ANTELOPE: TOWARDS ON-BOARD ANOMALY DETECTION IN TELEMETRY DATA USING DEEP LEARNING

Jakub Nalepa[1,2], **Michal Myller**[1], **Pawel Benecki**[1,2], **Jacek Andrzejewski**[1], **and Daniel Kostrzewa**[1,2]

[1]*KP Labs, Konarskiego 18C, Gliwice, Poland*
[2]*Silesian University of Technology, Akademicka 16, Gliwice, Poland*
jnalepa@ieee.org

ABSTRACT

Detecting anomalies in telemetry data captured on-board a spacecraft is a critical aspect of its safe operation, and it allows us to effectively and timely respond to failures and hazards. There exist three main types of anomalies that should be considered for such complex missions. In point anomalies, telemetry values fall outside the nominal operational range. The collective anomalies refer to the overall sequences of consecutive telemetry values that are anomalous (a single data point does not necessarily manifest an anomaly), whereas in contextual anomalies, the single values are anomalous within their local neighborhood. We present how deep learning can be employed to this task, where recurrent neural networks act as signal predictors. Then, the predicted signal is compared with the actual telemetry, and—based on the difference between them—a particular time point within a signal can be annotated as an anomalous event. This approach is being implemented as part of Antelope—our on-board computer with predictive maintenance capabilities.

Key words: anomaly detection; deep learning; recurrent neural networks; on-board telemetry processing.

1. INTRODUCTION

Detecting anomalies on-board satellites plays a critical role in ensuring their safe operations [1, 2, 3]. There have been various approaches for automating the process of detecting anomalies from telemetry data [4, 5, 6, 7]. The basic yet widely exploited algorithms include the out-of-limit techniques that are built upon the assumption that we have the prior expert knowledge allowing us to exploit a rule-based approach for detecting unexpected events [7]. There are machine learning algorithms for this task, including extreme learning machines [8], denoising autoencoders [9], generative adversarial networks [10], and various tensor-based systems [11], but they are commonly heavily parameterized and require large amounts of ground-truth (manually delineated) data, ideally with captured anomalies [12]. Since acquiring such data is infeasible in practice, unsupervised techniques have attracted the research attention, as they do not require having large training samples to train well-generalizing models [13].

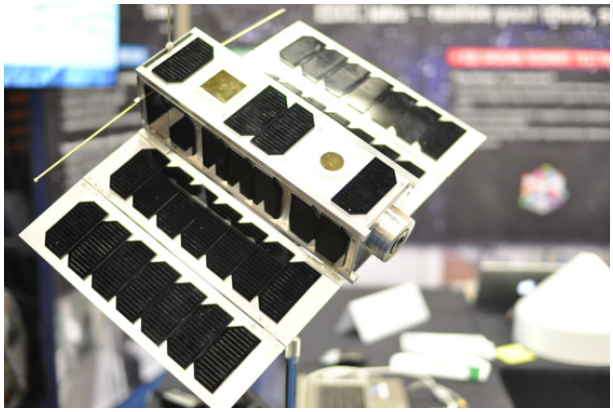


Figure 1. The OPS-SAT satellite can be an excellent source of real-life telemetry data that can be used for both verifying the unsupervised techniques for detecting anomalous events from such data, but also for developing new algorithms for this task. This figure comes from the ESA's OPS-SAT webpage: `https://www.esa.int/Enabling_Support/Operations/OPS-SAT`.

In this paper, we present our approach for this task—being developed as part of our Antelope on-board computer with predictive maintenance capabilities—which exploits recurrent neural networks. We are currently building upon Telemanom [13], and are utilizing long short-term memory networks to model the expected telemetry signal. Such prediction models can be trained from a set of the simulated nominal telemetry signals (e.g., using the software or hardware-in-the-loop simulators), or from a set of real-life telemetry presenting the nominal operation. Importantly, we can learn from the correct examples that do not contain anomalous events—it allows us to abstract from the type of anomalies that we want to target. Once the expected signal is elaborated, it is confronted with the actual one, and the obtained error triggers the alert showing that the anomaly has appeared. We additionally show how to verify the anomaly detection techniques in a quantitative way, and what kind of metrics reflect the underlying abilities of such deep learning techniques. Finally, we present our visualization tools that help us better understand the advantages and shortcomings of various anomaly detection methods

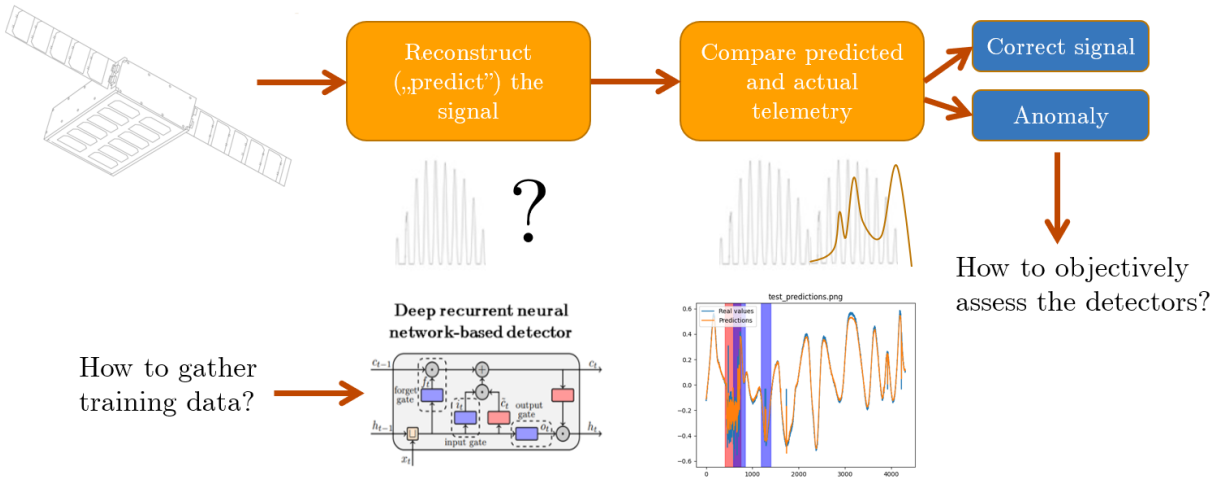


Figure 2. A high-level flowchart of our current approaches towards detecting anomalies from the telemetry data.

and will discuss our experiments performed over benchmark one-dimensional signal, and real-life telemetry captured on-board the European Space Agency's OPS-SAT satellite (Figure 1). Since the Antelope will be exploited on-board a satellite, our resource-frugal models will ultimately help us respond to the events quicker and could be used to reduce the amount of data to transfer back to Earth through annotating the most important parts of the signal that enable further analysis and interpretation.

This paper is structured as follows. Our approach for automatic anomaly detection from telemetry data is discussed in Section 2. In this section, we also present our Antelope Toolbox that be conveniently used for simulating telemetry data, and for performing quantitative and qualitative analysis of the existing and emerging event detection techniques. In Section 3, we present our initial experiments over the real OPS-SAT telemetry data. Finally, Section 4 concludes the paper.

2. TOWARDS ON-BOARD ANOMALY DETECTION USING DEEP LEARNING

In our current approach towards detecting anomalies from on-board telemetry, we exploit recurrent neural networks (RNNs)—more specifically, long short-term memory unit-based (LSTM) RNNS—to predict (reconstruct) the telemetry signal based on the historical data (the high-level flowchart is rendered in Figure 2). Such predicted signal may be easily compared with the actual telemetry, and the reconstruction error can be calculated as e.g., the mean square error within a considered time window. Once this error exceeds an assumed threshold, we can conclude that the signal has started becoming "anomalous", as it significantly deviates from the expected telemetry. This thresholding has to be, however, fine-tuned—having a too small threshold could easily lead to numerous false-positive (FP) errors (i.e., correct events that have been erroneously annotated as anomalies) that would negatively impact the practical utility of the automated anomaly detection tool, as numerous FPs would have to be filtered by the Operations Team. On the other hand, too large thresholds could increase the probability of omitting anomalous events, and incorrectly treating them as a correct operation of the spacecraft.

To train the reconstruction part of the pipeline (i.e., an LSTM network), we can exploit the available benchmark, real-life or simulated time-series data. There indeed exists a manually-annotated ground-truth dataset (referred to as *Telemanom* [13], and available at `https://github.com/khundman/telemanom`). It includes spacecraft anomaly data and experiments from the Mars Science Laboratory and SMAP missions. Such time-series data is divided into training and test subparts, with the latter containing the ground-truth anomalous events. Although this dataset might be treated as a solid start for experimentation, there are known issues related to the important aspects of Telemanom, including its normalization (for details, see `https://github.com/khundman/telemanom/issues/23`; last date of accessing the webpage: June 12, 2021). Finally, it is important to note that various quality metrics are commonly used in different anomaly detection papers (there also exist custom quality metrics [12]), hence confronting state-of-the-art techniques is not trivial.

Unfortunately, capturing real-life telemetry and elaborating the corresponding ground-truth manual annotations is a challenging task due to the large amount of telemetry data that would have to be transferred to the ground stations. Additionally, the process of generating ground-truth is time-consuming, very user-dependent and prone to human errors, and incorrect or noisy ground-truth data can easily deteriorate the generalization performance of supervised learners. such data can (and should) be, however, utilized for verifying and validating detection engines—an example set of OPS-SAT telemetry channels is presented in Figure 3. We can appreciate that it is possible to conveniently capture different signals that may

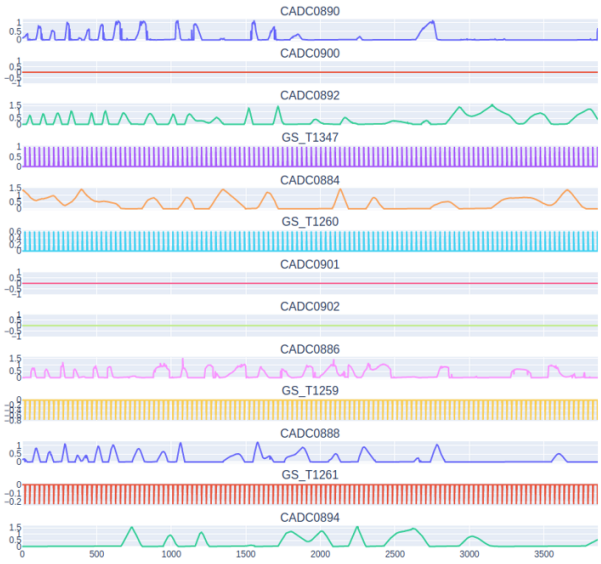


Figure 3. A set of example OPS-SAT telemetry channels

manifest anomalous behavior of the spacecraft.

To address the issue of lacking ground-truth anomaly datasets (and the problem of their varying quality in numerous regards), we can exploit simulated time-series data. Here, we can benefit from building a "digital twin" of the considered satellite in order to generate simulated telemetry (both nominal, without any anomalous events, and with events). As we can precisely simulate events of specific characteristics that will happen in a known time point, such simulations can bring us accurate and high-quality ground truth.

In Figure 4, we render an example simulated telemetry channel in our Antelope Toolbox, alongside an annotated anomalous event. The nominal telemetry may be therefore used for training the predictors (LSTM networks), whereas the channels with anomalous events can be utilized for verification of the detection process—similarly, we can perform this investigation for any benchmark or real-life time-series data.

To thoroughly understand the operational performance of any machine learning model, we commonly perform quantitative, qualitative and statistical analysis that shed more light on the capabilities of such learners. In the context of anomaly detection from telemetry, we can calculate various overlap metrics (see an example in Figure 5, in which we can utilize the Antelope Toolbox for qualitative analysis of ground-truth and automatically determined anomalous events), such as the DICE coefficient or the Jaccard's index, if the corresponding ground truth exists, alongside numerous measures based on the confusion matrix. These metrics may not, however, necessarily reflect the practical utility of event detectors, as in practical scenarios we may want to detect (or even predict) an anomaly as fast as possible (ideally before it has started happening, but in a reasonable time window). It could give us more time to respond to such situations, and to take appropriate actions, e.g., to save the spacecraft through switching it into its safe operation mode.

3. DETECTING ANOMALIES IN THE OPS-SAT TELEMETRY

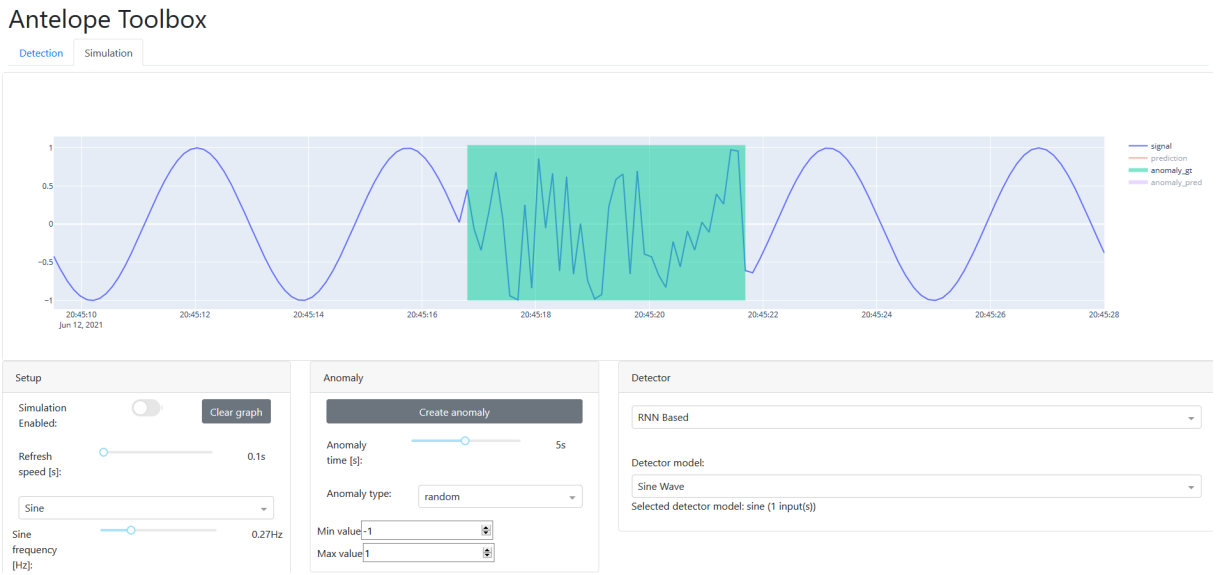
Although we do not have the ground-truth OPS-SAT data with annotated anomalies, we can exploit the existent and simulated periodic time series to train the reconstructing RNNs, and deploy such models over real OPS-SAT telemetry. In this scenario, calculating quantitative detection quality metrics is impossible (we can, however, manually annotate potential "anomalous events" in the process of visual investigation).

Figure 6 renders example OPS-SAT telemetry channels that were used for training and verifying our LSTMs and event detectors. We can appreciate that the recurrent networks are able to accurately reconstruct both training and test samples (hence generalize well over the unseen time-series data)—the orange signals present our predictions, whereas the actual telemetry is visualized in blue. Although we were not validating the detected anomalies with the OPS-SAT Operations Team, the quantitative analysis indicates that there might be some "suspicious" part of the signals that could correspond to the on-board events (we have annotated them in blue, in the process of manual investigation). In Figure 6, we can observe that the LSTM detectors, followed by the thresholding-based detection were able to capture such variations in the initial part of the test signal. Finally, it is interesting to note that such unsupervised detectors can be ultimately used for making the process of creating new ground-truth (annotated) samples easier. Here, such detectors could be used for elaborating candidate anomalous events within the existent telemetry data which would be later verified by the experienced Operations Team, in order to prune incorrectly annotated FPs. It could significantly decrease the time of creating new ground-truth datasets, and make the entire procedure less tedious and semi-automatic.

4. CONCLUSIONS

In this paper, we discussed our approach for detecting anomalies from the satellite telemetry data using recurrent neural networks followed by the thresholding-based error analysis. Since the number of existing time-series datasets that could be used for training and verifying event detection techniques is extremely limited, simulating such data—hence building a "digital twin" of a spacecraft—is an inevitable step towards fast adoption of data-driven on-board anomaly detection in Fault Detection, Isolation and Restoration systems. We also discussed the Antelope Toolbox—our tool for simulating time-series data alongside anomalous events of various characteristics, and for performing quantitative and qualitative analysis of existing and emerging event detection systems. We believe that bringing artificial

(a) Simulating a period time-series data using the Antelope Toolbox



(b) Loading an existing time-series data into the Antelope Toolbox for further analysis

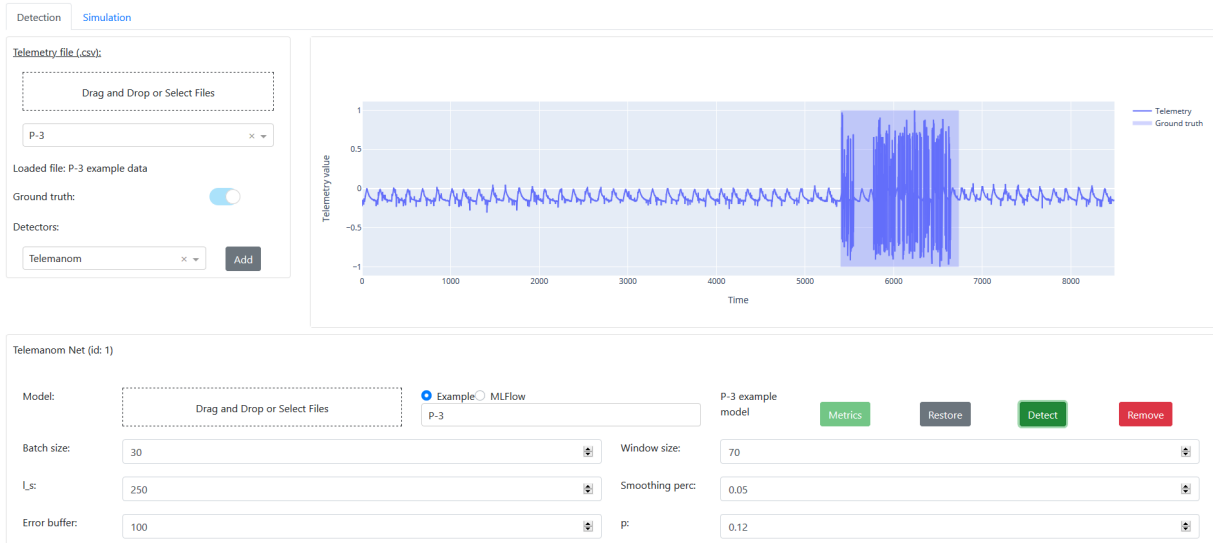


Figure 4. In the Antelope Toolbox, we can simulate not only nominal periodic time-series data, but also various anomalies of different characteristics. The green area in the upper plot (a) indicates the ground-truth anomaly within a periodic simulated channel, whereas the simulated telemetry is presented in blue. Additionally, we can load an existing time-series data (b) for further investigation (e.g., for verifying pre-trained anomaly detectors over benchmark or real data).

intelligence-based anomaly detection on-board the spacecrafts is an important step towards data-driven Fault Detection, Isolation and Restoration systems that will ultimately improve the safety of emerging satellites, and reduce their operational costs through delivering clear insights into the behavior of the entire spacecraft and its pivotal (sub)systems. To make this step possible, we need not only accurate, well-generalizing, and robust (e.g., against various noise distributions [14, 15]) algorithms, but we also have to be able to deploy them in hardware-constrained execution environments [16]. Designing, implementing, verifying, and validating such resource-frugal learners constitute our current research and development efforts.

ACKNOWLEDGMENTS

We would like to thank the OPS-SAT Flight Control Team for all the support.

This work was supported by the Polish National Centre for Research and Development (POIR.01.01.01-00-

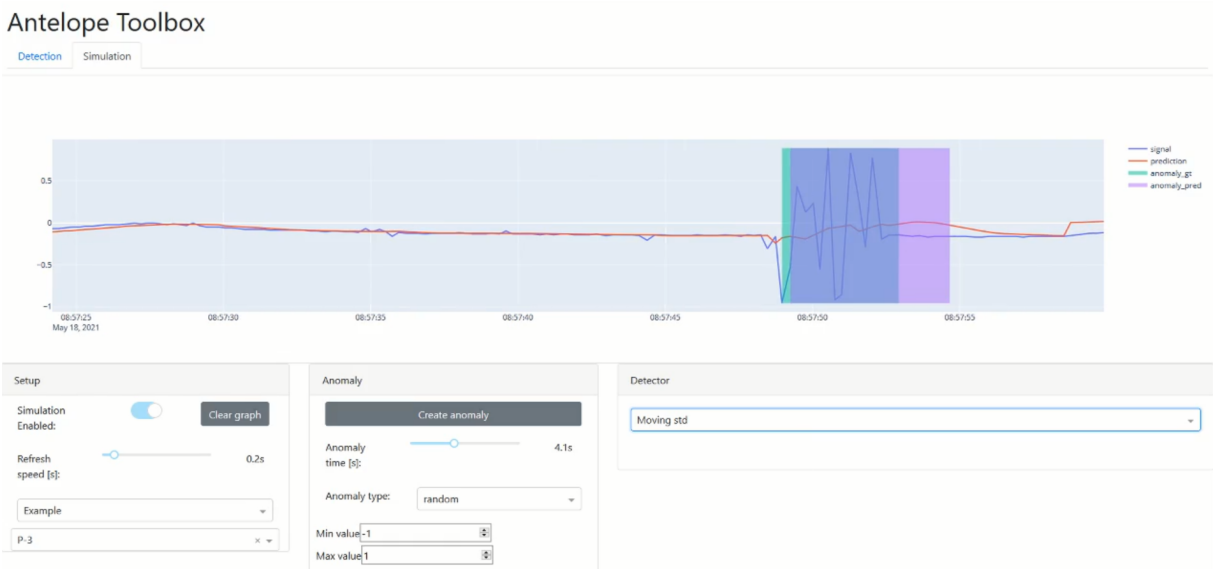


Figure 5. Analyzing the overlaps between the predicted and actual anomalies can be one of the indicators that quantify the performance of event detectors. The overlap indices may not, however, correspond to the practical utility of the automated anomaly detection systems, as we often wish to detect (or even predict) anomalies as fast as possible. Our Antelope Toolbox may be conveniently exploited to investigate the capabilities of not only deep learning-based techniques, but also classical methods, based e.g., on analyzing the signal and error characteristics within the moving window (in this example, we can compare the ground-truth and automatically annotated events in green and violet, respectively).

0853/19), and by the Silesian University of Technology grant for maintaining and developing research potential (BKM21) and Rector's grant (02/080/RGJ20/0003).

REFERENCES

[1] Xiaoyu Zhang, Jiusheng Chen, and Quan Gan. Anomaly detection for aviation safety based on an improved kpca algorithm. *Journal of Electrical and Computer Engineering*, 2017:4890921, Mar 2017.

[2] M A Ivanushkin, S S Volgin, I V Kaurov, and I S Tkachenko. Analysis of statistical methods for outlier detection in telemetry data arrays, obtained from "AIST" small satellites. *Journal of Physics: Conference Series*, 1326:012029, oct 2019.

[3] Hui Li, Jing He, and Fuqiang Cheng. Research on anomaly detection method for satellite power supply based on bayesian model. *IOP Conference Series: Materials Science and Engineering*, 782:032034, apr 2020.

[4] Lishuai Li, Maxime Gariel, R. John Hansman, and Rafael Palacios. Anomaly detection in onboard-recorded flight data using cluster analysis. In *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*, pages 1–31, 2011.

[5] Du Ying, Wang Fei, Sun Chao, Bao Jie, and Yang Qi. Anomaly detection of orbit satellite telemetry sequence based on two-window mode. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 1064–1068, 2018.

[6] Ryan Mukai, Zaid Towfic, Monika Danos, Mazen Shihabi, and David Bell. Msl telecom automated anomaly detection. In *2020 IEEE Aerospace Conference*, pages 1–6, 2020.

[7] Haixu Jiang, Ke Zhang, Jingyu Wang, Xianyu Wang, and Pengfei Huang. Anomaly detection and identification in satellite telemetry data based on pseudo-period. *Applied Sciences*, 10(1), 2020.

[8] Sara Abdelghafar, Ashraf Darwish, Aboul Ella Hassanien, Mohamed Yahia, and Afaf Zaghrout. Anomaly detection of satellite telemetry based on optimized extreme learning machine. *Journal of Space Safety Engineering*, 6(4):291–298, 2019.

[9] Weihua Jin, Bo Sun, Zhidong Li, Shijie Zhang, and Zhonggui Chen. Detecting anomalies of satellite power subsystem via stage-training denoising autoencoders. *Sensors*, 19(14), 2019.

[10] Yue Song, Jinsong Yu, Diyin Tang, Danyang Han, and Sen Wang. Telemetry data-based spacecraft anomaly detection using generative adversarial networks. In *Proc. ICSMD*, pages 297–301, 2020.

[11] Youjin Shin, Sangyup Lee, Shahroz Tariq, Myeong Shin Lee, Okchul Jung, Daewon Chung, and Simon S. Woo. Itad: Integrative tensor-based anomaly detection system for reducing false positives of satellite systems. In *Proc. CIKM*, page 2733–2740, New York, NY, USA, 2020. Association for Computing Machinery.

[12] Pawel Benecki, Szymon Piechaczek, Daniel Kostrzewa, and Jakub Nalepa. Detecting anomalies in spacecraft telemetry using evolutionary

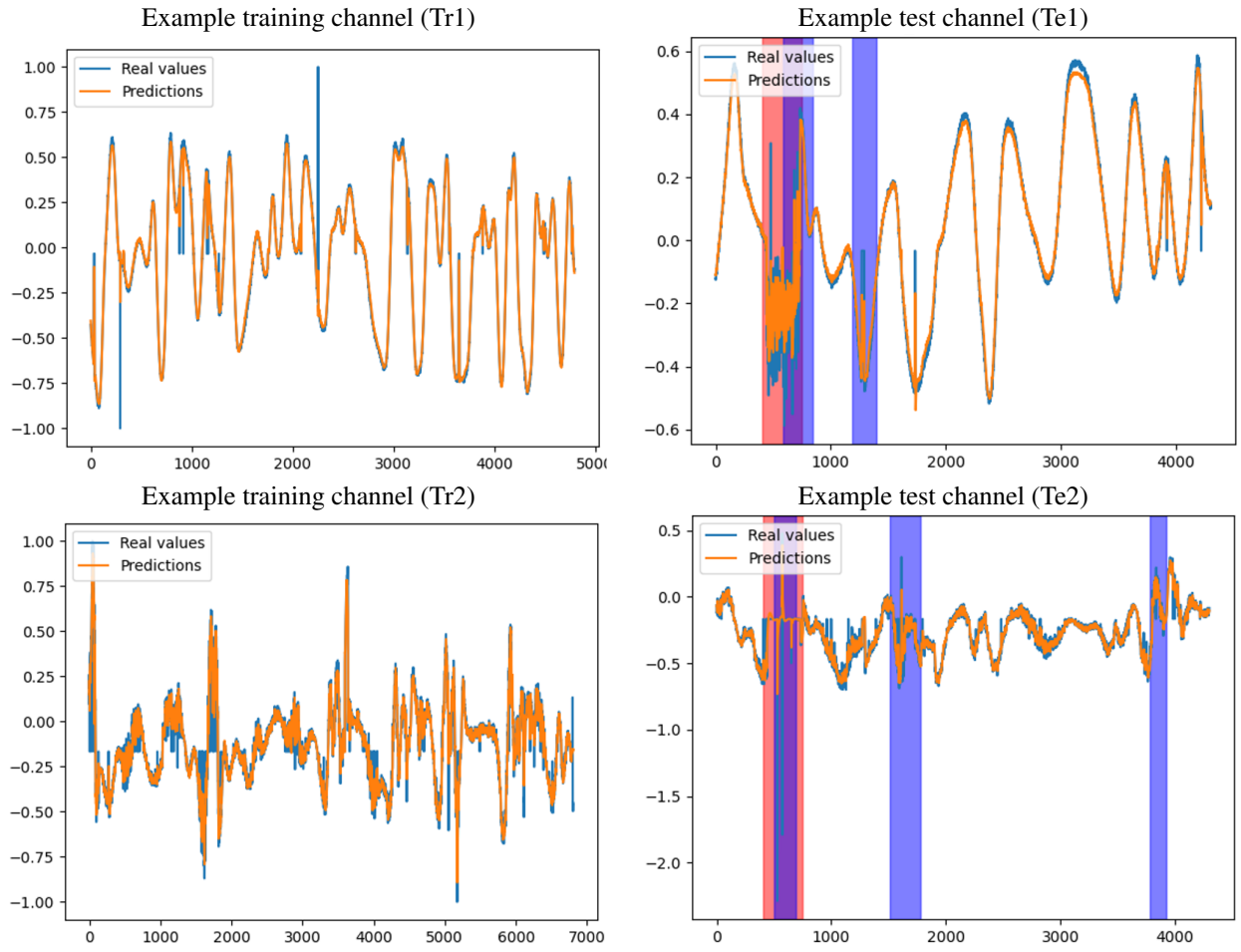


Figure 6. Example training and test telemetry channels captured on-board OPS-SAT. Although we do not precisely know if there were any anomalous events, the visual investigation may help us determine "suspicious" parts of the time series that could indicate on-board anomalies. It, however, will have to be double-checked with the Operations Team.

thresholding and LSTMs. In *Proc. GECCO*, page 1–2, New York, NY, USA, 2021. Association for Computing Machinery. in press.

[13] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proc. ACM SIGKDD*, pages 387–395, 2018.

[14] Jakub Nalepa and Marek Stanek. Segmenting hyperspectral images using spectral convolutional neural networks in the presence of noise. In *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2020, Waikoloa, HI, USA, September 26 - October 2, 2020*, pages 870–873, 2020.

[15] Jakub Nalepa, Michal Myller, Marcin Cwiek, Lukasz Zak, Tomasz Lakota, Lukasz Tulczyjew, and Michal Kawulok. Towards on-board hyperspectral satellite image segmentation: Understanding robustness of deep learning through simulating acquisition conditions. *Remote Sensing*, 13(8), 2021.

[16] Jakub Nalepa, Marek Antoniak, Michal Myller, Pablo Ribalta Lorenzo, and Michal Marcinkiewicz. Towards resource-frugal deep convolutional neural networks for hyperspectral image segmentation. *Microprocessors and Microsystems*, 73:102994, 2020.