18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.

SpaceOps-2025, ID # 217

AI-based Novelty Detection in Space Operations: Three Years of Operational Experience and Progression at GSOC

Kathrin Helmsauer [a]*, Agnese Del Moro [a], Tobias Göttfert [a], Clemens Schefels [a], Leonard Schlag [a]

[a] *German Aerospace Center (DLR), Space Operations and Astronaut Training, 82234 Weßling, Bavaria, Germany*
* Corresponding Author, kathrin.helmsauer@dlr.de

Abstract

Anomaly detection in satellite telemetry is critical for ensuring operational reliability and early fault detection. This paper presents the integration of the AI-based *Automated Telemetry Health Monitoring System* (ATHMoS) into the operational workflows of the *German Space Operations Center* (GSOC) at the *German Aerospace Center* (DLR). We discuss the challenges encountered during deployment and the solutions implemented to enhance ATHMoS' effectiveness. Key improvements, informed by engineer feedback, include refined parameter classification—particularly expanded support for highly periodic parameters with little to no noise and certain discrete parameters like counters—as well as a user-driven reclassification workflow to reduce false positives from nominal events such as maneuvers and maintenance activities. Additionally, we introduce a continuous integration (CI) pipeline that automates configuration testing across multiple satellite telemetry datasets, streamlining performance evaluation, optimization, and comparison with the operational ATHMoS system. These advancements enable broader applicability of ATHMoS across diverse satellite missions, including both large-scale scientific and communication missions as well as resource-constrained platforms such as CubeSats. Furthermore, ongoing developments focus on a real-time, onboard version of ATHMoS, laying the foundation for future advancements in AI-driven telemetry health monitoring.
Keywords: Telemetry, Time Series, Artificial Intelligence, Machine Learning, Data Analysis, Space Operations

Acronyms/Abbreviations

AI for Automation of Satellite Health Monitoring and Ground Operations (AISHGO)
Anomaly Report Ident (ARI)
Automated Telemetry Health Monitoring System (ATHMoS)
Cubesat to Accommodate Payloads and Technology Experiments (CAPTn-1)
Continuous integration (CI)
Centre National d'Études Spatiales (CNES)
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
German Aerospace Center/Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
European Space Agency (ESA)
European Space Agency Anomaly Detection Benchmark (ESA-ADB)
Failure Detection, Isolation and Recovery (FDIR)
Fast Fourier Transform (FFT)
German Space Operations Center (GSOC)
Gravity Recovery and Climate Experiment-Follow-On (GFO)
Intrinsic Dimensions (ID)
Low Earth orbit (LEO)
Mars Science Laboratory (MSL)
Outlier Probability Via Intrinsic Dimension (OPVID)
Probabilistic Intrinsic Dimension Outlier Scores (PIDOS)
Recommendation Ident (RI)
Soil Moisture Active Passive (SMAP)
Space Experiment for Satellite Artificial intelligence Monitoring (SESAM)
Telemetry (TM)
Visualization and Data Analysis software (ViDA)

1. Introduction

In satellite operations, telemetry data is crucial for monitoring system health and detecting anomalies that may indicate potential failures. Modern satellites generate vast amounts of telemetry data, often tracking tens of thousands of parameters. For example, each *GRACE Follow-On* (GFO) satellites, operated by the *German Space Operations Center* (GSOC) at the *German Aerospace Center* (DLR), monitors approximately 80,000 unique housekeeping parameters. The scale and complexity of this data present significant challenges for system engineers and operators, highlighting the need for automated analysis tools.

To address this challenge, space operation centers are increasingly turning to machine-learning-based tools for anomaly detection. At GSOC, we have developed the *Automated Telemetry Health Monitoring System* (ATHMoS), a semi-supervised novelty detection system based on the *Outlier Probability Via Intrinsic Dimension* (OPVID) algorithm. ATHMoS has been operational since 2022 for the TerraSAR-X and TanDEM-X satellite missions, automatically analysing over 1000 telemetry parameters and assisting engineers in identifying system irregularities.

This paper begins with a brief overview of related work from other space agencies and organizations that are advancing the use of machine learning for anomaly detection in Section 2. Section 3 revisits the original ATHMoS workflow as deployed at GSOC in 2022. In Section 4, we present the enhancements made to ATHMoS based on operational experience, including improvements in parameter classification—particularly for challenging cases such as counters—and the introduction of a reclassification workflow to reduce false positives from nominal events like manoeuvres and maintenance operations. We also describe the development of a continuous integration (CI) pipeline that enables automated evaluation of ATHMoS configurations across multiple satellite telemetry datasets. This streamlines our system optimization and performance tracking.

As a result of these enhancements, ATHMoS has been expanded to support additional missions at GSOC, including both large-scale science and communication satellites as well as resource-constrained platforms like CubeSats. Looking ahead (Sec. 5), we outline our efforts to develope a real-time, onboard version of ATHMoS, with a first proof-of-concept planned for the *Cubesat to Accommodate Payloads and Technology Experiments* (CAPTn-1) mission, scheduled for launch in early 2026.

2. Related Work

The use of machine learning and its potential to detect anomalies has many different areas of application where novelty detection is of importance, such as intrusion detection in computer systems or networks or video surveillance [1]. While the capabilities of machine learning can be used for satellite health monitoring and anomaly detection, they can also be utilized in other areas related to space operations such as mission planning or the processing of image data [2].

Various space agencies and companies are actively researching and operationalizing their novelty and anomaly detection approaches. A recent experiment by the *Centre National d'Études Spatiales* (CNES) named *Space Experiment for Satellite Artificial intelligence Monitoring* (SESAM) combines a One-Class Support Vector Machine with a clustering method based on *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) in an ensemble approach [3]. The goal of SESAM to quickly detect on board a satellite has been shown difficult to achieve. Similarly, Airbus is also investigating new *Failure Detection, Isolation and Recovery* (FDIR) methods and evaluated various autoencoder-based neural network architectures [4]. A dataset composed of six telemetry channels representing currents obtained from different solar array sections was used in their analysis.

With respect to continually improving existing algorithms as well as aiding in the research and evaluation of new approaches, high-quality labelled datasets close to the domain of satellite operations play a vital role. Many publications rely on a labelled dataset extracted from the *Soil Moisture Active Passive* (SMAP) satellite and the *Mars Science Laboratory* (MSL) [5]. More recently, the *European Space Agency Anomaly Detection Benchmark* (ESA-ADB) was published [6, 7], which provides over 3 billion data points from more than 30 years of data from three satellite missions and aims to shape new and upcoming anomaly detection approaches in the coming years.

3. ATHMoS – Automated Telemetry Health Monitoring System (Initial Design)

The *Automated Telemetry Health Monitoring System* (ATHMoS), developed at GSOC, is a modular framework for anomaly detection in satellite telemetry data. Rather than a single algorithm, ATHMoS comprises a sequence of pre-processing, detection, and post-processing steps that together form a comprehensive analysis workflow. Its design enables flexible adaptation to the wide variety of telemetry data encountered in satellite operations.

This section provides an overview of the system's core architecture and processing workflow as initially developed in 2022. Subsequent enhancements informed by real-world use and engineering feedback are described in detail in

Section 4.2. For further technical details on the underlying anomaly detection method, which is based on the *Outlier Probability via Intrinsic Dimension* (OPVID) algorithm, see [8]. The transition from a research prototype to a stable operational tool is discussed in [9].

ATHMoS is supported by the *Visualization and Data Analysis* (ViDA) framework [10], which offers a web-based front-end for interactive exploration of telemetry data and detection results. Over the past four years, ViDA has evolved significantly, incorporating tools from the *AI for Automation of Satellite Health Monitoring and Ground Operations* (AISHGO) project [11], as outlined in [12].

While ATHMoS is referred to as a "system" in its acronym, it is not tied to any specific software platform or hardware environment. Its effectiveness relies on the complete processing pipeline—from initial data cleaning and parameter classification to anomaly scoring and result interpretation.

3.1 Pre-processing

ATHMoS begins by processing raw satellite telemetry data to prepare it for training and anomaly detection. This step is essential to ensure data completeness, extract meaningful features, and enhance model accuracy. Our pre-processing consists of three main steps:

- Gap Detection and Data Cleaning: Time-series data gaps can disrupt statistical analysis and feature extraction. ATHMoS detects these gaps and removes the affected time windows to maintain data integrity. Additionally, we interpolate the time series to standardize timestamps, ensuring uniform time steps for more consistent feature extraction.
- Parameter Classification and Feature Optimization: Each telemetry parameter is automatically classified according to its type, enabling the selection of optimized features for subsequent analysis. Parameters representing discrete flags (e.g., system status indicators) are handled differently from more or less noisy, continuous physical attributes (e.g., temperatures or voltages) to enhance anomaly detection. To address cases where standard classifications led to suboptimal detection—particularly for parameters that didn't fit cleanly into discrete or continuous categories—we have introduced additional parameter types, as detailed in Section 4.2.2.
- Noise Reduction and Smoothing: In addition to raw time-series data, a smoothed version is generated to reduce noise for continuous parameters. Both the smoothed data and the noise serve as an additional input for our feature extraction in later steps. Initially, we applied a smoothing algorithm based on the Fast Fourier Transform (FFT), which performed well for identifying dominant frequencies and preserving them during smoothing. However, this approach introduced issues later in the workflow, particularly for periodic parameters without noise—these challenges and our improved solution are discussed in Section 4.2.1.

3.2 Initial Training and Model Update

The first stage of ATHMoS establishes a baseline model that represents the nominal or typical behavior of satellite telemetry parameters. This is a resource-intensive process, particularly in the initial setup, as it typically requires one year of historical telemetry data per parameter to capture seasonal and long-term variations. Key training steps include:

- Feature Vector Extraction: A sliding window approach is applied to the pre-processed data to generate feature vectors. For low Earth orbit satellites, a window size of 1.5 hours with a step size of 30 minutes results in approximately 17,500 feature vectors over a one-year period. This window size encompasses the typical orbital period of *low Earth orbit* (LEO) satellites. Fig. 1 illustrates the features for both discrete and continuous parameters.

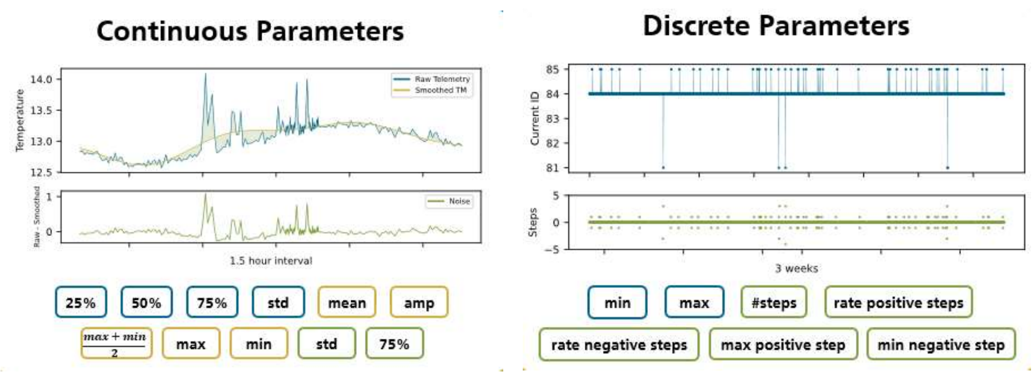


Fig. 1. Features for Discrete and Continuous Parameters

- Outlier Filtering with Clustering: Since the training data is not labeled, a variant of *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) is used to remove obvious outliers, improving model robustness. Fig. 2 illustrates the results of applying our version of DBSCAN-Clustering to a three-dimensional feature set.

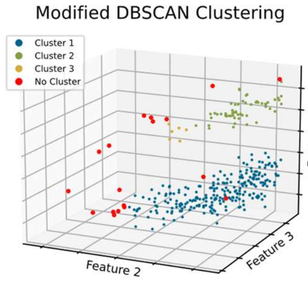


Fig. 2. Modified DBSCAN-Clustering

- Model Generation: The core ATHMoS algorithm as described in [8] is applied to the filtered feature vectors. Intermediate results, including both the *Intrinsic Dimensions* (ID) (c.f. Fig. 3), and the *Probabilistic Intrinsic Dimension Outlier Scores* (PIDOS), are stored to enable efficient inference during the testing phase.

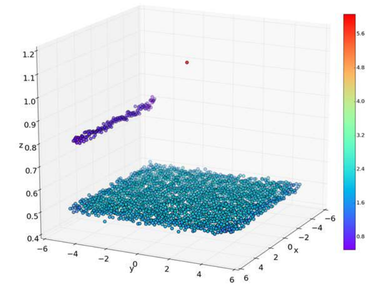


Fig. 3. Feature Vectors in Separate Fuzzy Hyperplanes, Each Characterized by its Intrinsic Dimension

To maintain model accuracy and adapt to evolving telemetry behavior, ATHMoS employs a dual-scale training approach consisting of:

- A long-term model, trained on one year of data, which is updated weekly to reflect long-term trends.
- A short-term model, trained on one month of data, which is updated daily to capture recent variations.

Both models contribute to anomaly detection by providing two independent novelty scores, allowing ATHMoS to distinguish between persistent anomalies and short-term deviations: The long-term model serves as the baseline representation of nominal satellite telemetry behavior, incorporating seasonal patterns and long-term variations. Every week, the model is retrained, where the oldest week of data is discarded and the newest week of data is added to ensure the model remains up to date. Feature vectors used in retraining exclude high-priority anomalies detected in previous runs, ensuring the model focuses on nominal behavior. In contrast, the short-term model is designed to detect recent deviations that may not (or not yet) be reflected in the long-term model. It is trained on the most recent month of telemetry data, offering finer sensitivity to emerging trends. This model is updated daily, where the oldest day's data is removed and the latest day's data is incorporated into training.

3.3 Model Inference and Anomaly Detection

ATHMoS evaluates new satellite telemetry data by comparing it to the trained models, assessing deviations from expected behavior. Unlike real-time anomaly detection systems, ATHMoS operates on archived housekeeping data, making it best suited for daily batch processing of the most recent telemetry records. To ensure consistency, the inference process follows the same pre-processing and feature vector computation steps as the training phase. Once feature vectors are generated, anomaly detection proceeds as follows:

- Novelty Score Computation: Each feature vector from the test data is compared to the trained model (see Section 3.2 and Fig. 4), and its novelty score is computed using the *Outlier Probability Via Intrinsic*

Dimension (OPVID) and *Probabilistic Intrinsic Dimension Outlier Scores* (PIDOS). This score quantifies the deviation of the new data from the learned nominal behavior.

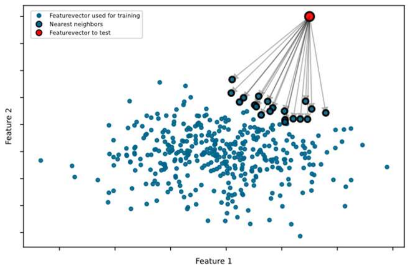


Fig. 4. Compare Each New Feature Vector to its Nearest Neighbours

- Anomaly Labeling: Novelty scores determine whether a time window is classified as an anomaly. Since feature vectors are extracted from overlapping time windows, ATHMoS leverages this redundancy to improve robustness: isolated threshold breaches that are not corroborated by neighboring windows are not immediately classified as high-priority anomalies, as can be seen in Fig. 5. Instead, only sustained deviations trigger high-priority detections, reducing the likelihood of false positives.

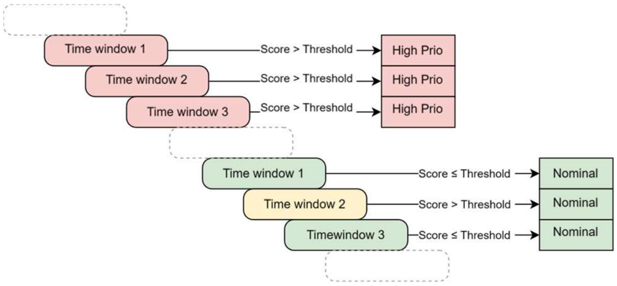


Fig. 5. Flagging High Priority Novelties

By applying this structured approach, ATHMoS ensures that anomalies are detected with both sensitivity and stability, minimizing spurious alerts while effectively identifying deviations that require further investigation.

4. Collecting and Incorporating the Engineers' Feedback

After deploying ATHMoS operationally, we have identified several areas for improvement based on both systematic testing and user feedback. As we gained more experience, discussions with engineers highlighted specific aspects that required refinement. To address this, we have established a structured feedback collection process, enabling us to systematically identify shortcomings and prioritize enhancements. Additionally, we implemented a rigorous testing framework to evaluate changes before deployment. This includes a CI pipeline that automates testing across different configurations and telemetry datasets. Looking ahead, we plan to introduce a second operational workflow, designed to facilitate the smooth transition of major updates while minimizing disruptions to ongoing anomaly detection. By combining these elements, our feedback process enables continuous refinement of ATHMoS while maintaining an uninterrupted operational workflow.

4.1 Collecting Feedback

To systematically improve ATHMoS while minimizing disruptions to operations, we have established a structured feedback process that allows engineers to provide input with minimal effort while ensuring thorough evaluation of proposed changes. This process consists of three key components:

- Internal Wiki for Quick Notes and Screenshots: Engineers can easily document observations, issues, or suggestions on a dedicated internal wiki page (Fig. 6). The wiki serves as a low-effort, low-barrier way to quickly capture feedback, including screenshots and/or brief descriptions of unexpected detections and/or missed anomalies. This ensures that valuable insights are not lost and can be systematically reviewed in later discussions.

ATHMoS User Feedback

Please fill in your feedback on ATHMoS detections in the tables below.

TDX feedback

nr	user	parameter (spid if needed)	time of occurrence	false positive	needs discussion	missed anomaly	comment
0	Example User	AST99999	2023-05-17 13:50 UTC	☐	☐	☑	example: This anomaly was not flagged, maybe add ref to ARI and give some context

Fig. 6. Screenshot of Wiki Page: ATHMoS User Feedback

- Regular Review Meetings with Engineers: Periodic review meetings provide a structured setting to collect feedback, discuss recurring issues, and prioritize improvements. Engineers present their experiences with ATHMoS, while we report on locally tested modifications and their impact. Decisions are made collaboratively on which changes should be implemented, ensuring that updates align with operational needs.
- On-Demand One-on-One Discussions: In addition to scheduled review meetings, ad hoc discussions with individual engineers help clarify specific requests in more detail. These one-on-one sessions allow for a deeper understanding of how ATHMoS is used in daily operations and provide valuable context for refining anomaly detection strategies. This flexible approach ensures that feedback is both comprehensive and actionable.

4.2 Enhancements to Parameter Handling

Initially, ATHMoS categorized telemetry parameters as either discrete or continuous, tailoring the anomaly detection process—particularly the construction of feature vectors—accordingly. For most continuous parameters, such as voltages and temperatures, this approach has proven highly effective, successfully identifying all relevant anomalies within the analysed timeframe. However, our primary focus now is on minimizing false positives, ensuring that (ideally) every detected novelty is meaningful to our engineers. The challenge has been more pronounced for discrete parameters, which have exhibited a higher rate of false-positive detections. As a result, our enhancements have been primarily directed toward improving their classification and handling.

4.2.1 Continuous Parameters

For continuous parameters, our pre-processing workflow begins with a smoothing algorithm to help distinguish the true signal from noise. The computed features are derived from three key sources: the raw telemetry data, the smoothed signal (representing the underlying "true" behaviour of the signal), and the noise (the difference between the raw and smoothed signal). This approach has been highly effective for parameters with inherent noise. However, unexpected challenges arose for parameters that exhibit little to no noise, particularly periodic signals such as positional data (c.f. Fig. 7). In these cases, anomalies must be evaluated relative to their expected cycles rather than absolute deviations.

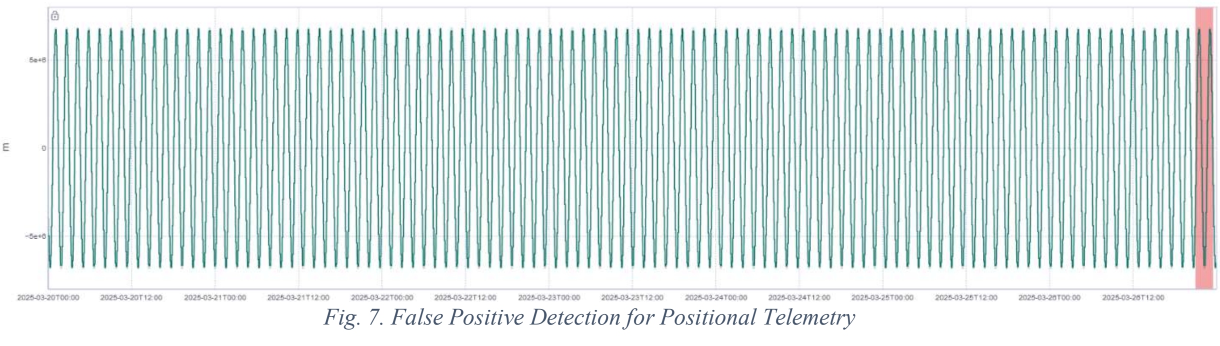


Fig. 7. False Positive Detection for Positional Telemetry

One of the primary causes of false positive detections for periodic parameters stemmed from our initial choice of smoothing algorithm. However, FFT introduced significant artefacts at the edges of the analysed timeframe, especially when the signal's period was incomplete. To mitigate the abrupt boundary effects caused by FFT, we initially applied a strategy of loading additional data beyond the required timeframe and trimming the edges. While this helped, it did not fully resolve the issue. As illustrated in Fig. 8, the smoothing process still introduced distortions that affected anomaly detection.

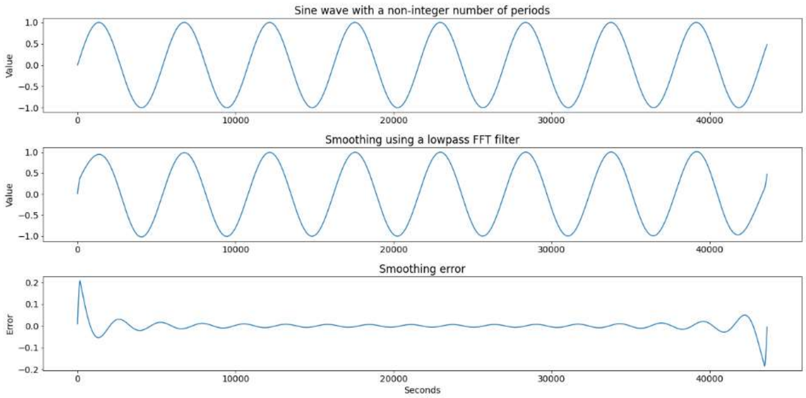


Fig. 8. Extended Edge Effects of FFT-Smoothing for Periodic Parameters without Noise

Additionally, FFT-based smoothing led to an unintended consequence: real novelties were spread across a much larger time window than expected. Artificially injecting anomalies into the data revealed that ATHMoS flagged an excessively broad timeframe as anomalous, which reduced the precision of detections (see Fig. 9). This behaviour stemmed from the same root cause—the way FFT smoothing distributes errors over time (Fig. 10).

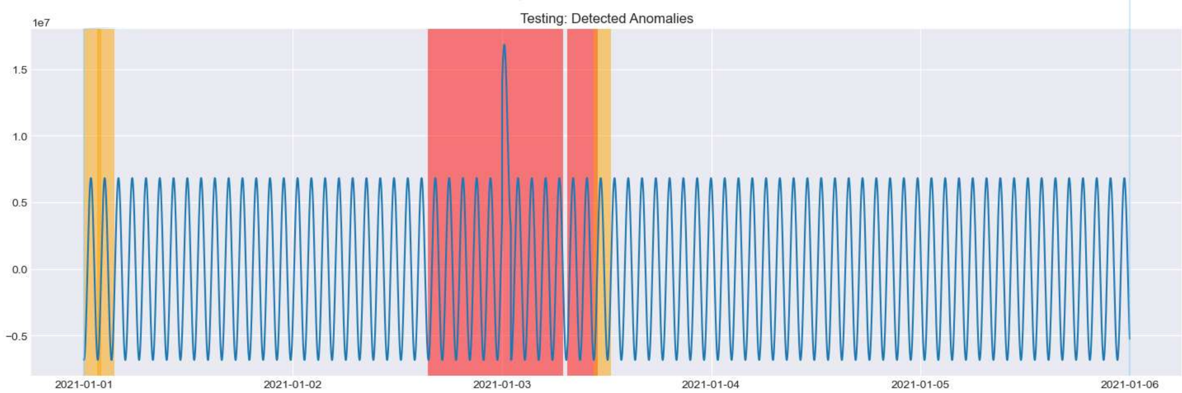


Fig. 9. Extended Spread of Detected Novelty

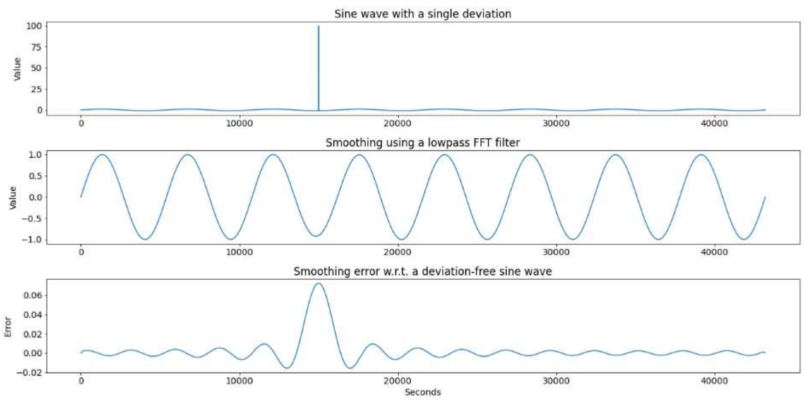


Fig. 10. Extended Spread Caused by FFT Smoothing

To address these issues, we replaced FFT smoothing with Savitzky-Golay filtering. This method better preserves local signal characteristics while reducing edge distortions and unnecessary temporal spread in anomaly detection. The benefits were immediately apparent: Fig. 11 demonstrates that boundary artefacts are significantly reduced, in turn reducing the edge effects. As seen in Fig. 12, detected novelties are now confined to a more focused and meaningful timeframe.

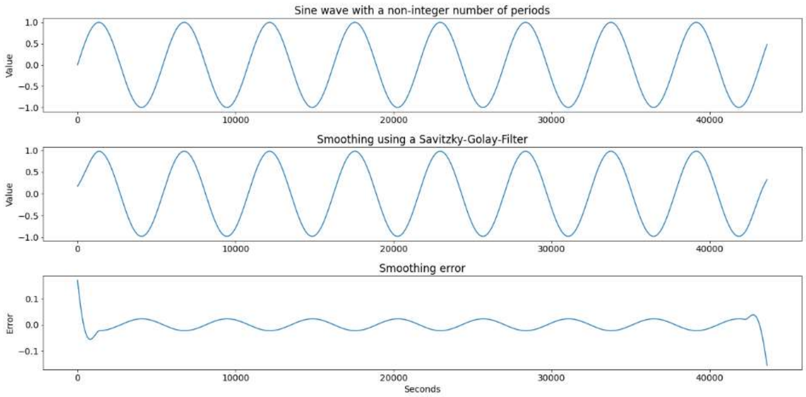


Fig. 11. Reduced Edge Effects for Savitzky-Golay Smoothing

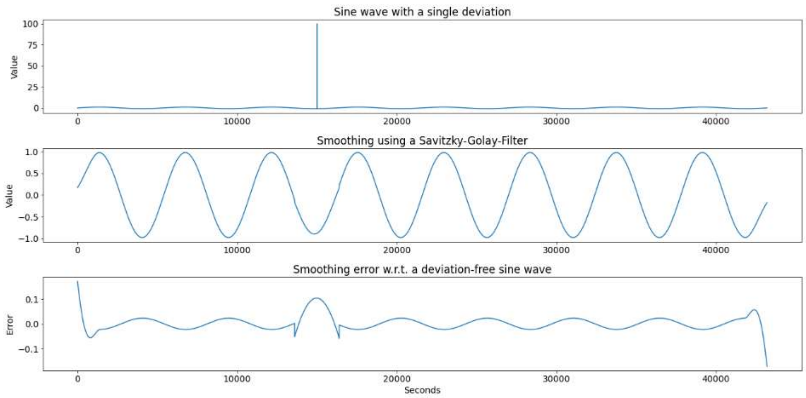


Fig. 12. More Uniform Spread for Savitzky-Golay Smoothing

As with the FFT-based approach, we continue to load additional data beyond the required analysis window to buffer the boundaries. However, with Savitzky-Golay smoothing, the edge effects are more localized. By trimming the smoothed signal back to the original timeframe after filtering, we can effectively eliminate residual distortions near the edges. This step ensures that any boundary artefacts do not influence the actual detection results, leading to more robust and reliable outputs within the analysis range.

Importantly, switching to Savitzky-Golay smoothing did not negatively impact novelty detections for noisy parameters. Additionally, it virtually eliminated false detections at boundaries (Fig. 13) and improved the precision of high-priority anomaly flagging (Fig. 14).

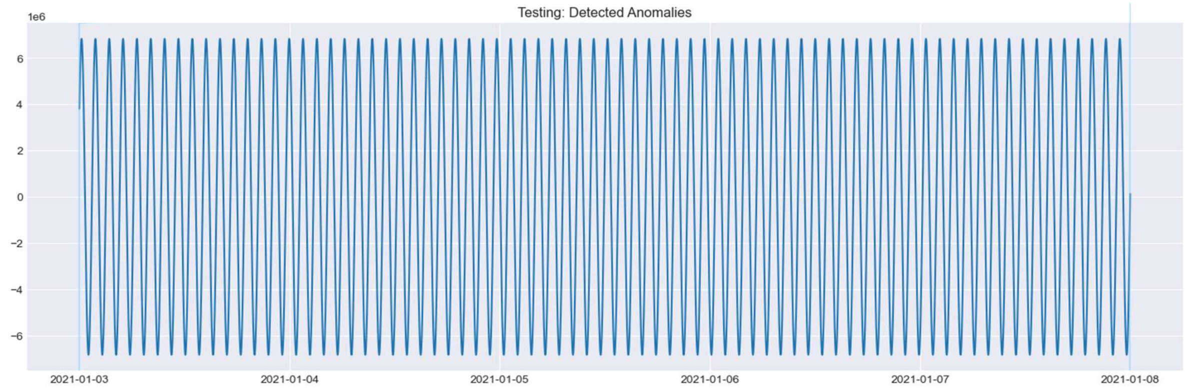


Fig. 13. ATHMoS Detections with Savitzky-Golay Smoothing for Non-Noisy Periodic Parameters

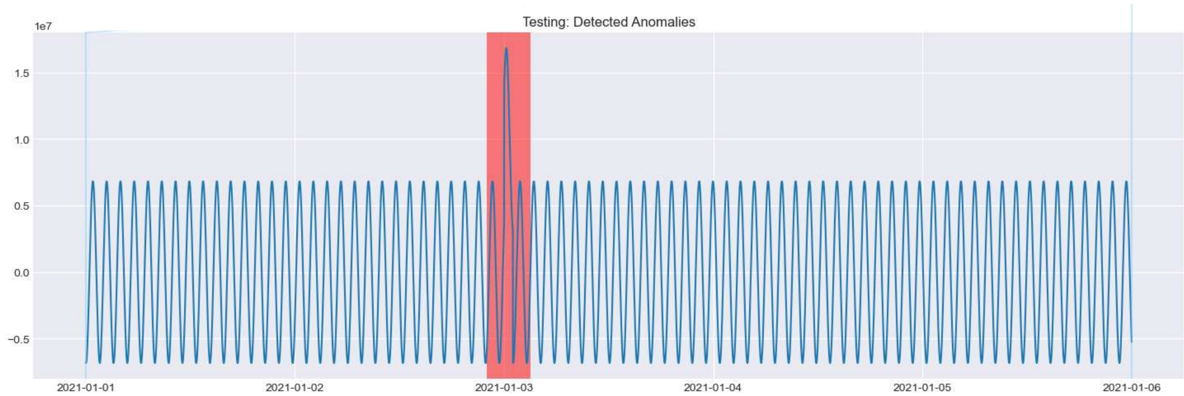


Fig. 14. Flagging High Priority Novelties using Savitzky-Golay Smoothing

By refining our smoothing approach, we have significantly reduced false positives while maintaining high sensitivity for true novelties, bringing us closer to our goal of detecting only meaningful anomalies for engineers.

4.2.2 Discrete Parameters

Unlike continuous parameters, many discrete telemetry parameters do not exhibit a strong correlation with the standard 90-minute orbital period of LEO satellites. As a result, analysing these parameters within such a short timeframe can lead to false positive detections, as anomalies may not be adequately contextualized. Fig. 15 illustrates a false-positive detection in a discrete parameter, which upon closer inspection (Fig. 16) appears to result from an insufficient time window rather than an actual anomaly. Since discrete parameters often reflect system state transitions or event-driven changes rather than continuous trends, a longer observational window is required to properly assess their behaviour.

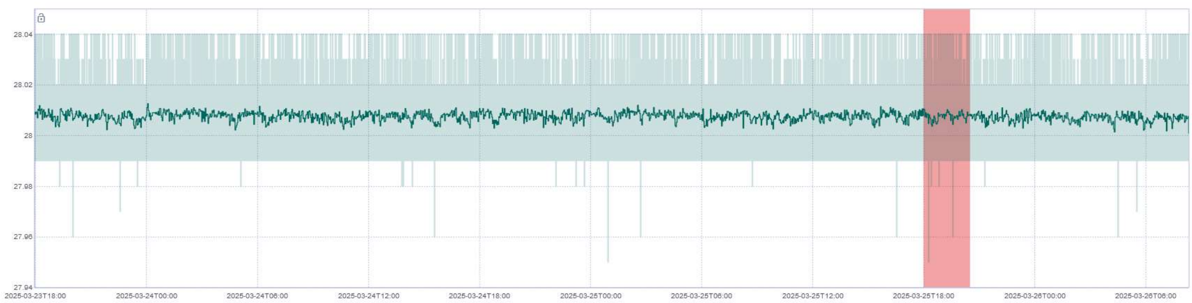


Fig. 15. False-Positive Novelty in Discrete Parameter

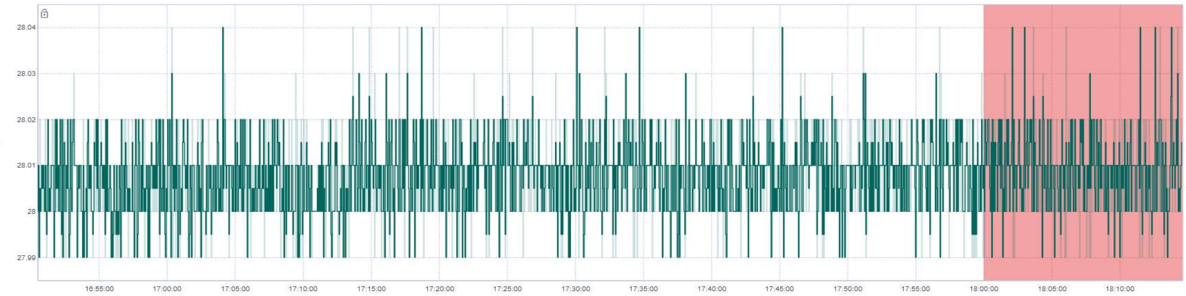


Fig. 16. Zooming into Detected Novelty

To improve anomaly detection for discrete parameters, we plan to increase the window size to 24 hours, with an 8-hour sliding step. This adjustment ensures that anomalies are assessed over a more representative timescale, capturing potential patterns that might otherwise be missed in shorter windows. As illustrated in Fig. 16, this expanded windowing approach provides better temporal context, reducing the number of false positives while maintaining sensitivity to meaningful deviations. By refining the time window strategy for discrete parameters, ATHMoS can achieve more reliable anomaly detection across a wider range of telemetry data types.

4.2.3 New Parameter Type: Counters and Pseudo-Counters

Initially, ATHMoS categorized most monotonous parameters with or without resets (e.g., counters) as continuous parameters due to their wide range of values within each time window. As a result, these parameters were smoothed and analysed using the standard feature set for continuous telemetry parameters. However, this approach led to misleading results, particularly at the edges of analysed timeframes, as shown in Fig. 17.

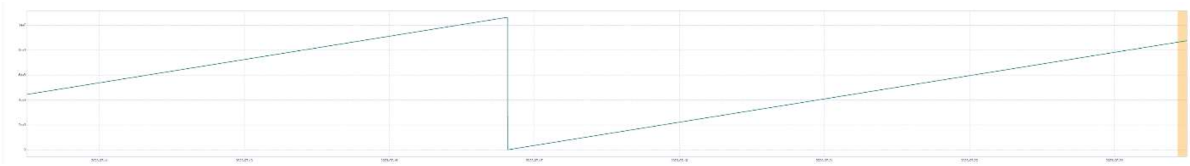


Fig. 17. Edge Effects for Counters

To address this issue, we introduced a dedicated parameter type for counters. Instead of smoothing the signal, we now extract features based on absolute minimal and maximal differences between two consecutive values, and constant duration intervals within each time window. This modification significantly reduces false positive detections for counters (Fig. 18) while preserving sensitivity to meaningful anomalies.

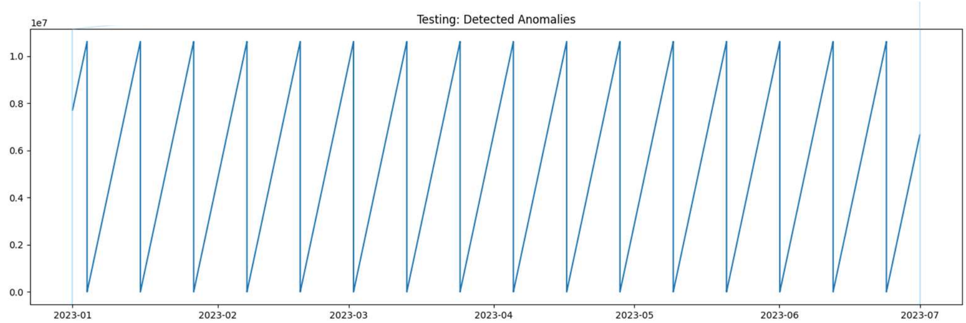


Fig. 18. Analysing Counters as a Separate Feature Type

To validate this approach, we injected synthetic anomalies into counter-like parameters, both with and without resets. As demonstrated in Fig. 19 and Fig. 20, this refined method successfully detects irregularities, including unexpected jumps and deviations in expected slopes.

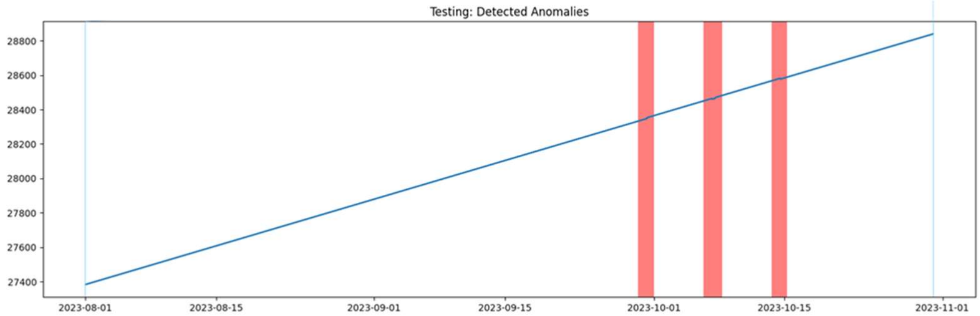


Fig. 19. Flagging Irregularities Using Counter Features

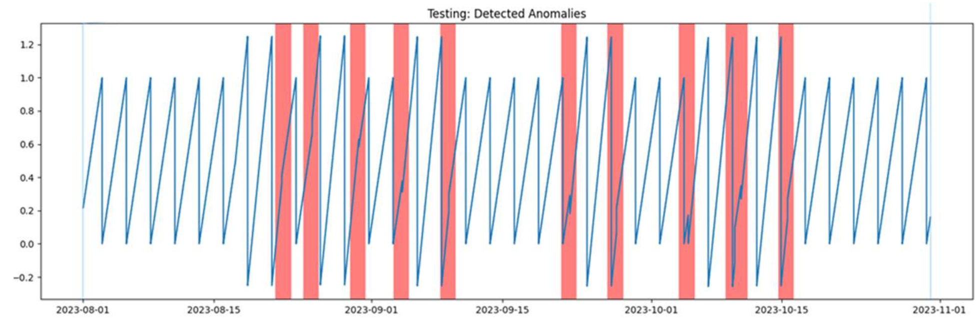


Fig. 20. Flagging Changes in the Slope Using Counter Features

A special class of parameters, which we refer to as pseudo-counters, posed a separate challenge. These include telemetry values such as fuel consumption, which should be non-decreasing but exhibit minor fluctuations due to sensor noise or measurement imprecision. The standard ATHMoS workflows for both continuous and discrete parameters resulted in excessive false-positive detections in operational use, see Fig. 21.

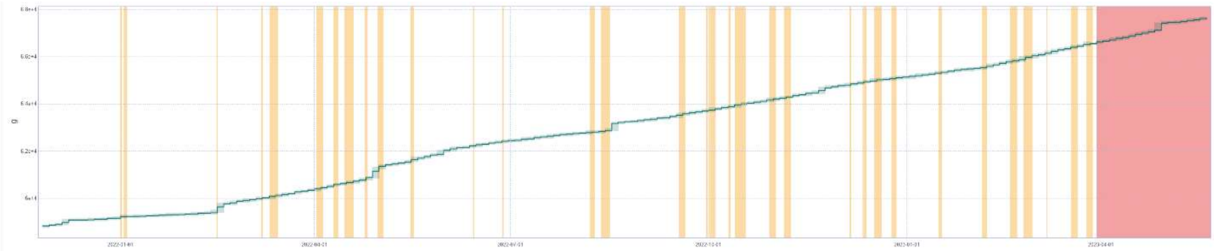


Fig. 21. Many False-Positive ATHMoS Detections for Pseudo-Counters

Inspired by the success of counter-specific features, we developed a relaxed feature set for pseudo-counters. Instead of using strict absolute values, we apply percentile-based thresholds to measure jump magnitudes and constant durations. This refinement reduces the number of false positives while still detecting true anomalies (Fig. 22).

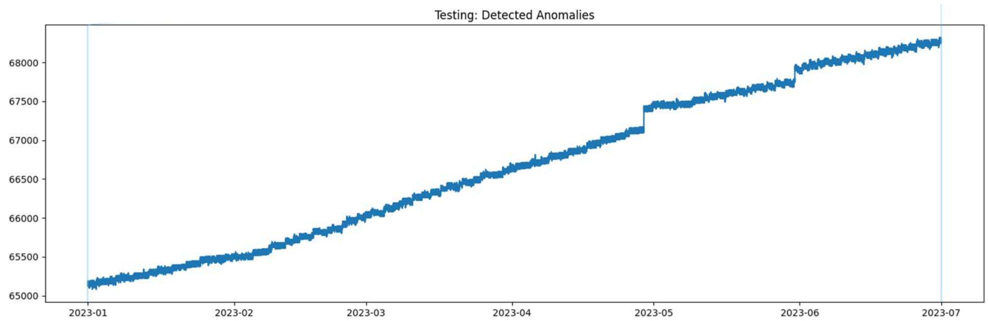


Fig. 22. Analysing Pseudo-Counters as a Separate Feature Type

18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.

To maintain the guiding principle of "one tool to rule them all," we extended the ATHMoS parameter classification algorithm. Now, instead of simply distinguishing between discrete and continuous parameters, ATHMoS can automatically classify counters and pseudo-counters as distinct parameter types as well. This enhancement ensures that each parameter type is analysed using the most suitable feature extraction techniques, further improving anomaly detection accuracy across all telemetry data.

4.3 Labelling by Users and Retraining Process

To enhance anomaly detection accuracy and minimize false positive detections, ATHMoS integrates direct user feedback via relabelling in ViDA, our front-end interface. This allows engineers to manually reclassify detected novelties as nominal when appropriate.

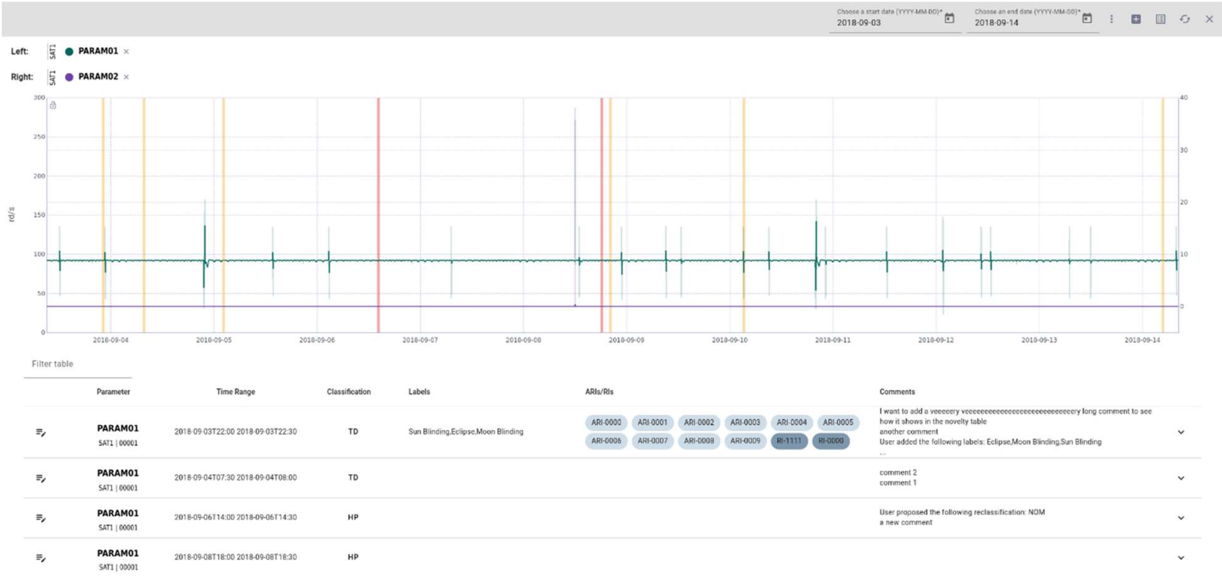


Fig. 23. Telemetry Time-Series Plot and Associated Novelty Table, Showing ATHMoS Classification, Assigned Labels, Comments and ARIs/RIs, as Examples

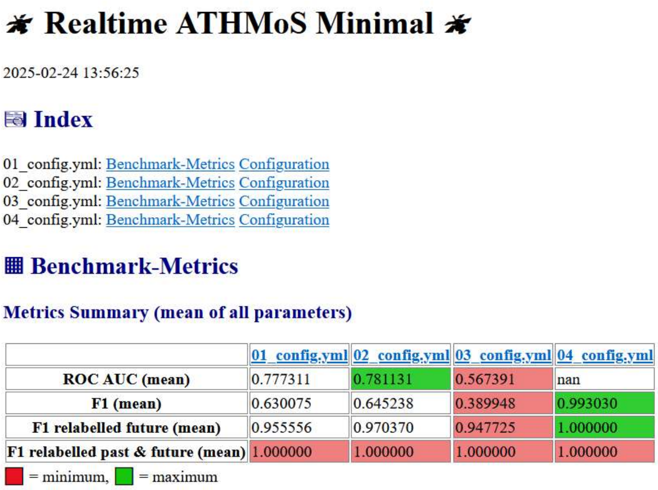
Once a user relabels a detection, the relabelled classification has to be manually verified. ATHMoS will then automatically incorporate the updated classification into the next retraining cycle, i.e., the relabelled data is included in the retraining of the short-term model during the following night and of the long-term model during the next weekend.

If a sufficient number of similar relabelling instances accumulate, ATHMoS is able to automatically classify future occurrences of the same behaviour as nominal. However, if only a limited number of instances exist or their rate of occurrence is very low, the model is not able to correctly label the new behaviour. To solve this problem, we are currently developing an automatic classification algorithm to compare new detections against previously relabelled cases. If the system determines a strong similarity, it will suppress the detection accordingly.

Beyond marking false positives as nominal, we plan to extend the relabelling functionality to allow users to classify nominal timeframes as anomalies. This ensures that such data is excluded from future training sets, preventing the models from learning undesired patterns. As with nominal relabelling, these classifications will automatically take effect in the next retraining cycle.

4.4 CI pipeline and Second Operational Workflow

To ensure the robustness and continuous improvement of ATHMoS, we have implemented a CI pipeline within GitLab. This pipeline allows users to configure a set of hyperparameters and upload sample telemetry data (both real and synthetic) to systematically test different configurations across various parameters. By automating these tests as well as generating informative reports, we can analyse the impact of changes in a controlled environment, making it easier to identify weaknesses, improve detection accuracy, and maintain system stability. Fig. 24 shows a sample summary for an example run for our real-time version of ATHMoS using four different hyperparameter configurations. The CI pipeline plays a crucial role in refining ATHMoS before deploying modifications into operational use.



Realtime ATHMoS Minimal

2025-02-24 13:56:25

Index

01_config.yml: Benchmark-Metrics Configuration
02_config.yml: Benchmark-Metrics Configuration
03_config.yml: Benchmark-Metrics Configuration
04_config.yml: Benchmark-Metrics Configuration

Benchmark-Metrics

Metrics Summary (mean of all parameters)

	01_config.yml	02_config.yml	03_config.yml	04_config.yml
ROC AUC (mean)	0.777311	0.781131	0.567391	nan
F1 (mean)	0.630075	0.645238	0.389948	0.993030
F1 relabelled future (mean)	0.955556	0.970370	0.947725	1.000000
F1 relabelled past & future (mean)	1.000000	1.000000	1.000000	1.000000

= minimum, = maximum

Fig. 24. Summary of the CI Pipeline Output (Screenshot)

In addition to the CI pipeline, we are planning a second operational workflow to facilitate smooth transitions when implementing larger changes. Following a blue-green deployment approach, this secondary workflow will be an identical copy of the current operational system, allowing us to apply and test modifications that have already been validated through the CI pipeline. By running both workflows in parallel, we can compare detection results across the full range of parameters under real operational conditions. If the updated workflow proves successful, we can seamlessly switch to it, using the previous version as a backup. Once the new workflow is considered stable, the backup workflow can then be used as a testbed for further improvements, ensuring an iterative and low-risk approach to deploying enhancements.

5. Conclusions and Future Work

ATHMoS has proven effective in detecting all relevant anomalies within the analysed timeframe, validating its utility as a robust anomaly detection tool for satellite telemetry data. To further improve its operational value, we are actively refining the system based on continuous feedback from our engineers. A key focus is reducing false positives—specifically, mathematically novel but operationally irrelevant detections—to ensure the results align more closely with our users' expectations. Two important next steps in this process are the extension of our automatic parameter classification algorithm (see Sec. 4.3) and the introduction of a second operational workflow tailored for nominal events such as manoeuvres and maintenance activities (see Sec. 4.4).

ATHMoS is now available to any mission at GSOC on request. As part of its broader operational deployment, we are currently in the process of setting up a workflow for our GFO mission, expanding the system's operational reach across mission types and operational contexts.

Looking ahead, we are also investigating the potential of real-time, onboard anomaly detection. A dedicated project is currently developing a proof-of-concept implementation of ATHMoS for the CAPTn-1 mission, with launch planned for 2026. The subsequent research phase in 2027 will focus on evaluating the feasibility and performance of onboard processing, laying the groundwork for autonomous anomaly detection in future satellite missions.

Acknowledgements

We would like to express our gratitude to our department head, Steffen Zimmermann, and our group lead, Tobias Göttfert, for their support throughout the development of ATHMoS. Additionally, we extend our thanks to all members of the DLR-MBT team for their valuable contributions and feedback.

Special thanks go to our working student Lena Kahle for her analysis of different smoothing algorithms and their influence on the corresponding novelty scores, which has significantly contributed to improving our data pre-processing approach.

References

[1] R. Chalapathy and S. Chawla, Deep Learning for Anomaly Detection: A Survey, https://doi.org/10.48550/arXiv.1901.03407, 2019

[2] P. Miralles et al., Machine Learning in Earth Observation Operations: A review, Proceedings of the 72nd International Astronautical Congress (IAC), Oct. 2021

[3] M. Zaroubian, P. Delande, P. Lambert, L. Coratger, P. Lefevre and C. Ouriachi, SESAM: An Experiment for AI-Based On-Board Satellite Monitoring, 17th International Conference on Space Operations, 2023

[4] F. Alesa et al., Spacecraft Failure Detection, Isolation and Recovery using Artificial Intelligence: A Study for Implementation on the Edge, 17th International Conference on Space Operations, 2023

[5] K. Hundman et al., Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Jul. 2018

[6] G. De Canio, K. Kotowski, and C. Haskamp, ESA Anomaly Dataset (1.0) [Data set], European Space Agency, https://doi.org/10.5281/zenodo.12528696, 2024

[7] K. Kotowski et al., European Space Agency Benchmark for Anomaly Detection in Satellite Telemetry, https://doi.org/10.48550/arXiv.2406.17826, 2024

[8] C. O'Meara, L. Schlag, L. Faltenbacher, and M. Wickler, ATHMoS: Automated Telemetry Health Monitoring System at GSOC using Outlier Detection and Supervised Machine Learning, Proceedings of the AIAA SpaceOps 2016 Conference, May 2016

[9] C. Schefels, L. Schlag, A. Del Moro, K. Helmsauer, T. Lesch, and T. Göttfert, Bringing a Machine Learning Based Novelty Detection Software Tool from Research to Production, 17th International Conference on Space Operations (SpaceOps 2023), Dubai, United Arab Emirates, 2023, 6–10 March

[10] A. Del Moro, M. Dauth, T. Lesch, C. Schefels, A. Braun, V. Filip, and T. Göttfert, A Modern Approach to Visualize Structured and Unstructured Space Missions' Data, 17th International Conference on Space Operations, Dubai (SpaceOps 2023), United Arab Emirates, 2023, 6–10 March

[11] N. Policella, J. Martinez, N. Salor Moral, S. Kotisis, K. Helmsauer, T. Göttfert, M. Cristoforetti, A. Gobbi, F. Hegwein and G. De Canio, Innovating Space Operations with AI: The AISHGO Project, ID #49, 18th Conference on Space Operations (SpaceOps 2025), Montreal, Canada, 2025, 26–31 May.

[12] A. Del Moro, T. Lesch, D. Leidreiter, C. Schefels, T. Göttfert and K. Helmsauer, Three (now four) years into operations for a satellite health monitoring system and its evolution, ID #272, 18th Conference on Space Operations (SpaceOps 2025), Montreal, Canada, 2025, 26–31 May