In-situ Science Data Analysis and Curation for the Europa Lander Autonomous Mission Prototype.

Yuliya Marchetti, Caleb Wagner, Philip Twu, Marissa Cameron, Glenn Reeves, Grace Tan-Wang, Steve Chien, Rebecca Castano, Kiri Wagstaff
Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91108.
(yuliya.marchetti@jpl.nasa.gov)

Introduction: Future landed missions on ocean worlds pose great technological and engineering challenges, in part due to very short durations and extremely limited communication bandwidths. These new challenges drive the need for autonomous systems focused on increasing science yield in the presence of highly unknown environments [1]. We present a new science-driven data analysis framework to inform autonomy decisions for a Europa Lander mission concept.

The two main components of the onboard data analysis system are: 1) observational data science value (SV) estimation and 2) content-dependent data reduction. The former is a quantitative measure translating and compressing a specific science question into a single metric. The latter reduces the data beyond classical data compression to preserve the highest-priority science information, e.g. [2]. SV estimation feeds into both onboard decision-making for actions to take next as well as content-dependent data reduction.

Methods: Based on a notional Europa Lander payload [1], we have developed a set of synthetic instrument data, implemented generic science value metrics and data reduction methods, and integrated these into a fully autonomous prototype flight software system.

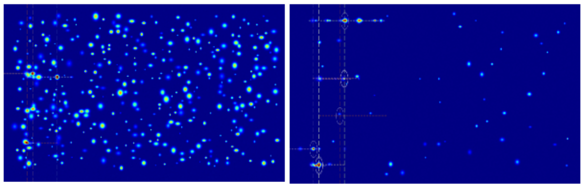
Synthetic data and instrument simulations. The generic science instruments include a mass spectrometer (GCMS), Raman spectrometer, microscope, seismometer, and science imager. Synthetic instruments emulate specified data volumes and potential data complexity and phenomena using a wide set of parameters.

Science value estimation. Onboard science data assessment is a combination of both various instrument-adapted signal processing and detection algorithms as well as statistical and information theoretic measures to "compress" the information about the detection as a normalized SV.

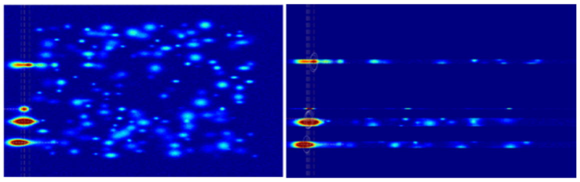
Content-dependent data reduction. We implement two increasing levels of data reduction: 1) subset data and 2) summarized data. The former extracts only the identified interesting data subsets, while the latter returns summary statistics. The choice of the reduction method depends on SV obtained from the original data. Each reduced data product has an associated *reduced* SV that reflects the loss of information, e.g. [3].

Results and Discussion: The data assessment methods and metrics are derived from and driven by instrument-specific science goals and questions.

GCMS example. The SV is represented by two science-defined metrics: 1) abundance, representing presence and prominence of peaks, i.e. simulated compounds, and 2) pattern similarity that compares a spectrum to a known library of *biosignatures* [1]. The former uses clustering entropy to represent how well the data is separated into convex groups. The latter is computed with the weighted cosine similarity. Figure 1 shows some simulated data scenarios: (a) many peaks corresponding to a high SV; (b) simulated biosignature (a row of peaks, top left), producing a high SV; (c) low data volume with a low SV; and (d) subset data based on detections in (c), resulting in ~92% size reduction.



(a) High abundance of compounds **(b)** Simulated biosignatures (left)



(c) Small data size **(d)** Subset data derived from (c)

Figure 1: Simulated GCMS examples with x- and y-axis representing the spectral (m/z) and time dimensions (secs), respectively.

Conclusions: Content-dependent data reduction has the potential to provide orders of magnitude reduction in downlink volumes while ensuring that the most critical data reaches the scientists. Future work will continue on integrating retrospective data assessment, novelty detection, prediction of SV for "unknown" data, and adaptive selection of reduction algorithms based on science goals and engineering constraints.

References:

[1] Hand, K. et al. (2022) *Science Goals and Mission Architecture of the Europa Lander Mission Concept.* Planet. J 3(1), 22; [2] Thompson, D. et al. (2009). *Adaptive Subsampling of Image Sequences for Remote Exploration*. ESA Spec. Pub., 673 (12); [3] Wagstaff, K. et al. (2004, March). *Science-based region-of-interest image compression*. In LPSC (p. 1934).

Acknowledgments: The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.