Onboard Autonomous Summarization and Prioritization of CE-ESI MS Data for the Ocean Worlds Life Surveyor

Jake Lee, Steffen Mauceri, Jack Lightholder, Mark Wronkiewicz, Gary Doran, Lukas Mandrake, Zuzana Cieslarova, Miranda Kok, Maria F. Mora, and Aaron Noell
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Introduction

The **Ocean Worlds Life Surveyor (OWLS)** is a field prototype instrument suite designed to autonomously search for evidence of water-based life, developed in preparation for potential future missions to ocean worlds such as Enceladus and Europa. One instrument included in this suite is a Capillary Electrophoresis-Electrospray Ionization Mass Spectrometer (CE-ESI MS), which can detect the presence of organic molecules and other potential biosignature compounds. Due to the extreme energy costs involved in communication from these distant worlds, a mission's downlink bandwidth is insufficient to return raw data from even a single sample.

We developed two onboard capabilities to address this constraint:

- **Compression via knowledge summarization**
- **Prioritization for the most scientifically useful observations.**

CE-ESI MS generates a two-dimensional array of ion counts resolved by their m/z and migration times. The most scientifically valuable information in this data are the locations and properties of ion peaks, which indicate the presence of specific compounds. **ACME performs knowledge summarization and compression by describing these peaks.**

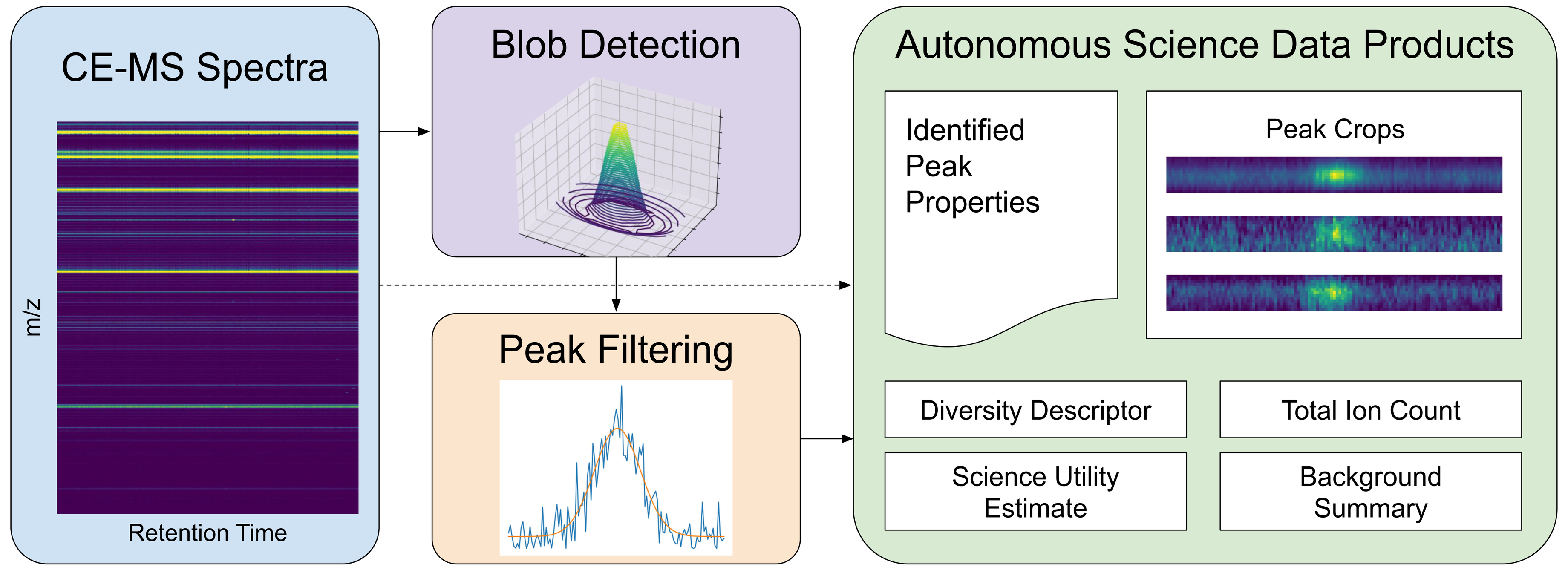


Figure 1. ACME system diagram.

TOTAL MISSION BUDGET	SINGLE SAMPLE RAW DATA SIZE	AUTONOMOUS SUMMARY SIZE
70MB	100MB	0.1MB

Methodology

ACME first preprocesses the raw data by removing background noise with a median filter and enhancing signals with a Difference of Gaussians filter. Then, ACME identifies potential peaks with blob detection, which results in tens of thousands of peak candidates, the majority of which is noise. To separate scientifically valuable peaks from noise artifacts, we calculate several peak characteristics for each candidate. These characteristics include peak location, height, width, signal-to-noise ratio, and the mean squared error of a Gaussian fit (shown in Figure 2 and Table 1).

Finally, we use a classifier that applies a series of thresholds to keep only target peaks that indicate the presence of a compound. These target peak candidates have autonomous science data products generated for downlink.

Each step of this method is designed to be transparent and configurable, even inflight during a mission, for adaptability to unforeseen scenarios (such as instrument failure or domain shift).

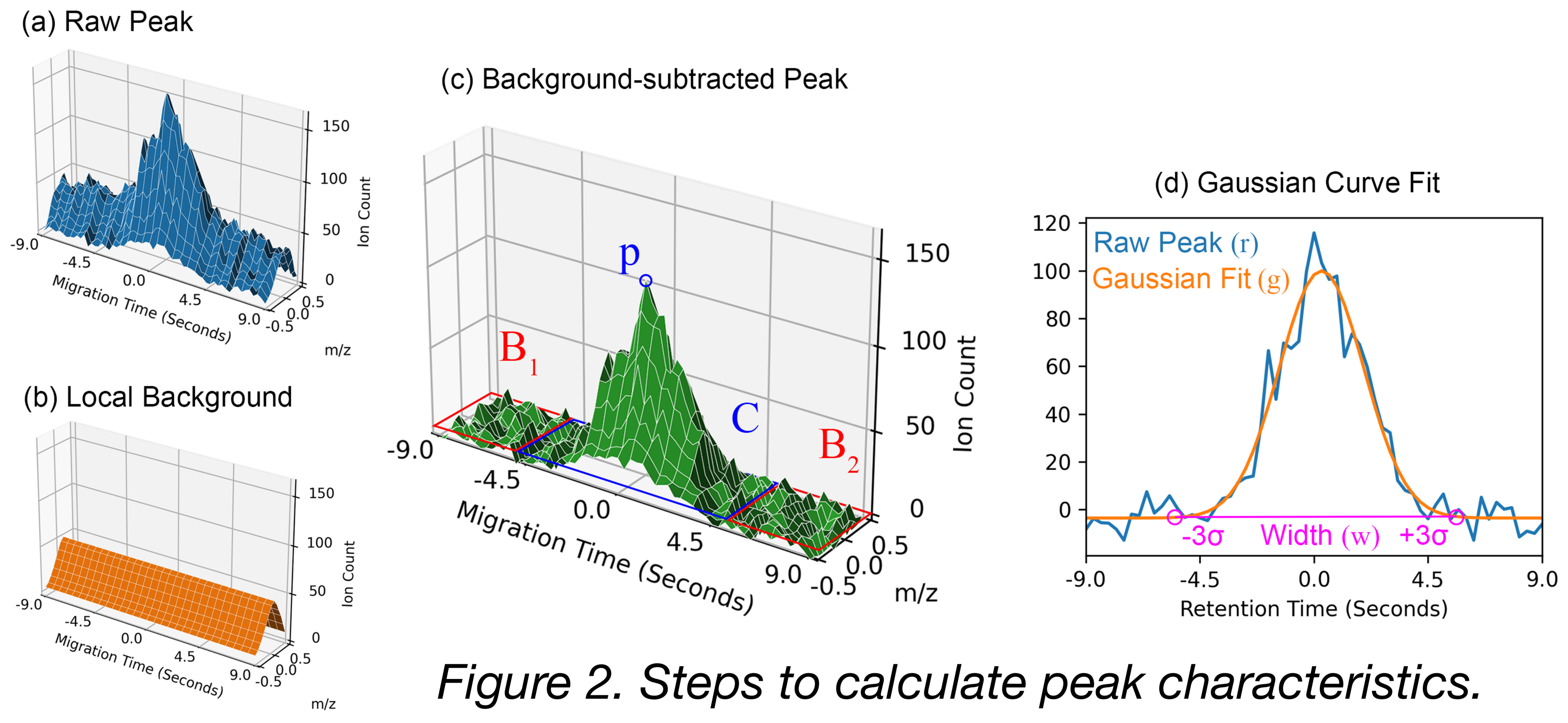


Figure 2. Steps to calculate peak characteristics.

Table 1. List of peak characteristics. Variables defined in Fig. 2.

Peak Property	Equation	Peak Property	Equation
Peak Height	$p = \max(C)$	Background Level	$\max(\hat{B}_1, \hat{B}_2)$
Peak Volume	$\int C$	Background Std.	$\max(\sigma_{B_1}, \sigma_{B_2})$
Peak Width	w	Background Diff.	$\text{abs}(\hat{B}_1 - \hat{B}_2)$
Peak Z-Score	$p / \max(\sigma_{B_1}, \sigma_{B_2})$	Background Ratio	$\min(\hat{B}_1/\hat{B}_2, \hat{B}_2/\hat{B}_1)$
Gaussian Loss	MSE(r, g)/p		

$\hat{B}_1$ is the median of B_1.
σ_{B_1} is the standard deviation of B_1.

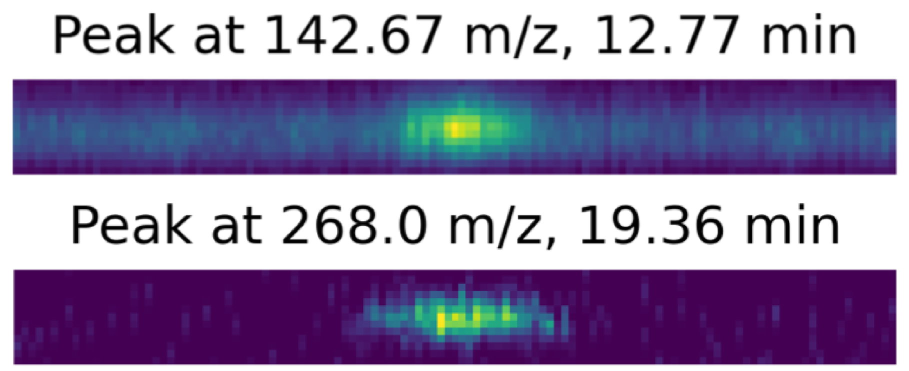
Autonomous Science Data Products

Peak Properties (~3 kB)

The primary science product of ACME is the list of all target peak locations and their calculated properties. In addition to scientific analysis, the properties provide transparency for the onboard classifier's decisions.

Peak Crops (~0.1 MB)

ACME crops a small window of the raw data around each peak location, allowing scientists to directly review and analyze the raw data to confirm and validate ACME's findings.



Background Summary (~60 kB)

ACME generates a summary product that captures the mean and variance of the background noise, providing context for each identified peak. It also serves as a diagnostic tool for sample quality and instrument health.

Total Ion Count (~15 kB)

ACME integrates over the m/z and full migration time axes separately. This produces the mean mass spectra and mean electropherogram, providing an insight into sample content and instrument health.

Science Utility Estimate/Diversity Descriptor (0.3 kB)

Prioritization necessitates the capability to determine which samples would be the most scientifically informative to return. ACME generates the Science Utility Estimate to quantify a sample's value relative to a specified scientific mission goals, and the Diversity Descriptor to enable the priortization of unique samples that are outliers.

Results

ACME was validated on lab-collected data to confirm that scientists are able to successfully analyze and make valid scientific conclusions using only ACME's science products, compared to analyzing the raw data directly. A configuration designed for a hypothetical Europa mission scenario had a peak detection performance with **66% precision and 99% recall**. ACME is available as an open-source repository at: https://github.com/JPLMLIA/OWLS-Autonomy

Acknowledgements

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.