Acoustic Data-based Characterization of Incipient Boiling for Space Applications

Hailey Ernest, Colin Hassett, Elliot Kunz, Cameron Hafer, Dr. Kelly Bodwin, Dr. Alex Dekhtyar
Departments of Statistics, Computer Science, and Mathematics
California Polytechnic State University, San Luis Obispo

Introduction / Background / Goals

Localized heat leaks can lead to a multitude of issues, including loss of fuel or explosion of the fuel tank. Detecting incipient boiling caused by localized heat leaks is important for space-based cryogenic fuel systems in order to manage heat leaks. This project explores whether **acoustic accelerometer data** could be used to **distinguish meaningful boiling behaviors** through unsupervised machine learning.

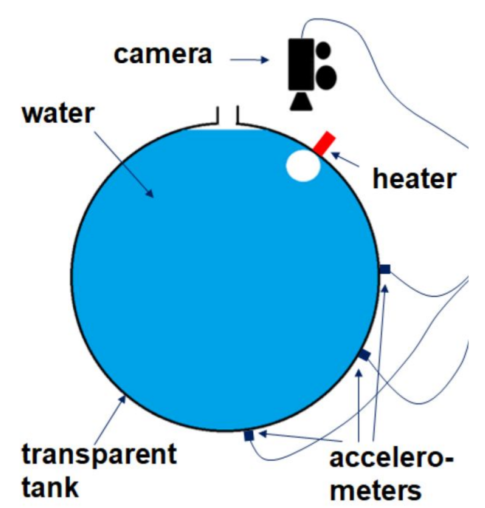

Figure 1: Experimental setup used to obtain the data (1)

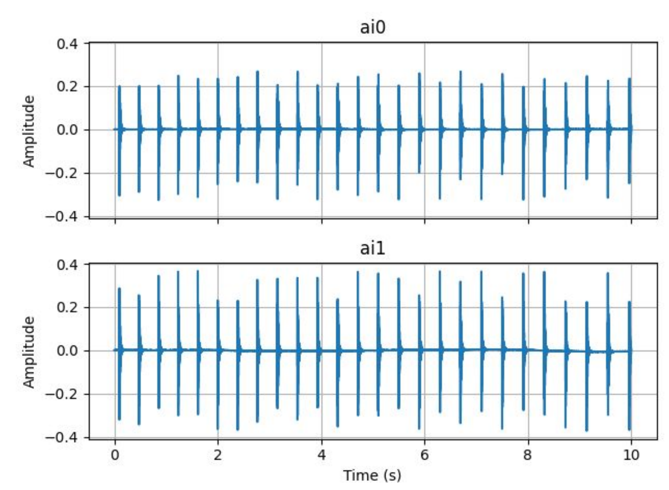

Figure 2: Raw signal from a representative boiling experiment

Methods

LAB: 446 runs ranging from 0.1 to 30 seconds. Both accelerometers used. A noise-floor threshold was applied to filter out runs with little or no boiling activity.

FEATURE ENGINEERING: 36 features engineered from time and frequency domains. Features focused on signal structure, not raw amplitude. Detrending applied to remove runtime-correlated effects. Pairwise features added of peak magnitude differences and cross-accelerometer correlation.

DIMENSIONALITY REDUCTION & CLUSTERING: Compared PCA, diffusion maps, and UMAP. UMAP produced clearest separation. Compared K-Means, HDBSCAN, and spectral clustering. HDBSCAN chosen for handling variable density and outliers. **Final pipeline: UMAP + HDBSCAN.**

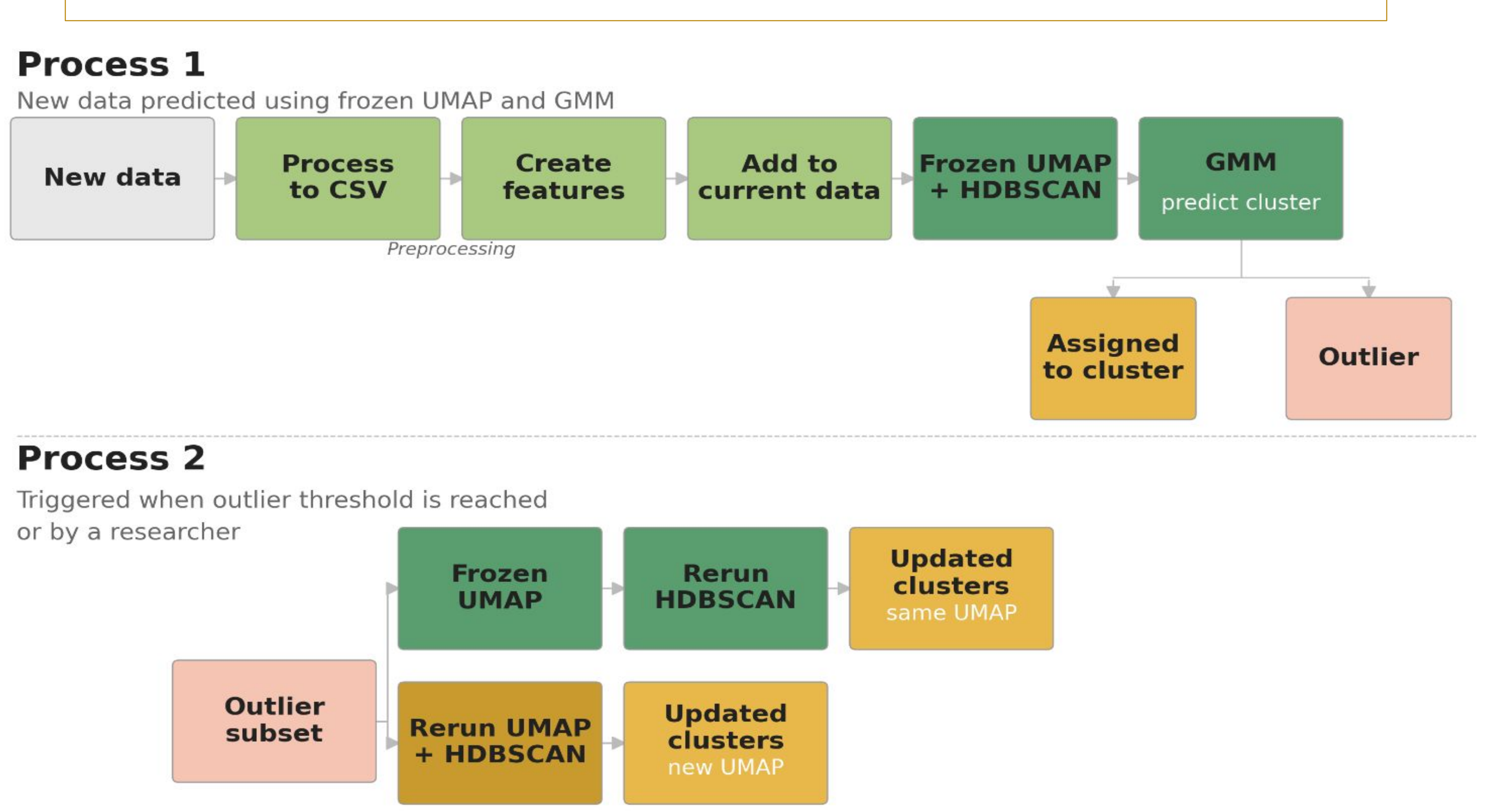

Figure 3: App pipeline workflow. Process 1 classifies new runs through preprocessing, frozen UMAP, and GMM. Process 2 is triggered by outliers and branches into two retraining options.

Results

Main clusters are determined by a noise-floor threshold separating runs into the following:

1. Lack of Boiling — runs with little or no boiling activity
2. Active Boiling — rhythmic and non-rhythmic boiling behaviors
   - **Sporadic**
   - **Chaotic**
   - **Rhythmic**

Subclustering within the rhythmic cluster revealed distinct boiling subtypes:

- **Single Rhythmic**
- **Double Rhythmic**
- **Rhythmic With Climax**

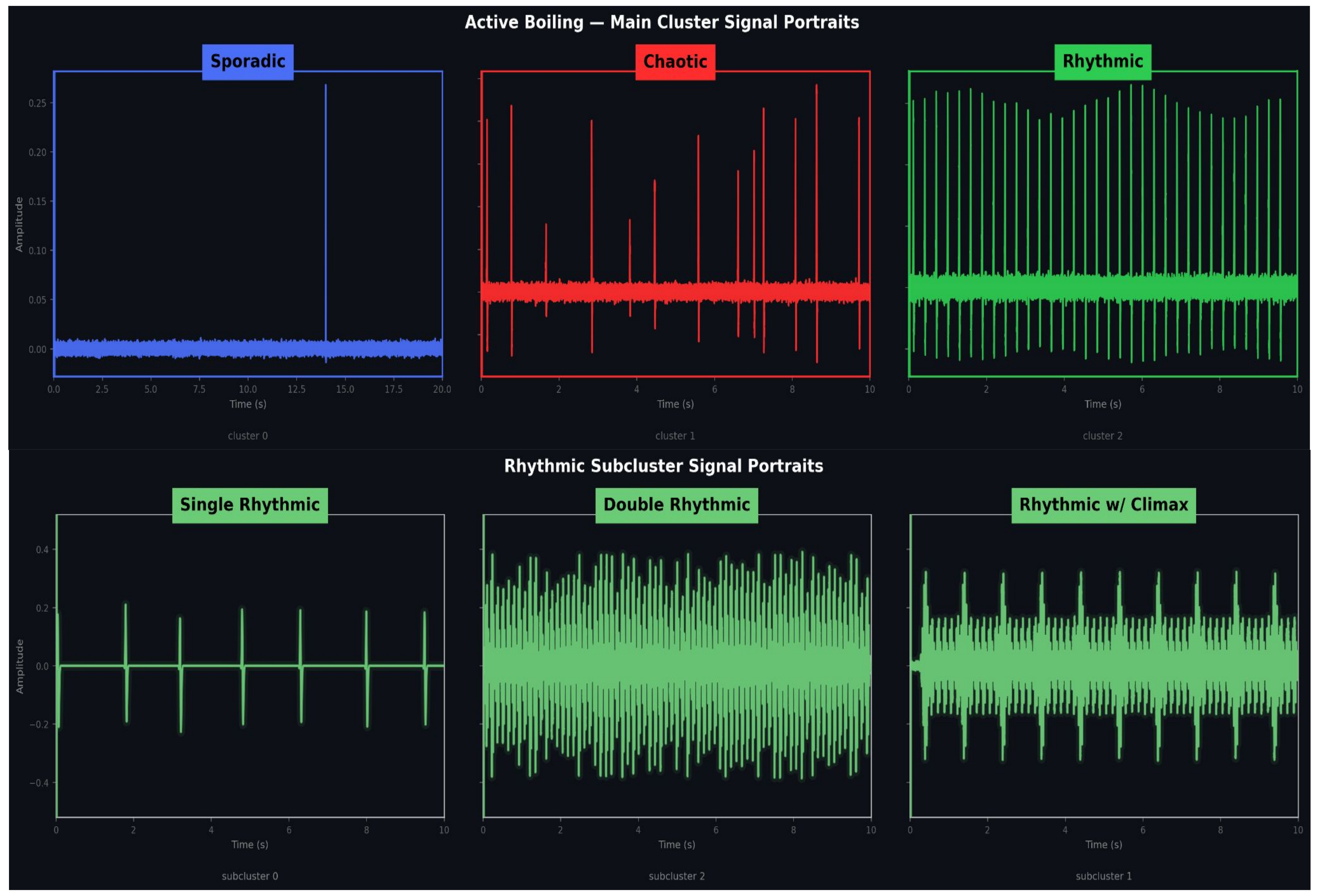

Figure 4: Representative signal portraits by boiling regime. Top: active boiling main clusters showing Sporadic (rare isolated spikes), Chaotic (irregular bursts of varying height), and Rhythmic (consistent repeating spikes). Bottom: rhythmic subclusters including Single Rhythmic, Double Rhythmic, and Rhythmic w/ Climax.

Conclusions / Discussions

We showed a clear distinction between rhythmic and non-rhythmic boiling using UMAP + HDBSCAN with a noise-floor threshold

Subclustering within the rhythmic cluster reveals that rhythmic boiling consists of multiple distinct regimes with different timing and intensity.

Our feature engineering improvements and the use of both accelerometers produced clearer cluster separation than previous analyses.

- **Clustering agreement with human labels**: 73% Rand Index.
- **Clustering stability:** 99% across parameter changes.

Future Directions / Next Steps

The next steps are to continue interpreting the rhythmic subclusters and to test the pipeline on future incoming data using **frozen UMAP transformations and Gaussian mixture models.** We also plan to finalize the visualization **application** and research report.

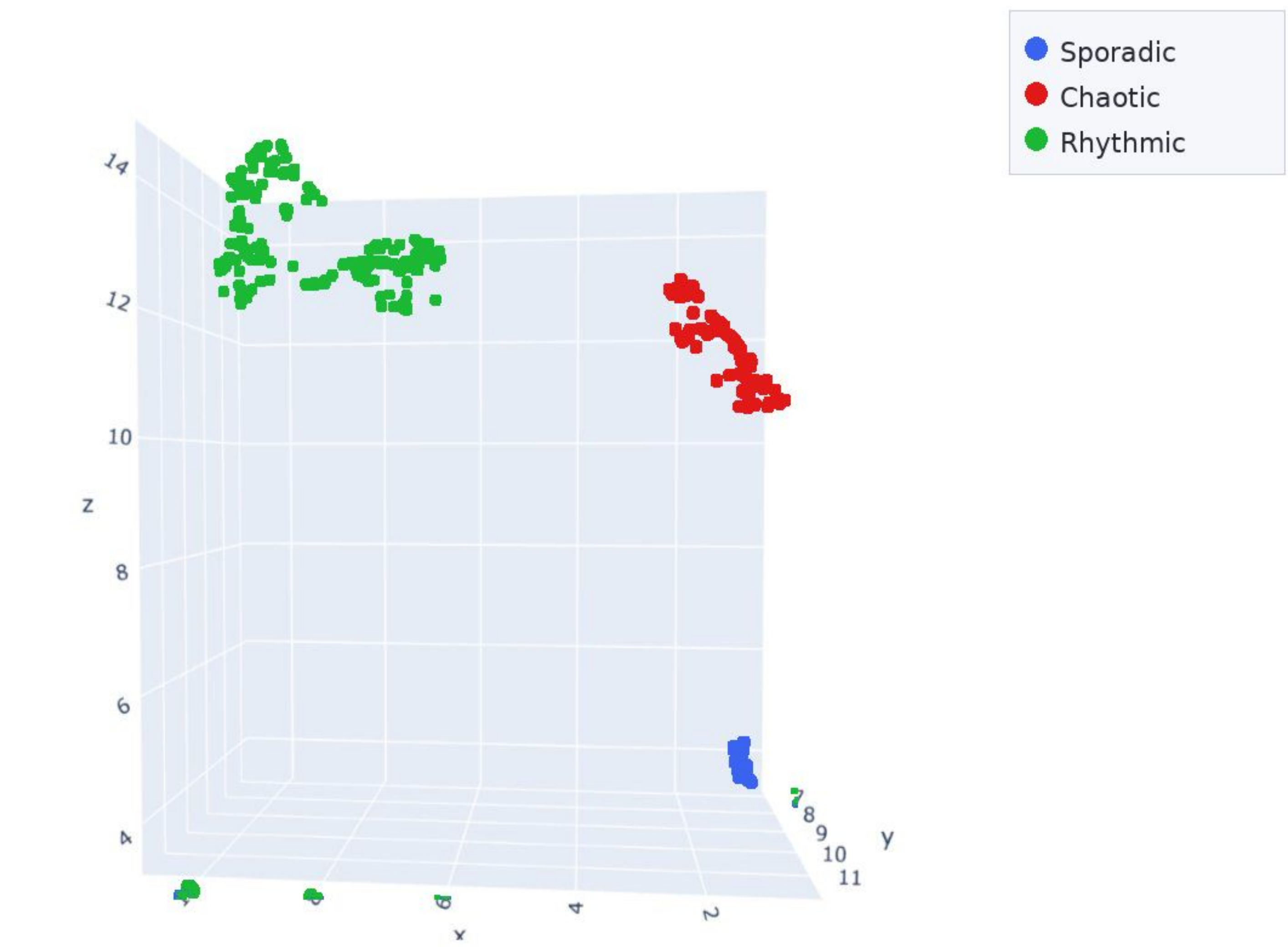

Figure 5: Active boiling clusters. 1) Green cluster represents rhythmic boiling 2) Red cluster represents chaotic boiling 3) Blue cluster represents sporadic boiling.

References

1. M. Khasin, J. Rogers, and V. Osipov, "Acoustic Data-based Characterization of Incipient Boiling for Space Applications," NASA Technical Memorandum NASA/TM-20250009668, September 2025.
2. Gupta, K., et al., "Interpretable Machine Learning for Acoustic Classification of Incipient Boiling Regimes," NASA Technical Memorandum NASA/TM-20250011359, Dec. 2025.
3. M. Khasin, "Acoustic and Visual Data for Incipient Boiling at Local Heat Leaks in a Water Tank (v1.0)," NASA Intelligent Systems Division Datasets,
NASA Ames Research Center, last updated Nov. 25, 2025. Available: https://www.nasa.gov/intelligent-systems-division/datasets/
Appendix

Acknowledgements

- Dr. Kelly Bodwin
- Dr. Alex Dekhtyar
- Dr. Michael Khasin
- NASA Ames Research Center
- Cal Poly Bailey College of Science and Mathematics

Acoustic Data-based Characterization of Incipient Boiling for Space Applications

Hailey Ernest, Colin Hassett, Elliot Kunz, Cameron Hafer, Dr. Kelly Bodwin, Dr. Alex Dekhtyar
Departments of Statistics, Computer Science, and Mathematics
California Polytechnic State University, San Luis Obispo

Introduction / Background / Goals

Localized heat leaks can lead to a multitude of issues, including loss of fuel or explosion of the fuel tank. Detecting incipient boiling caused by localized heat leaks is important for space-based cryogenic fuel systems in order to manage heat leaks. This project explores whether **acoustic accelerometer data** could be used to **distinguish meaningful boiling behaviors** through unsupervised machine learning.

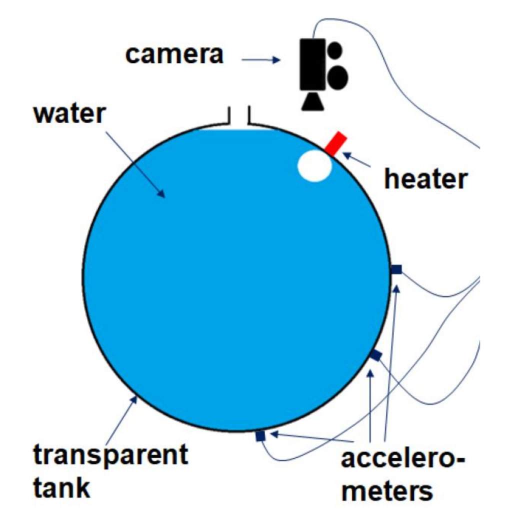

Figure 1: Experimental setup used to obtain the data (1)

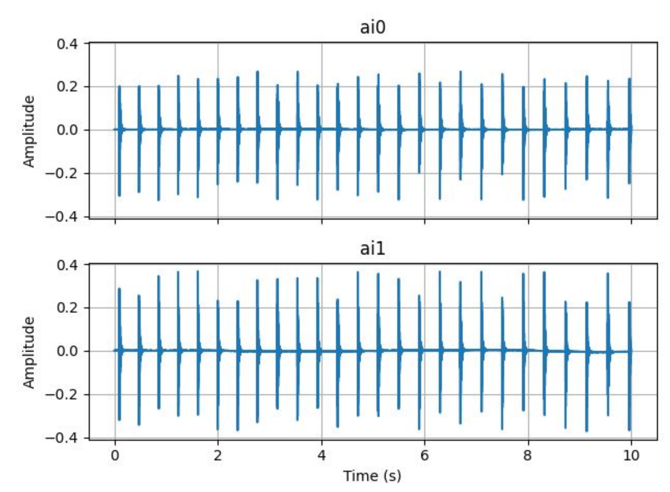

Figure 2: Raw signal from a representative boiling experiment

Methods

LAB: 446 runs ranging from 0.1 to 30 seconds. Both accelerometers used. A noise-floor threshold was applied to filter out runs with little or no boiling activity.

FEATURE ENGINEERING: 36 features engineered from time and frequency domains. Features focused on signal structure, not raw amplitude. Detrending applied to remove runtime-correlated effects. Pairwise features added of peak magnitude differences and cross-accelerometer correlation.

DIMENSIONALITY REDUCTION & CLUSTERING: Compared PCA, diffusion maps, and UMAP. UMAP produced clearest separation. Compared K-Means, HDBSCAN, and spectral clustering. HDBSCAN chosen for handling variable density and outliers. **Final pipeline: UMAP + HDBSCAN.**

References

1. M. Khasin, J. Rogers, and V. Osipov, "Acoustic Data-based Characterization of Incipient Boiling for Space Applications," NASA Technical Memorandum NASA/TM-20250009668, September 2025.
2. Gupta, K., et al., "Interpretable Machine Learning for Acoustic Classification of Incipient Boiling Regimes," NASA Technical Memorandum NASA/TM-20250011359, Dec. 2025.
3. M. Khasin, "Acoustic and Visual Data for Incipient Boiling at Local Heat Leaks in a Water Tank (v1.0)," NASA Intelligent Systems Division Datasets,
NASA Ames Research Center, last updated Nov. 25, 2025. Available: https://www.nasa.gov/intelligent-systems-division/datasets/
Appendix

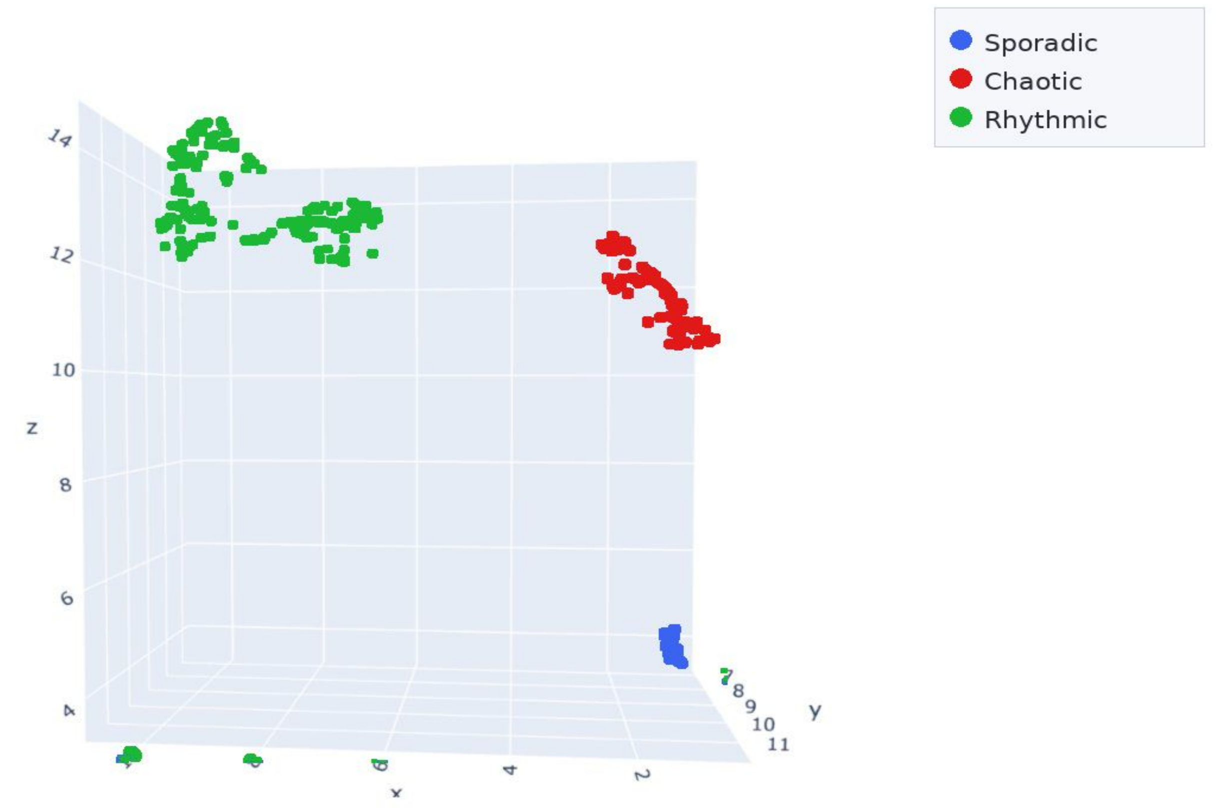

Figure 5: Active boiling clusters. 1) Green cluster represents rhythmic boiling 2) Red cluster represents chaotic boiling 3) Blue cluster represents sporadic boiling.

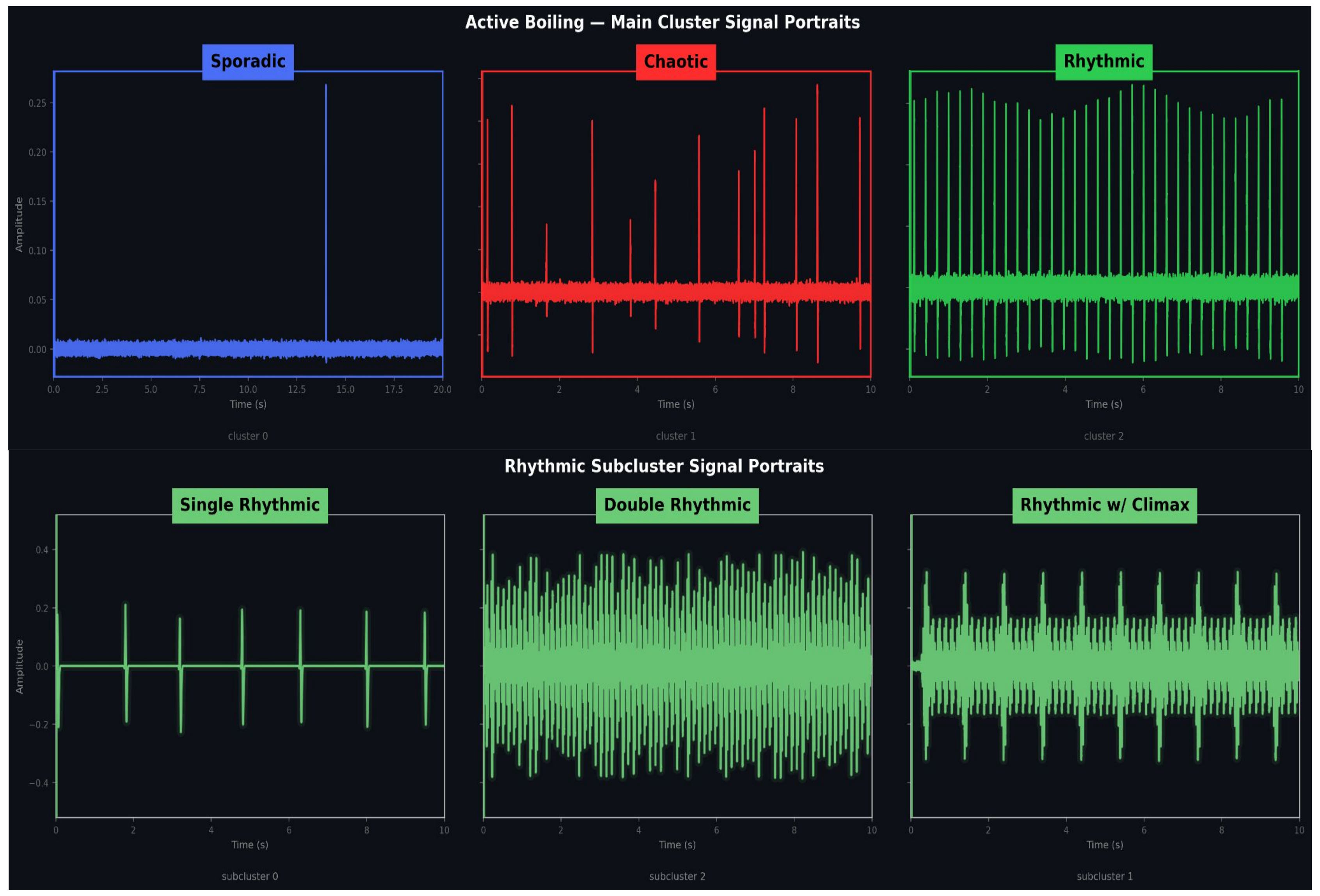

Figure 4: Representative signal portraits by boiling regime. Top: active boiling main clusters showing Sporadic (rare isolated spikes), Chaotic (irregular bursts of varying height), and Rhythmic (consistent repeating spikes). Bottom: rhythmic subclusters including Single Rhythmic, Double Rhythmic, and Rhythmic w/ Climax.

Results

Main clusters are determined by a noise-floor threshold separating runs into the following:

1. Lack of Boiling — runs with little or no boiling activity
2. Active Boiling — rhythmic and non-rhythmic boiling behaviors
   - **Sporadic**
   - **Chaotic**
   - **Rhythmic**

Subclustering within the rhythmic cluster revealed distinct boiling subtypes:

- **Single Rhythmic**
- **Double Rhythmic**
- **Rhythmic With Climax**

Conclusions / Discussions

We showed a clear distinction between rhythmic and non-rhythmic boiling using UMAP + HDBSCAN with a noise-floor threshold

Subclustering within the rhythmic cluster reveals that rhythmic boiling consists of multiple distinct regimes with different timing and intensity.

Our feature engineering improvements and the use of both accelerometers produced clearer cluster separation than previous analyses.

- **Clustering agreement with human labels:** 73% Rand Index.
- **Clustering stability:** 99% across parameter changes.

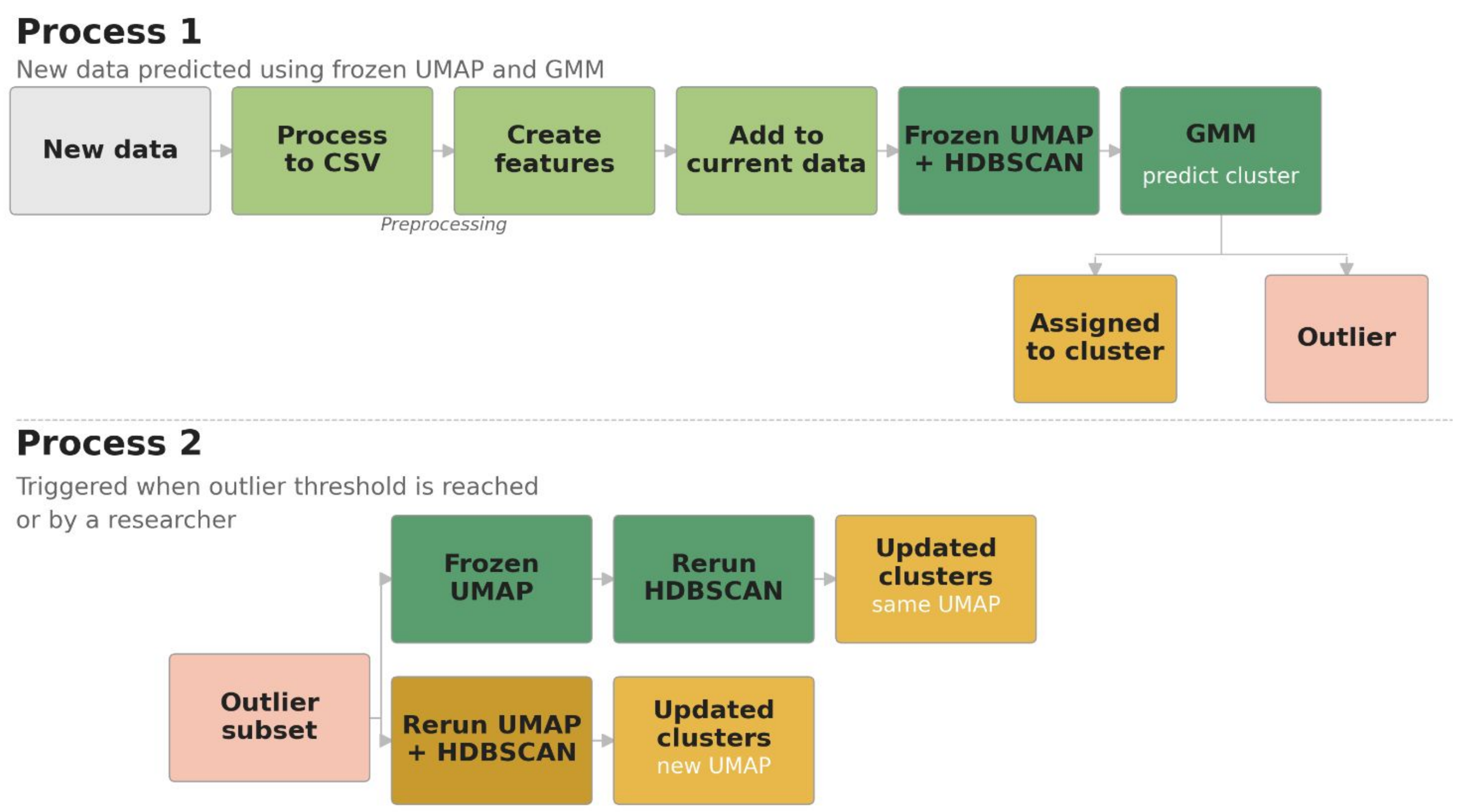

Figure 3: App pipeline workflow. Process 1 classifies new runs through preprocessing, frozen UMAP, and GMM. Process 2 is triggered by outliers and branches into two retraining options.

Future Directions / Next Steps

The next steps are to continue interpreting the rhythmic subclusters and to test the pipeline on future incoming data using **frozen UMAP transformations and Gaussian mixture models.** We also plan to finalize the visualization **application** and research report.

Acknowledgements

- Dr. Kelly Bodwin
- Dr. Alex Dekhtyar
- Dr. Michael Khasin
- NASA Ames Research Center
- Cal Poly Bailey College of Science and Mathematics