BREWING MACHINE LEARNING MODELS FOR ACCURATE ANALYSIS OF CURRENT AND FUTURE MISSION DATASETS. Nandita Kumari[1], Timothy D. Glotch[1], Benjamin T. Greenhagen[2]. (Nandita.kumari@stonybrook.edu) [1]Department of Geosciences, Stony Brook University, NY, [2]Johns Hopkins Applied Physics Laboratory, MD.

Introduction: Spectroscopy has been widely used to study the composition of planetary and interstellar bodies [1,2]. Spectroscopic techniques have helped us identify resources such as water [3,4] on the lunar surface as well. As we enter the Artemis era of lunar exploration, the data volume returned to the scientific community will be increasing multifold. As we gear towards the scientific exploration and in-situ resource identification on extra-terrestrial bodies, including the Moon, advanced methods for data archiving, processing, and analysis will be required. Here we have used a machine learning algorithm to estimate the Christiansen feature (CF) position from Channel 3, 4 and 5 (corresponding to 7.55 µm - 8.05 µm, 8.10 µm – 8.40 µm, and 8.38 µm - 8.68 µm, respectively) of the Diviner Lunar Radiometer Experiment [5]. The CF is an emissivity maximum or reflectance minimum that occurs where the real index of refraction passes unity and the imaginary index (extinction coefficient) approaches zero [6]. The position of the CF is indicative of silicate polymerization in a mineral and occurs at a shorter wavelength for silica-rich framework and chained-silicates and at longer wavelengths for silica poor minerals.

Background: The three mineralogy channels of Diviner have been used to understand the bulk silicate composition of the Moon in detail over the last decade. The calculated emissivities of these three bands are typically corrected for photometric effects, sun distance and solar incidence angle [7]. To calculate the CF position, the maximum of a three-point parabolic fit to the brightness temperatures or emissivities from these three channels is used. However, for highly silicic and ultramafic minerals, there is an ad-hoc cut-off at 6.9 µm and 9.6 µm, where the CF occurs well outside the wavelengths measured by Diviner. The parabolic fit method has proven to work well for regolith compositions with CF positions near 8 µm, but, as we move toward high or low-silica compositions dominated by minerals such as quartz or ilmenite/spinel, respectively, the difference between the estimated and real CF positions can be substantial. In this work, we have developed a machine learning algorithm to calculate CF directly from Diviner's 3 mineralogy spectral bands. The algorithm trains over a spectral library of different lunar mineral mixtures convolved to the band passes of different missions.

Dataset and Methods: The spectral library used in this study was collected in simulated lunar environment chambers at Stony Brook University [8] for different grain sizes of pure minerals and the Applied Physics Laboratory [9] for a variety of mineral mixtures, temperatures, and time. We have used a total of 539 spectra convolved to the three channels of Diviner. The CF of each full laboratory spectrum was estimated individually using a six-degree polynomial fit.

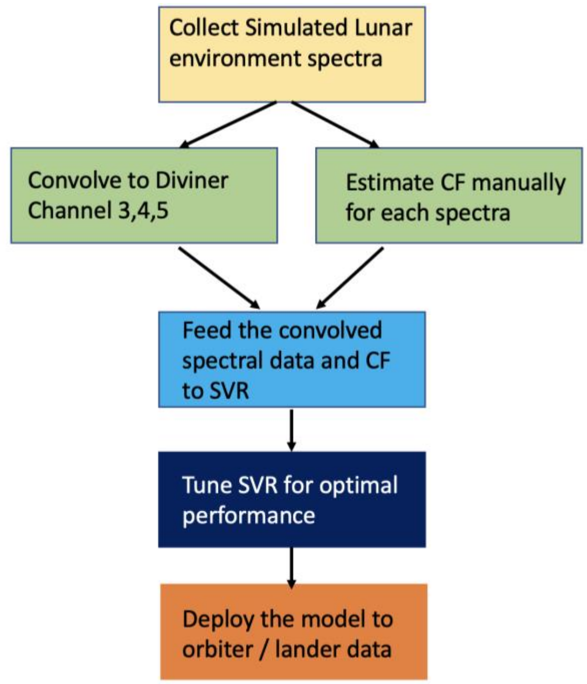

Figure 1. Flowchart of the machine learning model used to estimate CF wavelength position.

We have used a Support Vector Regression (SVR) algorithm trained over 70% of the spectra and then validated on the remaining 30% of the spectra. The pipeline uses cross-validated testing for optimization. SVR is a linear regression method that uses a hyperplane for fitting the data. The trained model is then deployed on the Diviner emissivity data for CF estimation (Figure 1). We also applied this algorithm on published bandpasses for the L-CIRiS [10], L-VISE [11] and Lunar Trailblazer thermal mapper [12] instruments to judge its performance and prepare for analyses of future datasets. We also performed the same for chained silicates via a cutoff for CF below 7.8 µm.

Results: The r^2 score of the validation set is 0.91 with an error in 0.14 µm in the predicted measurements. We further deployed this model on the Apollo 16 site for ground truth verification. We initially deployed the model on the standard emissivity value for

CF prediction for three different time-of-day ranges. The standard emissivity values without time-of-day correction display that the calculated composition varies from anorthositic to basaltic in line with previous studies [13] (Figure 3a).

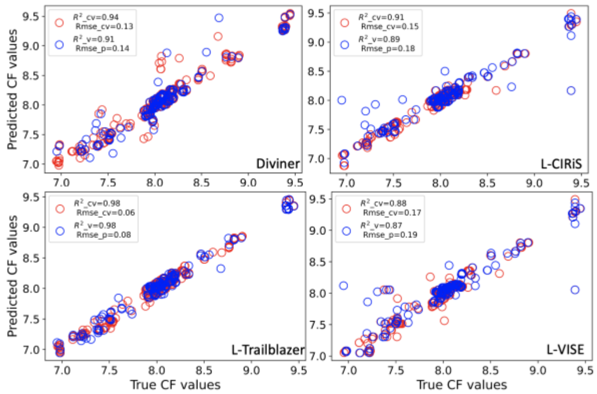

Figure 2. True versus predicted CF for a) *Diviner b) L -CIRiS c) Lunar Trailblazer and d) Lunar VISE*

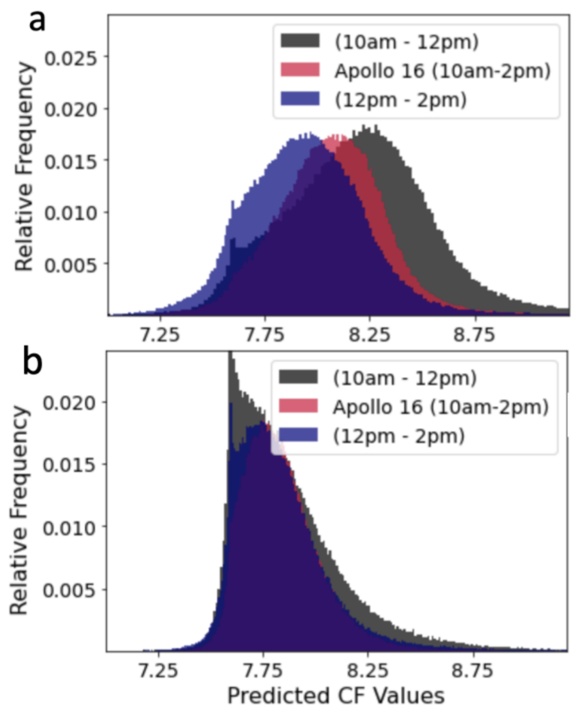

Figure 3. a) *CF estimated using standard emissivity at different time of day for same pixels. b) CF estimated using ML model for corrected emissivity at different time of day.*

We then also applied the ML model on the corrected CF values. The corrected CF values estimated through the ML model have a wider distribution with more distinguishable values compared to the parabolic model fit, which is concentrated around 8.15 µm. Some difference between the models also occurs as we move away from 8 µm in either direction, and the difference between the true and estimated CF increases due to the nature of parabolic fit. Furthermore, with the time-of-day correction, we observe that there is an overall shift in the CF values towards shorter wavelengths with narrower distribution of the values (Figure 3b). We also observe that after the correction, the estimated CF values overlap indicating the model predicts similar results as the previous parabolic model. Our corrected CF predictions are also in agreement with previously estimated values of Apollo 16 samples through lab spectroscopy by [13].

Future Missions: The trained model displays the lowest root mean squared error value (Figure 4a) and r-squared value (Figure 4b) for Lunar Trailblazer bands followed by the Diviner, L-CIRiS and L-VISE bands respectively. This is expected since Lunar Trailblazer has ten bands as opposed to only three bands for the rest of the instruments.

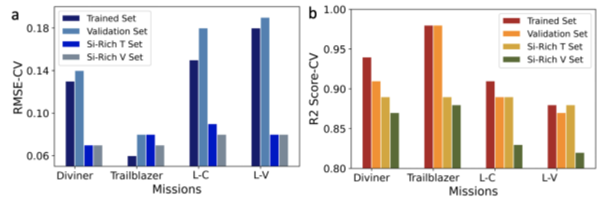

Figure 4. Model performance for each dataset using root mean squared error(a) and r-squared (b) using predicted and true CF values.

Future Work: In future, we plan to add ilmenite and basalt mixtures to the model so that it can be reliably deployed on the mare basalts.

References : [1] Tanaka et al., (1989), ApJ, 336,207-211 [2] Logan et al., (1973), JGR,78, 1896-1977 [3] Lucey et al., (2006), Rev. Miner. Geochem. 60, 83–219 [4] Li et al., (2018) *PNAS, 115(36)* 211 [5] Paige et al. (2009), Space Sci. Rev., 150, 125-160 [6] Conel, J.E. (1969), JGR, 74(6),1614-1634 [7] Greenhagen et al., (2010), Science, 329, 5998 [8] Shirley and Glotch (2019), JGR, 124(4),970-988 [9] Greenhagen et al., (2020), LPSC #2171 [10] Hayne et al., (2020), LPSC, 2707 [11] Hanna et al., (2022), LEAG Meeting [12] Ehlmann et al., (2022), LPSC, 2316 [13] Salisbury et al., (1973), Geo. et Cosmo. Acta, 3, 3191-3196