Discovering outliers in the Mars Express thermal power consumption patterns

Matej Petković[1,2,*] , **Luke Lucas**[3] , **Tomaž Stepišnik**[1,2] , **Panče Panov**[1,2] ,
Nikola Simidjievski[1,2,4] and **Dragi Kocev**[1,3*]
[1]Bias Variance Labs, Ljubljana, Slovenia
[2]Jožef Stefan Institute, Ljubljana, Slovenia
[3]LSE Space GmbH ,Gilching, Germany
[4]University of Cambridge, United Kingdom

Abstract

The Mars Express (MEX) spacecraft has been orbiting Mars since 2004. The operators need to constantly monitor its behavior and handle sporadic deviations (outliers) from the expected patterns of measurements of quantities that the satellite is sending to Earth. In this paper, we analyze the patterns of the electrical power consumption of MEX's thermal subsystem that maintains the spacecraft's temperature at a desired level. The consumption is not constant, but should be roughly periodic in the short term, with the period that corresponds to one orbit around Mars. By using long short-term memory neural networks, we show that the consumption pattern is more irregular than expected, and successfully detect such irregularities, thus opening possibility for automatic outlier detection on MEX in the future.

1 Introduction

Spacecraft' health and endurance depend on close monitoring and accurate analysis of their telemetry data. Analyzing these data is non-trivial since telemetry data are heterogeneous and complex, comprised from measurements and activity records from the different on-board equipment and sensors, typically noisy and incomplete. In turn, operators need to constantly monitor and analyze them, handling sporadic deviations (outliers) from the expected patterns of measurements that relate to the spacecraft's behavior.

Outlier (or anomaly) detection refers to identification and investigation of rare (and unexpected) events and patterns in the data, which do not conform to the underlying data distribution. In the context of spacecraft operations, typically such outliers are a result of an on-board equipment malfunction or unexpected (and/or novel) environmental effect. In this work, we analyze telemetry data from the Mars Express (MEX) spacecraft to detect anomalies in the electrical power consumption of MEX's thermal subsystem which maintains the spacecraft's temperature at a desired level.

MEX, a long-lasting mission of the European Space Agency, has been exploring Mars since 2004. It is responsible for a wealth of scientific data comprised of three-dimensional renders of the surface and a complete map of the chemical composition of Mars's atmosphere that has led to important scientific discoveries, such as the evidence of the presence of water. Given the age and the current condition of MEX, monitoring this consumption and identifying unexpected malfunction has a direct consequence on the longevity of the spacecraft and its mission [Lucas and Boumghar, 2017; Breskvar *et al.*, 2017; Petković *et al.*, 2019a; Boumghar *et al.*, 2018; Petković *et al.*, 2019b]. We propose a machine learning (ML) approach for identifying outliers in the MEX's thermal power consumption patterns. The proposed approach combines several state-of-the art unsupervised ML methods for anomaly detection to obtain accurate estimates of anomalous behavior. We evaluate the proposed approach on 11.5 years of MEX data showcasing its potential and practical utility with respect to the identified outliers.

*Contact Authors: matej@bvlabs.ai, dragi@bvlabs.ai

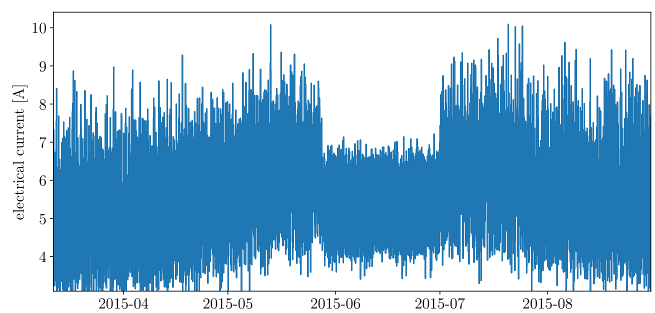

Figure 1: The conjunction of Mars and the Sun (June 2015) caused substantially different behavior of the MEX thermal subsystem.

2 Data

The data contains values of electrical currents running through the 33 electrical heaters on MEX that are part of the MEX thermal subsystem, spanning approximately 11.5 years, from 2008-08-22 to 2020-01-17. In our analyses, we sum the values of the individual heater lines in a total current $x(t)$. We analyze the time series $x(t)$ on the level of 15-minute intervals, as suggested in [Petković *et al.*, 2019a]. Each interval $[t_i, t_{i+1})$ ($t_{i+1} - t_i = 15\,\mathrm{min}$) is assigned a value $x(t_i)$, which is the average value of $x(t)$ for that time interval.

On short term, we assume that the values of electrical currents should be roughly periodic with the period that corresponds to one MEX orbit around Mars, which takes approximately 6.75 hours. We use this period as the unit of analysis and represent examples in the dataset as the vectors $x_i = [x(t_i), x(t_{i+1}), \dots, x(t_{i+N-1})]$ of $N = 27$ consecutive measurements of electrical current. The dataset contains 369,843 examples and note that consecutive examples overlap. Fig. 1 shows that the periodicity assumption is not true for the long term.

3 Related Work

Anomaly detection is a very active field of research [Chandola *et al.*, 2009; Pang *et al.*, 2020] that focuses on identifying point or collective anomalies (when either a single data point or consecutive data points are anomalous with respect to the entire signal) [Pilastre *et al.*, 2020]. In the context of analyzing telemetry data, anomaly detection methods typically focus on out-of-limits checking (comparing the values against predefined optimal operating ranges) and analysis of aggregated statistical features [Martínez-Heras *et al.*, 2012; Fuertes *et al.*, 2016; Martínez-Heras and Donati, 2014]. Other ML approaches include k-nearest neighbors (anomalies in XMM-Newton and Venus Express), support vector machines (for monitoring the status of a CNES-operated spacecraft) and others [Yairi *et al.*, 2017; Carlton *et al.*, 2018], including generative deep neural networks for detecting anomalies in the LUNar Attitude and Orbit Control System (AOCS) SIMulator (LUNASIM) [Ahn *et al.*, 2020]. Similarly, there are several recent attempts at using long short-term memory (LSTMs) networks for the task of anomaly detection [Hundman *et al.*, 2018; Pan *et al.*, 2020; Chen *et al.*, 2021].

4 Our Method

We propose constructing a heterogeneous ensemble combining different models for outlier detection. The models are derived from three unsupervised learning algorithms: k-means [MacQueen, 1967], isolation forest [Liu *et al.*, 2008] and long short-term memory (LSTM) autoencoders [Hochreiter and Schmidhuber, 1997]. The complete pipeline is presented in Fig. 2. The outlier score of k-means is the distance of an example to the closest center of the clusters, computed in k-means. Isolation forest outlier score is the average depth of a tree, which isolates an example from the others. The outlier score of LSTMs is measured as the reconstruction error (LSTMs learn the codes of the *normal* examples and then try to decode them back).

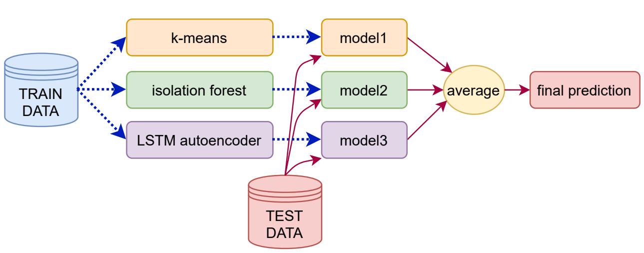

Figure 2: The proposed outlier detection approach: Train data is used to learn three different outlier detection models. During test time, the models are combined into an ensemble, by averaging the individual model prediction into the final, ensemble, prediction.

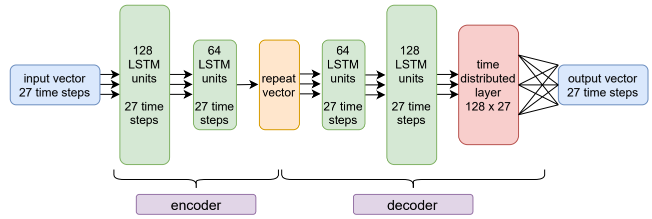

Figure 3: LSTM autoencoder architecture: an input layer, an encoder (comprised of two LSTM layers), a RepeatVector layer, a decoder (with two LSTM layers), time distributed layer and the output layer. Triple arrows denote that a sequence (of length $N = 27$) is passed to the next layer, while a single arrow denotes that only a single number is passed. The RepeatVector layer copies its input N times. The TimeDistributed layer is fully connected to the output layer. The value of N corresponds to the number of data points in one MEX orbit.

For the task of outlier detection, we combine the individual outlier scores from the three different methods in order to improve the overall performance, taking the classical idea behind the ensemble learning for better predictive models [Breiman, 1996]. Given the normalized scores s_{method} of the three methods (linearly mapped to the $[0, 1]$ interval), the final score of the ensemble is defined as

$$s_{\text{ensemble}}(x_i) = \frac{1}{3}(s_{\text{k-means}}(x_i) + s_{\text{IsoFor}}(x_i) + s_{\text{LSTM}}(x_i)).$$

5 Experimental Setup

Parametrisation. We set the number of clusters in k-means to $k = 50$, which allows for 50 prototypical curves of electrical current, since the preliminary experiments with the elbow method reveal that $k = 30$ prototypical curves could suffice, but however $k = 50$ decreases the amount of false positive alarms. The contamination parameter of Isolation Forests is set to 0.001, i.e., we expect 0.1% of the examples to be outliers. The remaining parameters are set to the values recommended in [Liu *et al.*, 2008]. The LSTM autoencoders are implemented using the Keras deep learning library [Chollet and others, 2015]. A more detailed overview of the autoencoder architecture is presented in Figure 3. We use *ReLU* function, since the preliminary experiments showed better performance than *tanh* activation. The models were trained for 1000 epochs using the Adam optimizer with a learning rate of 0.001 (and with the recommended parameters) and a batch size of 128 (chosen after evaluating batch sizes of $\{2^5, 2^6, \dots, 2^{14}\}$). We use the last 20% of examples in the train data for early stopping validation criteria: If no progress has been made in the last 50 epochs, the training stops. The objective function considered is mean-squared error.

Evaluation procedure. Since the data spans over 13 years, we create 12 train sets $\mathcal{D}_{\text{TRAIN}}$ with different lengths. All train sets start on 2008-22-8 but end on December 31st of each year ($y \in \{2009, 2010, \dots, 2020\}$). The respective test

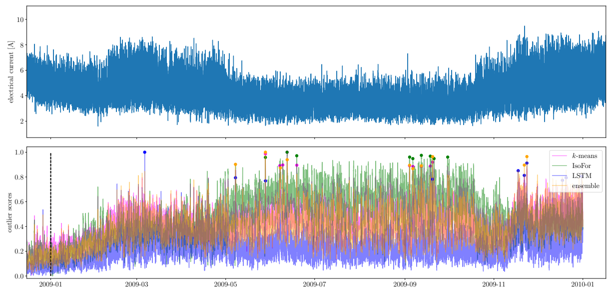

Figure 4: Outliers in the 2009 − 2010 period. The border between the train and test data is denoted by the vertical dotted line.

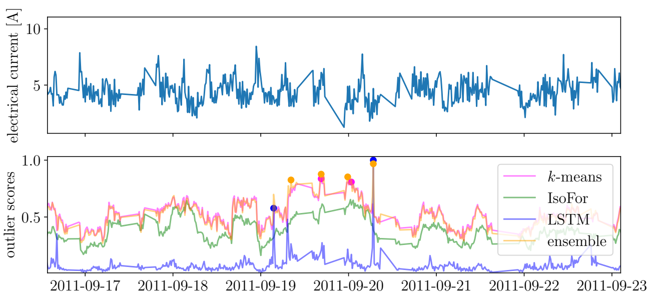

Figure 5: Outliers detected due to missing data in the 2011 − 2012 period.

sets start where the train set ended and is exactly one (Earth) year long. The exception is the 2020 test set, which is less than 2 months long. For each test set, we also identify the 10 orbits with the highest ensemble score s_{ensemble} which are then manually inspected and explained by a MEX spacecraft operator.

6 Results

We start with the test period 2009 − 2010, where the first conjunction appeared (note that no conjunction was present in the train data). Figure 4 shows that the ensemble outlier scores (as well as the single model scores) correctly identify this behavior as anomalous. Additionally, when the electrical current values were unusually high, this was again detected by the ensemble method (but not by some of the single models). Note that the conjunctions in later years are no longer considered anomalous, since the models can learn from the conjunction in 2009. In the testing period 2011 − 2012, most of the top-10 identified outliers are due to the missing values in the data (see Figure 5). Similar behavior can be also observed in the testing period 2013 − 2014.

The most abnormal patterns in the data (including the highest peaks in MEX's operation) have been detected in the 2017 − 2018 period. As shown in Fig. 6, the conjunction in that period has been classified by the ensemble as normal, while some of the individual models (e.g., isolation forest) have reported outliers. These outliers, however, quite diverge from the expected behavior but still remain challenging to be discovered by the models. This is further evident in the last test period 2018−2019 (Fig. 7), where the anomalous records of 2017 are used in the model-training process. In this scenario, the models report higher error on the training data than on the testing set (i.e., larger outliers are identified in the training set), meaning that the models while successful, struggle a bit to swiftly adapt to the new behavior.

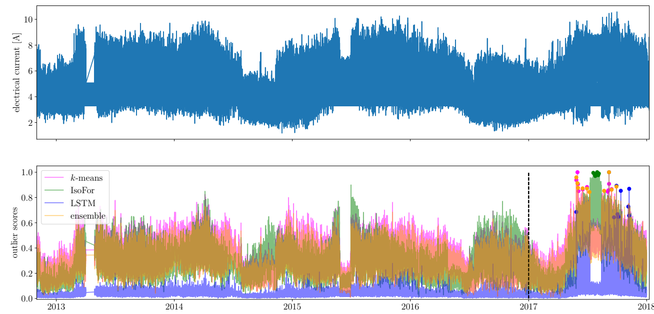

Figure 6: The most abnormal patterns in the data are detected in 2017.

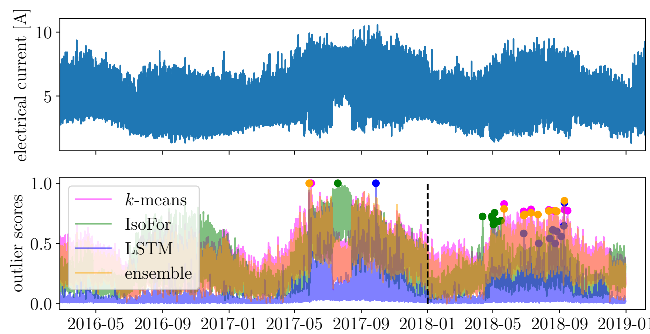

Figure 7: Outliers identified in the test set for the period 2018 − 2019. The models, while successful, slightly struggle to adapt to new anomalous behavior.

Another potential reason for this, is that the data in 2018 (and onward) differs from the previous years. In 2018, MEX operations underwent a fundamental change, which required a software upload. This process involved performing no science operations for weeks, implementing the new methodology, as well as flying in novel configurations. This had radical effects on the thermal power consumption. Which in turn makes learning from or comparing 2018 to any preceding year very difficult as it was so novel in so many ways.

7 Conclusion

In this paper, we propose an end-to-end ML approach for outlier detection in telemetry data – a heterogeneous ensemble of thee state-of-the art methods for outlier detection: k-means, isolation forest and LSTM autoencoders. We demonstrate the utility of the proposed approach on several tasks of identifying anomalous behavior in the electrical power consumption of the MEX's thermal subsystem. The results show that such an approach is able to accurately detect all major outliers (such as unusually high electrical currents, missing data and conjunctions). Moreover, this approach can provide additional insights into the spacecraft behavior during some rare events, such as the Siding Spring comet avoidance maneuvers.

References

[Ahn *et al.*, 2020] Hyojung Ahn, Dawoon Jung, and Han-Lim Choi. Deep generative models-based anomaly detection for spacecraft control systems. *Sensors*, 20(7):1991:1–20, 2020.

[Boumghar *et al.*, 2018] Redouane Boumghar, Luke Lucas, and Alessandro Donati. Machine learning in operations for the mars express orbiter. In *15th International Conference on Space Operations*, Marseille, France, 2018.

[Breiman, 1996] Leo Breiman. Bagging predictors. *Mach. Learn.*, 24(2):123–140, August 1996.

[Breskvar *et al.*, 2017] Martin Breskvar, Dragi Kocev, Jurica Levatić, Aljaž Osojnik, Matej Petković, Nikola Simidjievski, Bernard Ženko, Redouane Boumghar, and Luke Lucas. Predicting thermal power consumption of the mars express satellite with machine learning. In *Proceedings of the 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, pages 88–93, 2017.

[Carlton *et al.*, 2018] Ashley Carlton, Rachel Morgan, Whitney Lohmeyer, and Kerri Cahoy. Telemetry fault-detection algorithms: Applications for spacecraft monitoring and space environment sensing. *Journal of Aerospace Information Systems*, 15(5):239–252, 2018.

[Chandola *et al.*, 2009] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), 2009.

[Chen *et al.*, 2021] Junfu Chen, Dechang Pi, Zhiyuan Wu, Xiaodong Zhao, Yue Pan, and Qiang Zhang. Imbalanced satellite telemetry data anomaly detection model based on bayesian lstm. *Acta Astronautica*, 180:232–242, 2021.

[Chollet and others, 2015] François Chollet et al. Keras, https://keras.io, 2015.

[Fuertes *et al.*, 2016] Sylvain Fuertes, Gilles Picart, Jean-Yves Tourneret, Lotfi Chaari, André Ferrari, and Cédric Richard. Improving spacecraft health monitoring with automatic anomaly detection techniques. In *Proceedings of the 2016 SpaceOps Conference*, pages 1–16. American Institute of Aeronautics and Astronautics, Inc., 2016.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[Hundman *et al.*, 2018] 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, page 387–395, New York, NY, USA, 2018. Association for Computing Machinery.

[Liu *et al.*, 2008] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, ICDM '08, pages 413–422, USA, 2008. IEEE Computer Society.

[Lucas and Boumghar, 2017] Luke Lucas and Redouane Boumghar. Machine learning for spacecraft operations support - The Mars Express Power Challenge. In *Proceedings of the Sixth International Conference on Space Mission Challenges for Information Technology, SMC-IT 2017*, pages 82–87, 2017.

[MacQueen, 1967] J. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability - Vol. 1*, pages 281–297. University of California Press, Berkeley, CA, USA, 1967.

[Martínez-Heras and Donati, 2014] José-Antonio Martínez-Heras and Alessandro Donati. Enhanced telemetry monitoring with novelty detection. *AI Magazine*, 35(4):37–46, 2014.

[Martínez-Heras *et al.*, 2012] José-Antonio Martínez-Heras, Alessandro Donati, Marcus G. F. Kirsch, and Frederic Schmidt. New telemetry monitoring paradigm with novelty detection. In *Proceedings of the 2012 SpaceOps Conference*, pages 1–9. American Institute of Aeronautics and Astronautics, Inc., 2012.

[Pan *et al.*, 2020] 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, 2020.

[Pang *et al.*, 2020] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton van den Hengel. Deep learning for anomaly detection: A review. *CoRR*, abs/2007.02500, 2020.

[Petković *et al.*, 2019a] Matej Petković, Redouane Boumghar, Martin Breskvar, Sašo Džeroski, Dragi Kocev, Jurica Levatić, Luke Lucas, Aljaž Osojnik, Bernard Ženko, and Nikola Simidjievski. Machine learning for predicting thermal power consumption of the mars express spacecraft. *IEEE Aerospace and Electronic Systems Magazine*, 34(7):46–60, 2019.

[Petković *et al.*, 2019b] Matej Petković, Luke Lucas, Dragi Kocev, Sašo Džeroski, Redouane Boumghar, and Nikola Simidjievski. Quantifying the effects of gyroless flying of the mars express spacecraft with machine learning. In *Proceedings of the 2019 IEEE International Conference on Space Mission Challenges for Information Technology (SMC-IT)*, pages 9–16, 2019.

[Pilastre *et al.*, 2020] Barbara Pilastre, Loïc Boussouf, Stéphane D'Escrivan, and Jean-Yves Tourneret. Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning. *Signal Processing*, 168:107320, 2020.

[Yairi *et al.*, 2017] Takehisa Yairi, Naoya Takeishi, Tetsuo Oda, Yuta Nakajima, Naoki Nishimura, and Noboru Takata. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction. *IEEE Transactions* *on Aerospace and Electronic Systems*, 53(3):1384–1401, 2017.