Prompt Anomaly Detection for Small Satellites in Low-Earth Orbit Constellations: A Machine Learning Approach

Weicheng Qian, Joshua DeJong, Bryan Cooke
Calian Advanced Technologies
18 Innovation Boulevard, Saskatoon, SK, Canada; +1(306)931-3425
winchell.qian@calian.com

ABSTRACT

Small satellites in Low Earth Orbit (LEO) constellations are shaping the future of communications and Earth observation. Despite having inherent advantages such as lower latency and faster deployments due to quicker and cheaper access to space, monitoring the large number of LEO satellites imposes a significant burden when considering the long-established methods of human-in-the-loop anomaly detection and recovery. Timely responses to anomalies reduce operational outages and help maximize the availability of the network. Traditional rule-based detection and supervised learning have inherent limitations in monitoring the large numbers of space assets projected to be launched over the next decades because more human intervention will be required to ensure anomaly detection. Although studies demonstrate the potential of unsupervised learning in detecting system anomalies using time-series telemetry data from space missions, such studies are limited to small telemetry datasets (*i.e.* less than 100 mnemonics). Because it is common for satellites to have many hundreds or thousands of mnemonics, adapting such Machine Learning (ML) models for time-series anomaly detection at a larger scale remains challenging. We examined the TelemAnom model developed by NASA using years of historical satellite telemetry data and proposed improvements to the TelemAnom model for scalable time-series anomaly detection on operational platforms. We evaluated the predictability of our adapted model on empirical telemetry data and compared model-identified anomalies with known anomalies identified by subject matter experts. Our finding suggests that factors such as satellite design, availability of data, and Exploratory Data Analysis (EDA) processes are important considerations when aligning unsupervised models with traditional time-series anomaly detection methods.

1 Introduction

The rapidly increasing number of small satellites being deployed in Low-Earth Orbit (LEO) and Very Low-Earth Orbit (VLEO) constellations addresses the urgent demand for convenient and cost-effective broadband communication[1,2] and high-resolution and near real-time (nRT) Earth observation (EO) tasks.[3] Meanwhile, the unprecedented number of satellites presents two major challenges for satellite operators. On one hand, the expected rate of anomalies increases with the number of satellites.[4] On the other hand, manual and rule-based monitoring approaches become insufficient to handle the volume and complexity of telemetry data generated by large numbers of satellite constellations.[5,6] Leveraging artificial intelligence (AI)-based methods can improve automation, accuracy, and scalability in detecting anomalies during satellite operations.[7–9]

Studies have demonstrated the potential of deep learning methods to improve the automation of time-series anomaly detection (TSAD).[10–17] However, these models are developed using synthetic or small telemetry datasets, which usually consist of fewer than 100 mnemonics over less than a month, resulting in fewer than $500,000$ telemetry values in total. In contrast, the telemetry data of an operational satellite may have a thousand or more mnemonics and average hundreds or thousands of records per second, rendering the results of these studies informative but not representative of real operations. Therefore, the applicability of existing methods and the remaining challenges in applying deep learning methods to improve TSAD automation for satellite constellations remain unclear and can be summarized into the following questions:

- Are the models scalable to the number of mnemonics and the time-series duration typically found in practical telemetry datasets?
- Are the models' outputs usable for satellite operators to reduce service latency and improve their productivity?
- Are there special characteristics within certain telemetry datasets that have not been ad-

dressed by existing models?

We examined operational telemetry data from the Near Earth Object Surveillance Satellite (NEOSSat) from 2017 to 2022 to investigate these questions. Additionally, we learned from satellite operators about the anomaly response process, particularly regarding the type of model output that could enhance their daily workflow. Prediction-error-based, reconstruction-error-based, and dissimilarity-based approaches are three actively studied deep learning approaches for TSAD.[9] Both prediction-error-based and reconstruction-error-based approaches train models on telemetry data that reflects the nominal status of a spacecraft, and then detects anomalies based on the residuals between the model-predicted healthy telemetry data and observed telemetry data collected from the spacecraft. A well-known prediction-error-based TSAD model leveraging long short-term memory (LSTM) for prediction, developed by Hundman *et al.*,[12] was selected and adapted for detecting anomalies in NEOSSat telemetry data. We also investigated dynamic time warping (DTW) for estimating correlations among mnemonics.[10] We evaluated the performance of the adapted model and identified gaps in existing models for serving anomaly detection needs in satellite operations as they grow in scale and complexity.

Our findings suggest that satellite telemetry data are typically of much higher volume, and having a much larger feature space, than the datasets employed to develop the existing TSAD models. Typical approaches like one-hot encoding and dynamic time warping (DTW) become less straightforward to employ due to the increased temporal and spatial dimensions of operational telemetry data. Although the theoretical classification of anomalies into point anomalies, contextual anomalies, and collective anomalies aids in solving the general problem of anomaly detection,[9,18] in practice, this classification loses efficacy when applied to satellite operations due to the inherent protocols for responding to anomalies in-orbit. Existing models that exhibit decent performance in detecting the presence of anomalies still have trouble categorizing anomalies into relevant operational categories necessary for a corresponding recovery. Finally, we identified several typical anomaly patterns that seem to break assumptions of existing dynamic error threshold functions for future work. We believe that a more realistic synthetic benchmark dataset could be valuable in guiding TSAD model development to be applicable for satellite-related TSAD.

The rest of the paper is structured as follows: In Section 2, we employed exploratory data analysis (EDA) to compare operational telemetry data from NEOSSat versus the Mars Science Laboratory (MSL) rover and the Soil Moisture Active Passive (SMAP) satellite dataset used by Hundman *et al.*;[12] In Section 3, we summarized the differences between the theoretical anomaly categories versus the operational anomaly categories, and we illustrated the reason why this improvement is more applicable to an operational environment; In Section 4, we proposed an approach to reduce the feature space and map detected anomalies into operational anomaly categories by leveraging the structure of the telemetry data; In Section 5, we evaluated our approach by benchmarking the adapted TelemAnom model against anomalies identified during operation. We summarized a typical anomaly pattern that seems to break the assumptions in existing dynamic error threshold functions in Section 6. Finally, we discussed the gaps in existing TSAD methods for practical satellite operations in Section 7 and identified future work that can further promote the utility of TSAD for satellite operations.

2 Comparing Telemetry Datasets

We compared two open-access telemetry datasets with the NEOSSat mission telemetry dataset (hereafter NEOSSat dataset) in both data volume and feature space. The NEOSSat dataset demonstrates a significantly larger data volume and feature space compared to the datasets employed in the development and testing of existing TSAD methods. The scale of the NEOSSat dataset reveals limitations in the scalability of existing methods. We explored the structure of the telemetry data with dynamic time warping (DTW) correlation[10] and discussed the limitations and mitigations of applying DTW correlation to large-scale operational datasets.

2.1 Data Volume and Feature Space

The volume of data and the feature space are critical factors for effective TSAD. Large volumes of informative data can provide a more comprehensive view of the satellite's operational condition, reflecting the operational status of satellite fleets and enabling more thorough maintenance processes to be leveraged. Meanwhile, these large data volumes challenge ML models when considering the prediction throughput and the models' capacity to effectively utilize the vast amount of information embedded in the data.

Informed by satellite engineers, we extracted a few thousand channels that are closely related to operationally relevant anomalies, which results in more than a twenty-fold increase in the number of channels compared with the SMAP and MSL dataset. The total number of anomaly occurrences is similarly larger with up to a five-fold increase. As shown in Table 1, we found that the increase in data volume mainly consists in the available length of the time-series and the number of statuses each categorical variable can have. In the SMAP and MSL dataset, after one-hot encoding, each channel depends on mostly fewer than thirty dichotomous variables. However, for an operational telemetry dataset, a single categorical variable can have more than two hundred statuses, which translate into more than two hundred one-hot encoded dichotomous variables, greatly increasing the size of the feature space.

We estimated the average channel-perplexity for each feature space to further reflect the differences of feature spaces of existing datasets and practical satellite telemetry data. The perplexity PP of a discrete probability distribution of p over the support of status code (encoded as a categorical variable) for a channel is equal to two raised to the power of the information entropy of p,

$$PP(p) := 2^{H(p)} \tag{1}$$

where $H(p)$ is the information entropy in bits with base 2,

$$H(p) := \sum_{x} -p(x) \log_2 p(x) \tag{2}$$

We use linear-scaled perplexity instead of log-scaled information entropy to make it easier for readers to compare values in the table. The exponentially larger feature space, as in Table 1, necessitates dimension reduction techniques and careful use of one-hot encoding.

Dynamic Time Warping (DTW) is a widely used distance metric for comparing two time series due to its capability of aligning sequences with non-linear variations.[10,19–21] However, the time complexity of DTW without constraint would be $O(MN)$ where M and N are the length of two time-series, which results in time complexity over 2e7 for two mnemonics readings of practical satellite telemetry data. Constraints like Sakoe-Chiba band and Itakura Parallelogram[22,23] can reduce the time complexity for long time-series at the cost of overestimating the distances. Contextual information can help derive the radius parameter but such a radius parameter becomes time-series specific and thus less generalizable.[24–26]

In summary, the increased volume in feature space and the length of time series have yet to be reflected in the existing available datasets and may pose challenges to the usability of current TSAD methods for this application.

Table 1: Data Volume Comparison

Sub-dataset Name	SMAP	MSL	NEOSSat
Number of Channels	55	27	few thousand
Number of Anomalies	69	36	few hundred
Average Length	~8,600 in total		several thousand per day
Compressed Dataset Size	82MB in total		hundreds of megabytes per day
Average Perplexity of Feature Space	1 per channel		3e79 per channel

2.2 *Dynamic Time Warping Correlation*

We adapted Rao *et al.*'s dynamic time warping correlation method[10] to investigate the patterns in readings of mnemonics of operational telemetry dataset. Because categorical variables with hundreds of status codes frequently appear in the operational telemetry dataset, instead of min-max normalizing variables to be in the range of $[-1, 1]$ and then evaluating the DTW distance matrix on the normalized values, we computed DTW distances for each pair of readings, then used the variance of distances for each pair of readings as the DTW distance matrix. We then use the transform function

$$\text{trans}(x) = \frac{1}{1 + \log(x+1)} \tag{3}$$

to map DTW distance from $[0, +\infty)$ to $[1, 0)$. The transform function has a fat-tail to ensure that differences between pairs with high DTW distance remain visible.

We computed the DTW correlation of mnemonics from an example packet by appplying DTW correlation transform function Equation (3) to the DTW distance matrix calculated with Sakoe-Chiba band set to 10. The DTW correlation matrix is illustrated as a heatmap, with darker rectangles suggesting that there exist groups of mnemonics whose time-series readings have smaller DTW distances. These groups of mnemonics possess potential to be used to predict each other, or to be combined to reduce dimensionality when predicting other mnemonics.

Figure 1: DTW Correlation of Example Mnemonics for $1,000$ readings

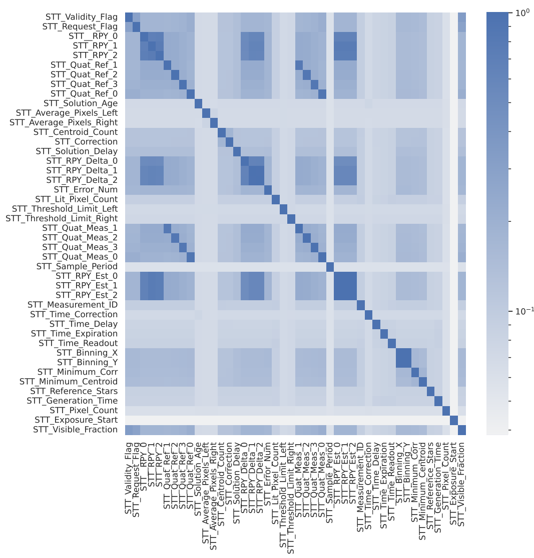

Figure 2: DTW Correlation of Example Mnemonics for $10,000$ readings

Despite its capability in aligning time series, the DTW approach is known to become computationally expensive with increases in time-series length.[24–26] We benchmarked the computational time of DTW-distance matrix for 164 mnemonics in parallel with 4 cores of AMD Ryzen Threadripper PRO 5955WX, with the Sakoe-Chiba band set to 10. Figure 3 suggests that the growth of computational time is superlinear with respect to the length of the time series. Applying DTW directly for the large data volume, such as several thousand records per day for each of few thousand mnemonics, of satellite telemetry data remains challenging.

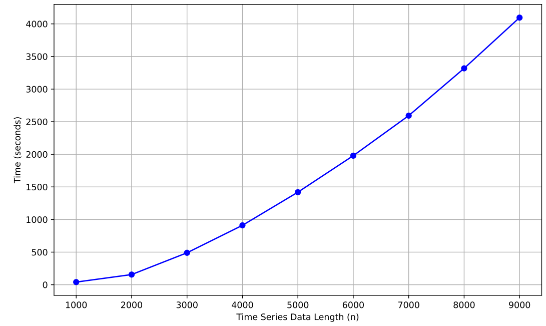

Figure 3: DTW-Distance Matrix Computational Time

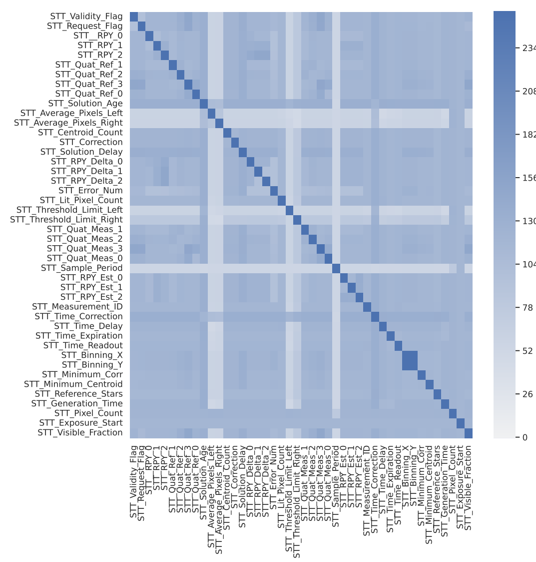

Figure 4: Relative Differences of DTW Distances of Example Mnemoncis Between $1,000$ and $10,000$ readings

We also investigated the variance of DTW correlation over the length of mnemonic readings and it seems that the DTW correlation of the length with $1,000$ readings (Figure 1) does not bear notable differences compared with the length with $10,000$ readings (Figure 2). We further investigated the relative differences of DTW distances, as shown in Figure 4. The overall similar color of the heatmap suggests that DTW correlation with a length of $1,000$ is similar to that with a length of $10,000$, except for an offset. This indicates that the pattern of DTW correlation can be relatively insensitive to the length of the time series. Since the time complexity of DTW increases with the length of the time series Figure 3, this finding suggests that the pattern of DTW correlation from a shorter period can be informative for a longer time period. We evaluated the computational time with Sakoe-Chiba band of radius 10 and it suggests that we cannot afford the time complexity to compute the DTW correlation for the entire time-series. Our finding suggests that we may consider using the DTW correlation with $1,000$ readings as an estimation while being aware that it may introduce biases.

Although the DTW correlation suggests potential groupings among mnemonics, we also observed that the darkened rectangles can appear for mnemonics from different missions (*i.e.* SMAP and MSL), as shown in the top-right and bottom-left corner in Figure 5. Such inter-spacecraft correlation becomes less intuitive, without knowing the meaning of these correlated mnemonics. Extra metrics in addition to DTW correlation can help further analyses.

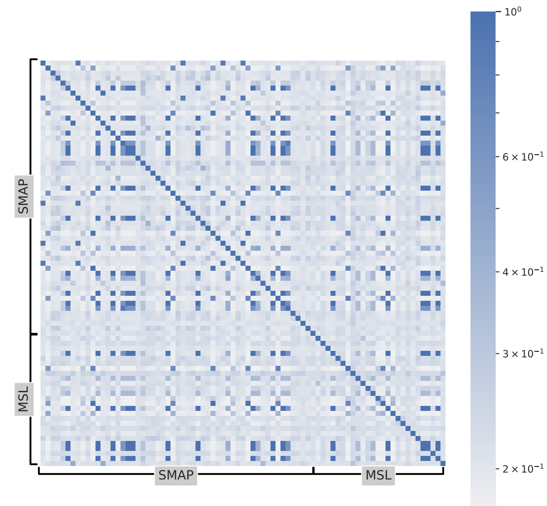

Figure 5: DTW Correlation of Mnemonics from SAMP and MSL

3 Response to Anomalies and Expected Model Output

Studies have categorized anomalies into point anomaly, contextual anomaly, and collective anomaly.[9,18,27,28] A point anomaly refers to individual data that appears anomalous relative to the rest of the data. A typical example of point anomaly is when a telemetry reading exceeds a constant threshold that defines the limit of healthy behavior. A contextual anomaly, also known as a conditional anomaly, refers to an instance where data appears anomalous given its context (or condition), often manifested as readings of related time-series variables. In time series, a contextual anomaly can occur when two events (hence their readings) happen in an unexpected order, even if both have healthy readings when inspected individually. A collective anomaly refers to a group of data that appear anomalous when considered together. For example, this could occur when a signal has readings above a threshold for a defined period of time.

The theoretical anomaly categories are abstract and summarize patterns of anomalies from the data perspective. Point anomalies, contextual anomalies, and collective anomalies may be tackled with different methods and techniques according to their characteristics. Existing TSAD models often focused on detecting the existence of anomalies while referring to these theoretical anomaly categories for performance analysis and algorithm development.[10,12–15,29]

However, for satellite operations applications, knowing the existence of anomalous data at a given time is insufficient. Satellite operators need to not only detect the existence of an anomaly but also to categorize the detected anomalies into operational categories. These operational anomaly categories are created to regulate response activities and track operational statistics. Different categories of anomalies may correspond to different response sequences and indicate varying severity levels. These categories often emphasize how to react to the anomaly rather than focusing on validating the theoretical patterns in the data.

4 TSAD into Operational Anomaly Categories

To the best of our knowledge, existing TSAD models[10,12–17,29] are focused on detecting existences of anomalies without further mapping the identified anomalies into operational anomaly categories. Rao *et al.*[10] evaluated the anomaly detection corresponding to each mnemonic. However, operational anomaly categories can be more complex than a one-to-one mapping to a single mnemonic as primary indicator. Furthermore, even for operational anomaly categories with a single indicator, it is unclear whether anomalies detected using a prediction-error-based approach with that mnemonic as the target can correspond to the operational anomaly category that uses the same mnemonic as the primary indicator.

We adapted the TelemAnom model and evaluated the performance of the adapted model in TSAD into operational anomaly categories for the NEOSSat mission. Compared with the SMAP and MSL datasets that provides paired channels of telemetry data and labeled anomalies, the adapted model needs to select useful channels from telemetry data for different operational anomaly categories. Only correctly detecting the existence of an anomaly and the binning into the corresponding operational anomaly category is considered as true positive.

5 Experiment

We trained the following models with 7 days of nominal (*i.e.*, non-anomalous) telemetry data and tested over 30 days of telemetry data. The test datasets include sampled anomaly days and randomly selected consecutive sequences of 30 days of nominal behavior from the five-year dataset.

5.1 Vanilla TelemAnom Model

The vanilla TelemAnom model is an LSTM model that takes an input of commands and monitors mnemonics, then predicts the expected readings of such a mnemonic if the underlying status of the system is normal. The differences between the LSTM estimated mnemonic readings under normal status and the actual mnemonic reading is then used to determine whether the system is currently experiencing anomalies. A high-level view of the Vanilla TelemAnom model can be summarized as:

$$\begin{aligned} \hat{\mathbf{y}} &= \mathrm{LSTM}\left(\mathbf{X}\right) \\ \mathbf{X} &:= [\mathbf{C}; \mathbf{y}]^\top \end{aligned} \tag{4}$$

where $\hat{\mathbf{y}}$ is the LSTM estimated mnemonic reading when the system has no anomalies, $\mathbf{y}$ is the actual reading of the mnemonic, $\mathbf{C}$ is defined as the matrix of command input to the spacecraft

$$\mathbf{C} := [\mathbf{c}^{(1)}; \mathbf{c}^{(2)}; \cdots; \mathbf{c}^{(m-1)}] \tag{5}$$

where $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \cdots, \mathbf{c}^{(m-1)}$ are one-hot encoded commands. Each $\mathbf{c}^{(i)}, i = 1, \cdots, m-1$ denote the active status of a unique command. For example, we may encode commands to operate a toy car by specifying "start the engine" as $\mathbf{c}^{(1)}$, "stop the engine" as $\mathbf{c}^{(2)}$, "accelerate" as $\mathbf{c}^{(3)}$, and "brake" as $\mathbf{c}^{(4)}$. Each bold symbol indicate a vector over the temporal span. For example the mnemonic reading vector $\mathbf{y}$ denotes

$$\mathbf{y} := [y_{t-l_s}, \cdots, y_{t-1}, y_t]^\top \tag{6}$$

where $y_{t-l_s}, \cdots, y_{t-1}, y_t$ each is a reading at a specific time.

The TelemAnom model further processes the prediction error

$$\mathbf{e} := \hat{\mathbf{y}} - \mathbf{y} \tag{7}$$

by smoothing it with Exponentially Weighted Moving Average (EWMA), then taken by dynamic error threshold function (DET) to detect ranges with associated anomaly scores by comparing z-score with a dynamically adjusted threshold. The threshold is aimed to minimize the mean and standard deviation of periods of readings that are considered healthy while being regularized by the penalty of the length of anomaly periods and number of anomaly periods.

5.2 Adapted TelemAnom Model

The essence of the maximum likelihood approach is to use other mnemonics related to the target mnemonics as an maximum likelihood estimator (MLE) of the commands sent to the spacecraft. That is, we altered the model as

$$\begin{aligned} \hat{\mathbf{y}}^{(i)} &= \mathrm{LSTM}\left(\hat{\mathbf{X}}^{(i)}\right) \\ \hat{\mathbf{X}}^{(i)} &:= [\mathbf{Y}_{\setminus \mathbf{y}^{(i)}}; \mathbf{y}^{(i)}]^\top \end{aligned} \tag{8}$$

where $\mathbf{Y}_{\setminus \mathbf{y}^{(i)}}$ is a matrix of other mnemonic vectors served as an MLE of one-hot encoded $\mathbf{C}$ commands.

$$\mathbf{Y}_{\setminus \mathbf{y}^{(i)}} := [\mathbf{y}^{(1)}; \cdots; \mathbf{y}^{(i-1)}; \mathbf{y}^{(i+1)}; \cdots; \mathbf{y}^{(m)}]^\top \tag{9}$$

where $\mathbf{y}^{(1)}; \cdots; \mathbf{y}^{(i-1)}; \mathbf{y}^{(i+1)}; \cdots; \mathbf{y}^{(m)}$ are mnemonics other than $\mathbf{y}^{(i)}$. An easy choice is to use all the rest of the mnemonics within the same packet, a structural unit of transmission that groups several mnemonics together.

5.3 Supervised LSTM Model

The TelemAnom model and our adaptation are unsupervised and only trained with datasets that define healthy telemetry data. We created a supervised LSTM model trained with labeled anomalies to serve as an upper bound, addressing the question of what performance an ML model can achieve for a well-known anomaly category when given the extra information of labeled data for historical occurrences of anomalies.

$$\mathbf{A} = \mathrm{LSTM}\left(\mathbf{X}\right) \tag{10}$$

where $\mathbf{X}$ is a matrix of all mnemonic vectors. All nominal variables are processed by one-hot encoding, and the remaining numerical variables are scaled in range $[-1, 1]$. The output $\mathbf{A}$ is a vector of dichotomous value -1 (healthy) and 1 (anomaly) to indicate whether the system has anomalies at the corresponding time.

6 Results

We compared the performance of our adapted TelemAnom model and the supervised LSTM model, where both models are evaluated with respect to operational anomaly categories for 30 days of healthy telemetry data, and several days containing anomalous signatures depending on the actual occurrences of each operational anomaly category (OAC). As shown in Table 2, the unsupervised adapted TelemAnom model has lower F1 scores compared with the supervised LSTM model. The F1 score of the adapted TelemAnom model for the practical telemetry dataset is also lower than the SMAP (0.793) and MSL dataset (0.855). Besides the extra challenge of detecting the operational anomaly category and the existence of anomalies, we found that the dynamic error threshold function (DET) seems to be one of the bottlenecks because its assumption does not expect long durations of anomaly occurrences, which can happen in practice.

Table 2: F1 Score for Operational Anomaly Categories

Operational Anomaly Category (OAC)	F1 Score (Adapted TelemAnom)	F1 Score (Supervised LSTM)
OAC-1	0.19	1.0
OAC-2	0.34	0.875
OAC-3	0.25	0.605

We performed error analyses, particularly for the false negatives when the model failed to detect existing anomalies. We plotted the LSTM model predicted healthy telemetry data $\hat{\mathbf{y}}$ (denoted as the blue line that partially overlapped with the orange line), the observed telemetry data (denoted as the orange line), and the residual between the predicted healthy telemetry data and the observed telemetry data (denoted as the green line) in Figure 6 to illustrate a typical false negative case of a long period with an anomalous signature. Specifically, this particular anomaly exists between the x-axis values of 750 and 1,500. Despite the clearly evident residual (the green error line) between predicted healthy telemetry data and observed telemetry data, this anomaly is overlooked by the DET algorithm due to its assumption that anomalous periods are usually shorter than healthy periods. While the assumption expecting shorter anomaly periods is reasonable in general, it can be violated during specific time intervals depending on the resolution of time, particularly in the context of satellite operations where anomalous scenarios may persist until they are proactively resolved.

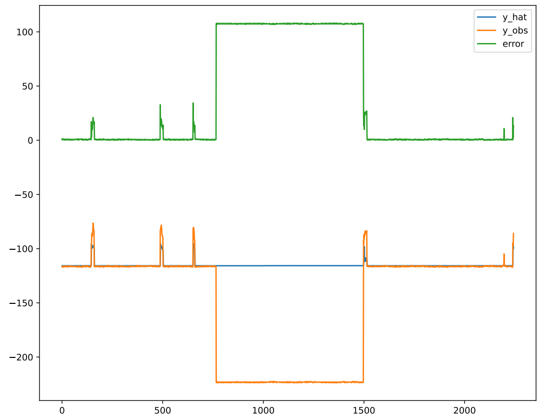

Figure 6: Example of DET False Negative

In summary, the adapted TelemAnom model, trained with healthy telemetry data, demonstrated its potential to detect anomalies and categorize them into bins applicable to an operational environment. Its performance can be further improved by enhancing the ability to identify anomalies from the residual of the model predicted healthy telemetry data compared to the observed telemetry data.

7 Discussion

The rapidly increasing number of satellites in Low Earth Orbit (LEO) and Very Low Earth Orbit (VLEO) constellations presents a challenge to satellite operators to efficiently and effectively manage responses to anomalies. The surge in satellite deployments not only generates a larger volume of telemetry data but also necessitates the creation of satellite-specific rules for the evaluation of telemetry data. These challenges demand robust and scalable solutions to maintain the operational integrity of orbiting assets.

Time-series anomaly detection (TSAD) models can greatly improve satellite operation teams' capacity to process large volumes of telemetry data while managing the fleet efficiently. Supervised learning models typically require training with labeled anomalies, and since rules for healthy telemetry data can be satellite-specific, the labeling process can be cumbersome and may not be adequately prepared for emerging anomalies. Therefore, it is ideal to develop models that do not rely on labeled anomalies, enabling more adaptive and efficient anomaly detection in dynamic satellite environments.

Theoretical anomaly categories concisely depict anomaly patterns as point anomalies, contextual anomalies, and collective anomalies. These anomaly patterns indicate the presence of non-nominal behaviour but become less informative for satellite operators responses. Operational anomaly categories based on response protocols and severity levels are more helpful for an environment that requires actionable steps in a rapid fashion. TSAD models that map detected anomalies into defined anomaly signature categories can significantly reduce satellite operations latency in responding to anomalies.

Prediction-error-based TSAD models demonstrated the potential to bin detected anomalies into operational anomaly categories by training only with datasets defining healthy telemetry data. Because these models do not require to be trained with labeled anomalies, they can be scaled more easily and are, therefore, more applicable when considering the expected surge in future satellite deployments. The existing prediction-error-based TSAD models are developed with small volume datasets and feature space. The output of existing models focuses on the existence of anomalies rather than the operational anomaly categories. To the best of our knowledge, we are among the first to report the adaptation and evaluation of prediction-error-based TSAD models with operational satellite telemetry datasets to identify and categorize anomalies for direct and actionable responses in the operational environment.

The existing prediction-error approach relies on prediction models to predict healthy telemetry data given the historical telemetry data, and an error detection function to identify anomalies from the residual of the predicted healthy telemetry data and the observed telemetry data. The prediction models need to account for the large feature space and compensate for noise in the telemetry data. The grouping patterns among mnemonics shown in Figure 1 can inform dimensionality reduction in feature space. The insensitivity of these grouping patterns, such as Figure 4, suggests opportunities to employ computationally expensive algorithms over several small sampled periods. Our finding suggests that the error detection function can be one of the bottlenecks for the prediction-error-based methods, as shown in Figure 6, due to the assumptions currently employed and the complexity of the variable time resolution and characteristics of anomalies in time series.

Our findings are limited to the dataset from the NEOSSat mission and may be subject to change given extra datasets from other missions. Space missions have varied purposes and operating conditions. Observed behaviour can be different than expected as can the identification, investigation and classification of anomalous signatures. However, common spacecraft components (such as attitude control, navigation, commanding and telemetry chains) can still benefit from a generic TSAD solution since the principles do not change greatly, reducing the level of effort to arrive at a mission-specific solution. This is particularly useful with large fleets of the same (or related) design, where fleet-common signatures can be shared alongside spacecraft-specific behaviours, simplifying management of the entire fleet.

TSAD with satellite telemetry data can be informed by the satellite design and the implementation of the various operations concepts. Establishing a systematic approach to enable models to learn from these inputs and adapt to a specific mission-type can be another critical consideration for TSAD models. This approach might involve the integration of satellite design, manufacturing, and management to arrive at a more holistic solution. By incorporating knowledge from all stages of a satellite's lifecycle, TSAD models can be better adapted to the unique requirements and behavior of the missions in question, leading to more accurate anomaly detection and more effective operational responses.

8 Conclusion

We examined operational telemetry data from the NEOSSat mission and compared with other open-access datasets. Using the NEOSSat dataset, we adapted and evaluated prediction-error-based TSAD models on detecting anomalies into operational anomaly categories. Our findings suggest that the large volume and feature space of operational telemetry data can be challenging for existing TSAD models. Existing TSAD models can be more informative for practical satellite operation by outputting operational anomaly categories instead of the presence of anomalies alone. Unsupervised TSAD models can be a scalable solution to serve the surge in satellite deployment but their performance needs to be evaluated and improved with practical, operationally representative telemetry data.

Acknowledgments

We thank Hanwen Zheng, Francis Senay, Ayman Shahriar, and Jarin Tasnim, for their contribution to the project. This project is undertaken with the financial support of the Canadian Space Agency.

References

[1] Iñigo Del Portillo, Sydney I. Dolan, B. Cameron, and E. Crawley. Architectural decisions for communications satellite constellations to maintain profitability while serving uncovered and underserved communities. *International Journal of Satellite Communications and Networking*, 41:82 – 97, 2022.

[2] Israel Leyva-Mayorga, B. Soret, and P. Popovski. Inter-plane inter-satellite connectivity in dense leo constellations. *IEEE Transactions on Wireless Communications*, 20:3430–3443, 2020.

[3] Mingsong Lv, Xuemei Peng, Wenjing Xie, and Nan Guan. Task allocation for real-time earth observation service with leo satellites. *2022 IEEE Real-Time Systems Symposium (RTSS)*, pages 14–26, 2022.

[4] J. D. Esary and F. Proschan. Relationship between system failure rate and component failure rates. *Technometrics*, 5:183–189, 1963.

[5] Kedar R. Naik, Andrew Holmgren, and J.C. Kenworthy. Using machine learning to automatically detect anomalous spacecraft behavior from telemetry data. *2020 IEEE Aerospace Conference*, pages 1–14, 2020.

[6] Sara Abdelghafar, A. Darwish, A. Hassanien, Mohamed Yahia, and A. Zaghrout. Anomaly detection of satellite telemetry based on optimized extreme learning machine. *Journal of Space Safety Engineering*, 6:291–298, 2019.

[7] Fares Fourati and Mohamed-Slim Alouini. Artificial intelligence for satellite communication: A review. *Intelligent and Converged Networks*, 2(3):213–243, 2021.

[8] Kathiravan Thangavel, Dario Spiller, R. Sabatini, S. Amici, N. Longépé, P. Servidia, Pier Marzocca, Haytham Fayek, and L. Ansalone. Trusted autonomous operations of distributed satellite systems using optical sensors. *Sensors (Basel, Switzerland)*, 23, 2023.

[9] Kukjin Choi, Jihun Yi, Changhwa Park, and Sungroh Yoon. Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines. *IEEE Access*, 9:120043–120065, 2021. Conference Name: IEEE Access.

[10] Jubilee Prasad Rao, John Pace, Jesse Williams, Ryan Mackey, Liang He, and Carlo Menegazzo. A Reusable Framework for Fault Detection and Isolation in Small Satellites. *Proceedings of the Small Satellite Conference*, Technical Poster Session 1, 2023.

[11] Zefan Zeng, Jiugang Lei, Guang Jin, Chi Xu, and Lu Zhang. Detecting Anomalies in Satellite Telemetry Data based on Causal Multivariate Temporal Convolutional Network. In *2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAI)*, pages 63–74, July 2022.

[12] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 387–395, New York, NY, USA, July 2018. Association for Computing Machinery.

[13] Dawei Pan, Zhe Song, Longqiang Nie, and Benkuan Wang. Satellite Telemetry Data Anomaly Detection Using Bi-LSTM Prediction Based Model. In *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6, May 2020. ISSN: 2642-2077.

[14] Zefan Zeng, Guang Jin, Chi Xu, Siya Chen, Zhelong Zeng, and Lu Zhang. Satellite Telemetry Data Anomaly Detection Using Causal Network and Feature-Attention-Based LSTM. *IEEE Transactions on Instrumentation and Measurement*, 71:1–21, 2022. Conference Name: IEEE Transactions on Instrumentation and Measurement.

[15] Shreshth Tuli, Giuliano Casale, and Nicholas R. Jennings. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data, May 2022. arXiv:2201.07284 [cs].

[16] Xuanhao Chen, Liwei Deng, Feiteng Huang, Chengwei Zhang, Zongquan Zhang, Yan Zhao, and Kai Zheng. DAEMON: Unsupervised Anomaly Detection and Interpretation for Multivariate Time Series. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 2225–2230, April 2021. ISSN: 2375-026X.

[17] Yifan Li, Xiaoyan Peng, Jia Zhang, Zhiyong Li, and Ming Wen. DCT-GAN: Dilated Convolutional Transformer-Based GAN for Time Series Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3632–3644, April 2023. Conference Name: IEEE Transactions on Knowledge and Data Engineering.

[18] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):15:1–15:58, July 2009.

[19] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11:561–580, 2007.

[20] Borui Cai, Guangyan Huang, N. Samadiani, Guanghui Li, and Chi-Hung Chi. Efficient time series clustering by minimizing dynamic time warping utilization. *IEEE Access*, 9:46589–46599, 2021.

[21] Eamonn J. Keogh. Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7:358–386, 2002.

[22] Z. Geler, V. Kurbalija, M. Ivanović, Miloš Radovanović, and Weihui Dai. Dynamic time warping: Itakura vs sakoe-chiba. *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6, 2019.

[23] Sri Harsha Dumpala, KNRK Raju Alluri, Suryakanth V Gangashetty, and Anil Kumar Vuppala. Analysis of constraints on segmental dtw for the task of query-by-example spoken term detection. In *2015 annual IEEE India conference (INDICON)*, pages 1–6. IEEE, 2015.

[24] V. Kurbalija, Miloš Radovanović, Z. Geler, and M. Ivanović. The influence of global constraints on dtw and lcs similarity measures for time-series databases. *Third International Conference on Software, Services & Semantic Technologies S3T 2011*, pages 67–74, 2011.

[25] V. Niennattrakul and C. Ratanamahatana. Learning dtw global constraint for time series classification. *ArXiv*, abs/0903.0041, 2009.

[26] R. M. Martins and A. Kerren. Efficient dynamic time warping for big data streams. *2018 IEEE International Conference on Big Data (Big Data)*, pages 2924–2929, 2018.

[27] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep Learning for Anomaly Detection: A Review. *ACM Computing Surveys*, 54(2):38:1–38:38, March 2021.

[28] Zahra Zamanzadeh Darban, Geoffrey I. Webb, Shirui Pan, Charu C. Aggarwal, and Mahsa Salehi. Deep Learning for Time Series Anomaly Detection: A Survey, December 2022. arXiv:2211.05244 [cs].

[29] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, January 2016. arXiv:1511.06434 [cs].