PREDICTING THE SEAWATER CHEMISTRY OF AN OCEAN WORLD USING MACHINE LEARNING ON ISOTOPIC MEASUREMENTS OF VOLATILE CO_2. B. P. Theiling[1], V. Da Poian[1,2], B. Powell[1] [1]NASA Goddard Space Flight Center, [2]Southeastern Universities Research Association.

Introduction: Given the long time intervals required for data transmission to and from ocean worlds targets, low bandwidth for data transmission, time required for data processing and analysis, and potentially extreme radiation environments (*e.g.*, Europa), it is clear that ocean worlds missions will need more autonomous flight instruments and software in order to achieve established science goals. Protracted time intervals for data analysis (*e.g.*, Europa Lander) strongly motivates the development of rapid, consistent and streamlined methods for interpreting data from flight mass spectrometers to *e.g.*, determine how mass spectra from a plume or surface liquid/ice relates to the surface/subsurface. Since mass spectrometry also has the potential to correctly identify biosignatures[1], it is imperative that such methods for interpreting data are consistent and accurate.

We used 848 isotope ratio mass spectra from laboratory analyses of CO_2 that interacted with ocean worlds-relevant seawaters as a 'training' dataset for 'unsupervised' machine learning. In unsupervised learning, characteristics of the data are not labeled or linked, and any similarities found only result from the neural network. CO_2 isotopologues analyzed for this dataset mimic the remote measurements of CO_2 by a flight mass spectrometer, and are detailed in Theiling [2]. From this dataset, we used measured features of the spectra, such as retention time, intensity, and (isotopologue) mass ratios as inputs for our autoencoder neural network. Our neural network was trained to find similarities in these and other spectral features for seawaters of a particular composition and amount of initial CO_2. Successful training then created an output of these similarities for various seawaters, which included $MgSO_4$, Na_2SO_4, NaCl, $MgCl_2$, KCl, and $NaHCO_3$, and combinations of these salts. We then applied dimensionality reduction techniques such as Principal Component Analysis (PCA), T-Distributed Stochastic Neighbor Embedding (TSNE), and Uniform Manifold Approximation and Projection (UMAP) to demonstrate latent data features as a two-dimensional projection in a unitless, high-dimensional space. In this projection, a data point represents the combined effect of spectral features such as intensity, retention time, and isotope ratio. Our initial UMAP demonstrates data clustering (organization of the data by the neural network) based on the amount of CO_2 that had initially interacted with each seawater. Further training using more 'supervised' learning techniques demonstrate strong clustering of preliminary data based on initial CO_2 concentration, seawater chemical composition, and ionic strength (salinity). Our preliminary work therefore suggests that machine learning has the potential to identify compositional variants of an ocean world seawater based on mass spectra from volatile CO_2 measurements.

Acknowledgments: This work was funded through a Strategic Task Group at NASA Goddard Space Flight Center. The training dataset was collected through funding from the Oklahoma Space Grant Consortium.

References: [1] Pappalardo, R. et al. (2013) *Astrobiology*, 13, 740–773. [2] Theiling (2020) *Icarus*, 114216.