Abstract for NASA Technology Showcase
https://nasa-techshowcase.seti.org

Victoria Da Poian, Eric Lyness, NASA Goddard Space Flight Center

Machine learning and data science for the future of planetary science missions exploring our outer solar system

Many upcoming and proposed missions to ocean worlds such as Europa, Enceladus, and Titan aim to evaluate their habitability and the existence of potential life on these moons. These missions will suffer from communication challenges and technology limitations as data volumes of instrumentation are growing at an exponential rate. Planetary instruments will soon generate orders of magnitude more data than can be returned to Earth. While a new communications infrastructure would help, it will be prohibitively expensive and will require years of development before being deployed for missions operations. A much cheaper solution is the development of smarter instruments or the concept we envision of *Science Autonomy*.

We investigate the applicability of data science and machine learning (ML) techniques on mass spectrometry (MS) data from laboratory analogs as a case study for development of new strategies for planetary science missions. Our driving science goal is to determine whether the mass spectra could contain information about the chemical composition of the sample and potential biosignatures. These developed data science and ML techniques aim at analyzing what inherent information the spectra contain and determining whether a data science pipeline could be designed to in situ analyze data from future planetary science missions that will explore further away in the solar system. Future icy ocean world missions will be equipped with highly developed mass spectrometers enabling isotope ratio measurements of the volatiles evolving from the exosphere (a very tenuous atmosphere) and from the plumes. A clear understanding of isotopic fractionation will be essential to evaluate potential isotopic biosignatures for future ocean worlds missions. For instance, the MAss Spectrometer for Planetary EXospheres (MASPEX) onboard the Europa Clipper mission will be able to analyze the isotopic composition of volatiles (e.g., CH_4, H_2O, NH_4, CO_2, etc.) (Brockwell et al., 2016) during fly-bys of Europa. We not only have to assess the capabilities of mass spectrometer instruments to perform this task during fly-bys, but also evaluate the potential ability of onboard software in future missions (e.g., Enceladus Orbilander) to autonomously pre-process and identify signals of higher interest to enable data transmission prioritization.

We present our proof-of-concept system using the laboratory data from the MOMA (Mars Organic Molecule Analyzer) instrument, the mass spectrometer onboard the 2028 ExoMars mission. We developed an Earth-based technology that classifies mass spectra to determine the general category of a physical sample, allowing the scientists to see the raw science data along with the system's categorization. This supports scientists in their analysis during the ExoMars mission operations that will be time-limited. A longer-term goal is to deploy this tool onboard a spacecraft for automated real-time in situ data analysis and decision-making. We envision the concept of *science autonomy* (Da Poian et al., 2022) where data analysis processes such as the ones discussed in this research could be used onboard a spacecraft to prioritize data transmission for bandwidth-limited outer solar system missions and could greatly enhance the quality of the limited data returned to Earth.

One of the main limitations during the development of this research has been the constraint of the available data for training ML models. We explore several ways to support the design of future instruments for planetary missions to face this data volume limitation such as:

- Work with data scientists to determine appropriate data strategy plan and determine volume necessary to train an ML network to answer mission's science questions.
- Designing flight instruments as closely as possible to commercial instruments. This allows data to be generated on less expensive instruments.
- Developing transfer learning techniques that can apply what is learned on the commercial data to the flight instrument data
- For outer planets, plan to generate data during the long cruise then transmit a trained "object" to the spacecraft before operations.
- Specify computing and memory for instrument appropriate for executing network (note that that training the network will be done on Earth, so required processing power is much lower)

Mission operations can also be optimized to enable extra data processing. One mission concept under study focuses on the Dragonfly mission (targeting Titan in the 2030s) where we study the possibility of running neural networks (NN) during the Titan night when the instruments are in low-power mode.

The main objective of our research is to share a refined method and provide lessons learned and insights in the data collection and the development of data analysis tools focusing on mass spectrometry data for future planetary missions for the development of onboard autonomy.