and the PC1 values was quantified. A scatter plot, along with a linear regression, are shown in Fig 3. The linear relationship for the gypsum-dolomite scatterplot, shown in Fig. 3, can be approximated by the following equation.

$$\% \text{ GYP} = 50.7 - 21.9 \times \text{PC1}$$

Results: With this mixture model, the spectra for other precent compositions can be generated. For this study, a 90% Gypsum and 10% Dolomite mixture was generated. First, the PC1 value for 90% Gypsum was computed to be -1.7881 using the equation for the regression line above, which is plotted in Fig. 3. Then, the principle component vector, in Fig. 2, is multiplied by this principle component value and added onto the mean vector, also in Fig. 2. In Fig. 4, the generated 10% Quartz spectrum is compared against a measured 10% Quartz spectrum that was omitted from the dataset during PCA. There are some minor differences between the spectra in the 2.4 to 2.6 μm wavelengths. The two spectra appear very similar. The error has an RMS value of 0.0130 and the MAPE is 0.0388.

Discussion: The technique from this study can be used with all PCA-based spectral mixture models. Mixture models with multiple endmember can be created for specific virtual mixing applications. There are two noticeable issues with the generated spectra. The spectra have a lot of high frequency noise that needs to be characterized and filtered. In this study and the last study [4], there has been a single large artifact in the lower wavelengths. Both of these issues are solvable with a median filter.

For this study, the percent composition and the principle component have a linear relationship. In most of the studies [1-3], there has been a linear relationship with the exception of the last Quartz-Gypsum study [4]. In this study, a linear regression was used to correlate the precent composition of gypsum with the PC1 value. Linear interpolation could have also been used with similar results, but is most useful for non-linear relationships.

Future Work: Studies in the near future will continue to model sets of lab mixtures in order to understand the correlation between principle components and dataset properties. Future studies will involve mineral mapping of hyperspectral images. A web-based tool for mixture modeling will also be developed.

Summary: A PCA-based mineral mixture model was applied to a gypsum-dolomite mixture dataset. A linear relationship between percent composition and the first principle component was quantified. A 90% Gypsum and 10% Dolomite mixture spectrum was generated. The generated spectrum is very similar to the measured spectrum with a MAPE of 0.0388.

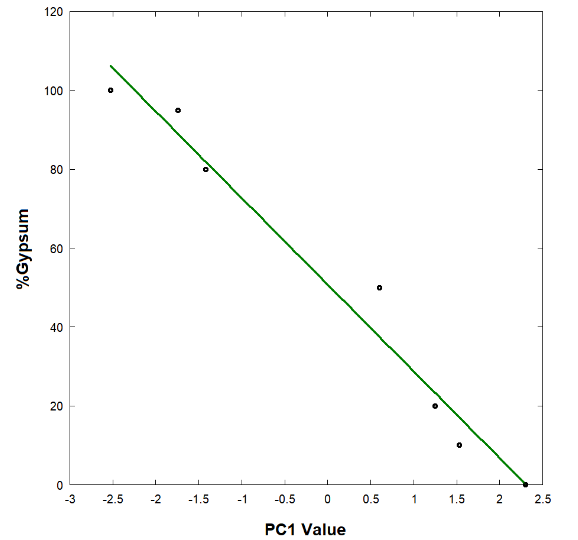

Fig. 3: A scatter plot and linear regression correlating percent gypsum composition and the first principle component value.

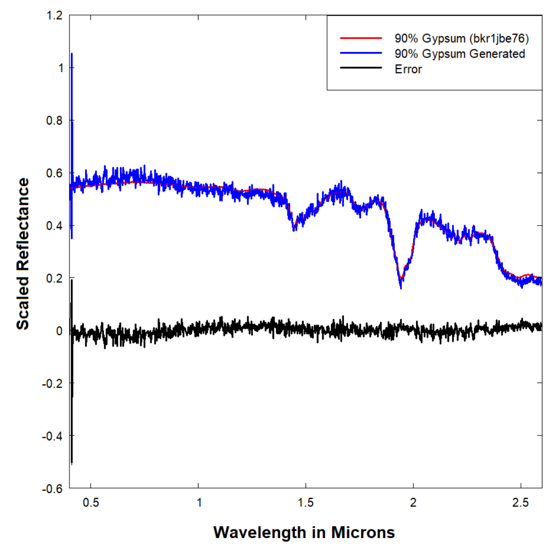

Fig. 4: A 90% Gypsum spectrum not included in the model training set and a 90% Gypsum spectrum generated with the PCA.

References: [1] J.S. Makarewicz and H.D. Makarewicz (2013) *IEEE Whispers* doi: 10.1109/WHISPERS.2013.8080604. [2] J.S. Makarewicz, H.D. Makarewicz, and J.L. Bishop (2018) LPSC, Abstract #1378, Poster #503. [3] D.J. Burnett and J.S. Makarewicz (2018) AGU Fall Meeting, Abstract #P41D-3767. [4] J.S. Makarewicz and H.D. Makarewicz (2018) AGU Fall Meeting, Abstract #IN11E-0665. [5] S.J. King, J.L. Bishop et al. (2014) LPSC, Abstract #2284, Poster #417.