Spacecraft Health Monitoring using a Weighted Sparse Decomposition

Barbara PILASTRE and Jean-Yves TOURNERET[1,2,3]

[1]INP-ENSEEIHT, University of Toulouse, 2 Rue Camichel, 31071 Toulouse, France
[2]IRIT, UMR5505, 118 Route de Narbonne, 31062 Toulouse, France
[3]TéSA, 7 Boulevard de la Gare, 31500 Toulouse, France

E-mail: barbara.pilastre@tesa.prd.fr,jean-yves.tourneret@enseeiht.fr

Loïc BOUSSOUF

Airbus Defence and Space, 31 rue des cosmonautes, 31400, Toulouse, France

E-mail: loic.boussouf@airbus.com

Stéphane D'ESCRIVAN

Centre National d'Études Spatiales (CNES), 18 avenue Édouard Belin, 18 av. Edouard Belin, 31400, Toulouse, France

E-mail: Stephane.d-Escrivan@cnes.fr

Abstract. In space operations, spacecraft health monitoring and failure prevention are major issues. This important task can be handled by monitoring housekeeping telemetry time series using anomaly detection (AD) techniques. The success of machine learning methods makes them attractive for AD in telemetry via a semi-supervised learning. Semi-supervised learning consists of learning a reference model from past telemetry acquired without anomalies in the so-called learning step. In a second step referred to as test step, most recent telemetry time-series are compared to this reference model in order to detect potential anomalies. This paper presents an extension of an existing AD method based on a sparse decomposition of test signals on a dictionary of normal patterns. The proposed method has the advantage of accounting for possible relationships between different telemetry parameters and can integrate external information via appropriate weights that allow detection performance to be improved. After recalling the main steps of an existing AD method based on a sparse decomposition [1] for multivariate telemetry data, we investigate a weighted version of this method referred to as W-ADDICT that allows external information to be included in the detection step. Some representative results obtained using an anomaly dataset composed of actual anomalies that occurred on several satellites show the interest of the proposed weighting strategy using external information obtained from the correlation coefficient between the tested data and its decomposition on the dictionary.

Keywords: Machine learning, spacecraft health monitoring, anomaly detection, sparse decomposition, dictionary learning

1. Introduction

Spacecraft health monitoring and failure prevention are major issues that can be handled by checking housekeeping telemetry time series. Housekeeping telemetry is composed of hundreds to thousands telemetry parameters describing the evolution over time of physical quantities such as temperature, pressure, voltage. Detecting anomalies in these parameters jointly can be a complicated task. Thus, a lot of state-of-the-art telemetry AD methods consider an univariate framework which consists of detecting anomalies in univariate time series separately via a semi-supervised learning. The idea is to learn nominal spacecraft behaviours from past telemetry composed of normal patterns (i.e., data without anomalies) and build a reference model to which new data can be compared. Then, any deviation from the nominal behaviour is considered as a fault. In this context, no assumptions are made about anomalies whose the type is not required for the learning. This represents a real advantage compared to other methods based on supervised learning in which labelled data are needed. Examples of univariate anomalies affecting telemetry time series are displayed in Fig. 1 (top, red boxes). Popular machine learning (ML) methods that have been investigated in this framework include the one-class support vector machine [2], nearest neighbour techniques [3, 4] or neural networks [5, 6]. However, these methods do not account for possible relationships existing between different parameters and thus cannot detect multivariate anomalies resulting from changes in the relationships between several telemetry time series. An example of such multivariate anomaly is displayed in Fig. 1 (bottom, red box). The detection of the abnormal behaviour in the top time series of this figure (outlined in the red box) will be detected more easily by using the bottom time series. This kind of situation is often referred to as "contextual anomaly", which has been recently investigated in [7, 8, 9].

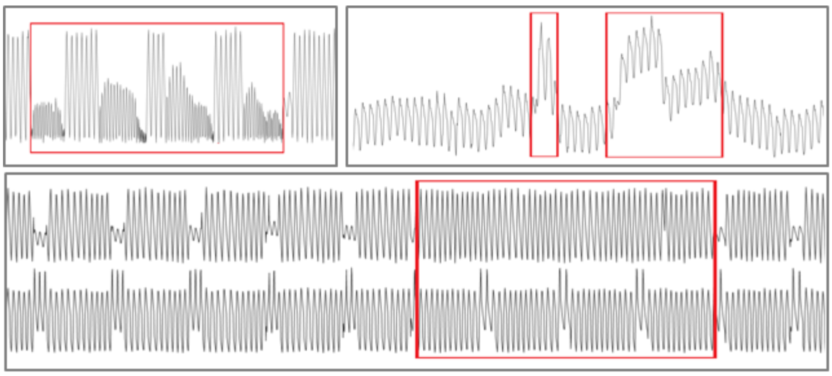

Figure 1. Examples of univariate (top) and multivariate (bottom) anomalies.

2. Proposed Anomaly Detection Method

2.1. Preprocessing

The proposed AD method first segments the different telemetry times series into windows of fixed size w as illustrated in Fig. 2. This segmentation step has been used successfully in many telemetry AD methods [2, 3]. Each resulting matrix is

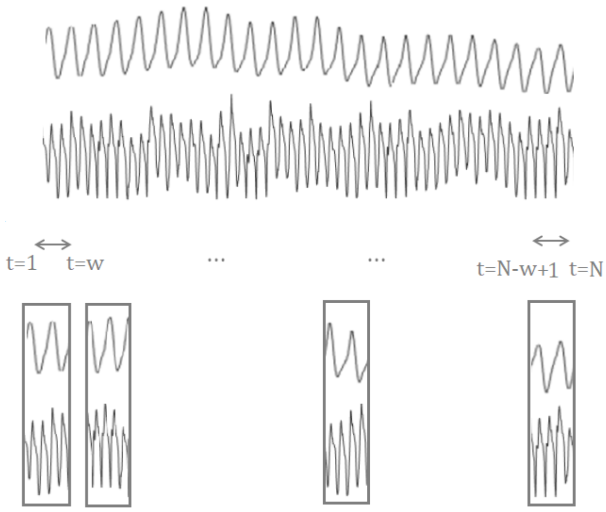

Figure 2. Segmentation of telemetry into windows

then transformed into a vector whose first w components correspond to the first time serie denoted as $\boldsymbol{y}_1$, whose components $w+1$ to $2w$ correspond to the second time serie denoted as $\boldsymbol{y}_2$ and so on. This preprocessing creates mixed signals composed of telemetry time series formed by the different parameters, i.e., $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_K^T]^T$ with $\boldsymbol{y}_k \in \mathbb{R}^w, \mathrm{k} = 1, ..., K$, where K is the number of telemetry parameters and w is the size of the time window.

2.2. Anomaly Detection using a Weighted Sparse Decomposition

Sparse decompositions have received an increasing attention in many signal and image processing applications. These decompositions were investigated for univariate AD in [10]. The AD strategy introduced in [10] decomposes each multivariate test signal as the sum of three signals: a nominal signal defined by a sparse representation into a dictionary of normal patterns, an anomaly signal and an additive noise such that

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e} + \mathbf{b} \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the multivariate test signal, $\boldsymbol{\Phi} \in \mathbb{R}^{N \times L}$ is a dictionary previously learned from data describing normal behaviours of spacecraft data, $\boldsymbol{x} \in \mathbb{R}^L$ is a sparse vector of coefficients, $\boldsymbol{e} \in \mathbb{R}^N$ is a possible anomaly signal (with $\boldsymbol{e} = 0$ in absence of anomaly) and $\mathbf{b} \in \mathbb{R}^N$ is an additive noise. The proposed multivariate framework considered in this work and the associated preprocessing divide the multivariate signals $\boldsymbol{y}$ and $\boldsymbol{e}$ as K blocks associated with the K telemetry parameters, i.e, $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_K^T]^T$ and $\boldsymbol{e} = [\boldsymbol{e}_1^T, ..., \boldsymbol{e}_K^T]^T$. The resulting sparse decomposition problem can be expressed as follows [1]

$$\min_{\mathbf{x},\mathbf{e}} \frac{1}{2} \parallel \mathbf{y} - \boldsymbol{\Phi}\mathbf{x} - \mathbf{e} \parallel_2^2 + a\|\mathbf{x}\|_1 + b\sum_{k=1}^{K} \|\mathbf{e}_k\|_2 \tag{2}$$

where $\|\boldsymbol{x}\|_1 = \sum_{l=1}^{L} |x_l|$ is the ℓ_1 norm of the vector $\boldsymbol{x} = (x_1, ..., x_L)^T$ and $\|\mathbf{e}_k\|_2$ is the euclidean norm of the vector $\boldsymbol{e}_k$. Note that (2) considers two distinct sparsity constraints for the coefficient vector $\boldsymbol{x}$ and the anomaly signal $\mathbf{e}$. This formulation reflects the fact that a nominal continuous signal can be well approximated by a linear combination of *few* atoms of the dictionary (sparsity of $\boldsymbol{x}$ using the ℓ_1 norm) and that anomalies are *rare* and affect few parameters at the same time (sparsity of $\mathbf{e}$ using the ℓ_{21} norm). This paper generalizes the AD strategy defined by (2) referred to as ADDICT [1] by the integration of external information via appropriate weights. The proposed model for AD in spacecraft telemetry is similar to the well-known group-lasso model [11] that assigns a weight to each group. In this work we consider K groups associated with the K telemetry parameters and propose to solve the following problem

$$\arg\min_{\mathbf{x},\mathbf{e}} \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}-\mathbf{e}\|_2^2 + a\|\boldsymbol{x}\|_1 + b\sum_{k=1}^{K} w_k\|\mathbf{e}_k\|_2. \tag{3}$$

where $w_k > 0$ is the kth weight associated with the kth parameter. The problem (3) can be solved with the alternating direction method of multipliers (ADMM) [12] by adding an auxiliary variable $\boldsymbol{z}$ and the constraint $\boldsymbol{z} = \boldsymbol{x}$ leading to

$$\arg\min_{\mathbf{x},\mathbf{e},\mathbf{z}} C(\mathbf{x},\mathbf{e},\mathbf{z}) = \frac{1}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{x}-\mathbf{e}\|_2^2 + a\ \|\boldsymbol{z}\|_1 + b\sum_{k=1}^{K} w_k\|\mathbf{e}_k\|_2 \quad \text{s.t.} \quad \mathbf{z}=\mathbf{x} \tag{4}$$

where "s.t" means "subject to". Note that, contrary to Problem (3), the first and second terms of (4) are decoupled, which allows an easier estimation of the vector $\boldsymbol{x}$. The ADMM algorithm associated with (4) minimizes the following augmented Lagrangian

$$\mathcal{L}_A(\mathbf{x},\mathbf{z},\mathbf{e},\mathbf{m},\mu) = C(\mathbf{x},\mathbf{e},\mathbf{z}) + \mathbf{m}^T(\mathbf{z}-\mathbf{x}) + \frac{\mu}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 \tag{5}$$

where $\mathbf{m}$ is a vector containing Lagrange multipliers and μ is a regularization parameter controlling the level of deviation between $\boldsymbol{z}$ and $\boldsymbol{x}$. The ADMM algorithm iteratively updates $\boldsymbol{x}$, $\mathbf{z}$, $\mathbf{e}$ and $\mathbf{m}$ as follows

Update of $\boldsymbol{x}$

$\boldsymbol{x}$ is classically updated as follows

$$\boldsymbol{x}^{k+1} = \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{y}-\mathbf{\Phi}\boldsymbol{x}-\mathbf{e}^k\|_2^2 + \mathbf{m}^k(\mathbf{z}^k-\boldsymbol{x}) + \frac{\mu^k}{2}\|\mathbf{z}^k-\boldsymbol{x}\|_2^2 \right\}. \tag{6}$$

Simple algebra leads to

$$\widehat{\boldsymbol{x}}^{k+1} = (\mathbf{\Phi}\mathbf{\Phi}^T + \mu^k I)^{-1}[\mathbf{\Phi}^T(\boldsymbol{y}-\boldsymbol{e}^k) + \mathbf{m}^k + \mu^k\mathbf{z}^k]. \tag{7}$$

Update of $\mathbf{z}$

The update of $\mathbf{z}$ is defined as

$$\widehat{\mathbf{z}}^{k+1} = \arg\min_{\mathbf{z}} \left\{ a\|\mathbf{z}\|_1 + (\mathbf{m}^k)^T(\mathbf{z}-\boldsymbol{x}^{k+1}) + \frac{\mu^k}{2}\|\mathbf{z}-\boldsymbol{x}^{k+1}\|_2^2 \right\}. \tag{8}$$

The solution of (8) is given by the element-wise soft thresholding operator $\widehat{\mathbf{z}}^{k+1} = S_{\gamma^k}\left(\boldsymbol{x}^{k+1} - \frac{1}{\mu^k}\mathbf{m}^k\right)$ with $\gamma^k = \frac{a}{\mu^k}$, where the thresholding operator $S_\gamma(\mathbf{u})$ is defined by

$$S_\gamma(\mathbf{u}) = \begin{cases} \mathbf{u}(n)-\gamma & \text{if } \mathbf{u}(n) > \gamma \\ 0 & \text{if } |\mathbf{u}(n)| \leq \gamma \\ \mathbf{u}(n)+\gamma & \text{if } \mathbf{u}(n) < -\gamma \end{cases} \tag{9}$$

where $\mathbf{u}(n)$ is the nth component of $\mathbf{u}$.

Updates of m and μ

The updates of $\mathbf{m}$ and μ are defined as

$$\widehat{\mathbf{m}}^{k+1} = \mathbf{m}^k + \mu^k(\mathbf{z}^{k+1} - \boldsymbol{x}^{k+1}) \quad \text{and} \quad \widehat{\mu}^{k+1} = \rho\mu^k \quad \text{with} \quad \rho > 1.$$

Update of e

Due to the reweighted scheme investigated in this paper, the update of the anomaly vector $\boldsymbol{e}$ changes from the non-weighted scheme studied in [1], leading to

$$\widehat{\mathbf{e}} = \arg\min_{\mathbf{e}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi}\mathbf{x} - \mathbf{e}\|_2^2 + b\sum_{k=1}^{K} w_k \|\mathbf{e}_k\|_2 \right\}. \tag{10}$$

The problem (10) can be divided into the following K independent problems

$$\widehat{\mathbf{e}}_k = \arg\min_{\mathbf{e}_k} \left\{ \frac{1}{2}\|\mathbf{h}_k - \mathbf{e}_k\|_2^2 + bw_k\|\mathbf{e}_k\|_2 \right\}. \tag{11}$$

where $\mathbf{h} = \mathbf{y} - \mathbf{\Phi}\mathbf{x}$ and $\mathbf{h}_k$ is the kth element of $\mathbf{h}$ associated with the kth telemetry time series for $k = 1, ..., K$. The solution of (11) is given by the group-shrinkage operator $\widehat{\mathbf{e}} = T_b(\mathbf{h})$ defined by

$$[T_b(\mathbf{h})]_k = \begin{cases} \frac{\|\mathbf{h}_k\|_2 - bw_k}{\|\mathbf{h}_k\|_2}\mathbf{h}_k & \text{if } \|\mathbf{h}_k\|_2 > bw_k \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

Note that the estimation of the anomaly signal $\boldsymbol{e}$ directly depends on the residual $\mathbf{h}$ resulting from the sparse decomposition of the test signal in the dictionary of normal patterns. Moreover, a value of the weight higher than 1 tends to reduce AD whereas a value lower than 1 promotes AD (for $w_k \geq 1$, the first equality is more difficult to satisfy, which leads to $\widehat{\boldsymbol{e}} = 0$ more frequently).

2.3. Anomaly Detection Strategy

The proposed AD strategy referred to as W-ADDICT (for Weighted Anomaly Detection using a sparse decomposition on a DICTionary) is defined by comparing the estimated anomaly signal $\boldsymbol{e}$ denoted as $\widehat{\boldsymbol{e}}$ to an appropriate threshold, i.e.,

$$\text{Anomaly detected if } \|\widehat{\boldsymbol{e}}\|_2^2 > S_{\text{PFA}} \tag{13}$$

where S_{PFA} is the threshold depending on the desired compromise between false alarm and good detection rates. This threshold can be determined using receiver operating characteristics (ROCs) if a ground-truth is available or using the user's experience. ROCs express the probability of detection PD as a function of the probability of false alarm PFA. In this work, these probabilities were computed using ground-truth time series and the threshold S_{PFA} was determined from the pair (PFA, PD) located the closest to the ideal point $(0, 1)$.

2.4. Weight updating using a correlation coefficient

Weighting the square norms $\|\mathbf{e}_k\|_2$ allows external information from expert knowledge to be considered. This paper investigates the usefulness of the correlation coefficient between the test signal $\boldsymbol{y}$ and its sparse decomposition $\boldsymbol{\Phi x}$, assuming that the smaller the value of the correlation coefficient, the higher the probability that $\boldsymbol{y}$ is affected by an anomaly. As explained before, weights interfere in the detection process in such a way that a small value of the weight promotes the detection of an anomaly. Conversely, a large value of the weight tends to reduce the number of detected anomalies. In order to respect this idea, we propose to build the weights with the following function

$$\begin{array}{cccc} f_\alpha & : & [-1,1] & \rightarrow & \mathbb{R}^+ \\ & & c_k & \mapsto & w_k = \frac{1}{((1+\alpha)-c_k)^2} \end{array} \quad (14)$$

which is an increasing function of c_k where c_k is the correlation coefficient between the test signal and its reconstruction associated with the kth parameter and α is an appropriate hyperparameter fixed by the user. The choice of this function is motivated by its simplicity since it only depends on one parameter α. Note that $c_k = \alpha$ corresponds to $w_k = 1$, i.e., to the unweighted model. For $c_k < \alpha$, the kth weight satisfies $w_k < 1$, which favours AD. Conversely, when $c_k > \alpha$, the kth weight is such that $w_k > 1$, which limits AD. Our experiments have shown that α can be tuned in the interval $]-1,+1[$ without loss of generality.

3. Experimental Results

The AD method (2) and its weighted version (3) (with weights built using the correlation coefficient as proposed in Section 2.4) have been evaluated on a representative anomaly dataset composed of $K = 7$ telemetry parameters with an available ground-truth. The hyperparameters a and b were tuned by cross validation using ground-truth and ROC curves. In all the experiments, the dictionary was composed of $L = 2000$ atoms learnt using the K-SVD algorithm [13] with two months of nominal telemetry (without anomalies), which represents approximately 5000 mixed training signals obtained after applying the preprocessing described in Section 2.1 with the parameter $w = 50$ (i.e., the signal length is $N = 350$). The anomaly database was built using 18 days of telemetry, i.e., was composed of 1000 signals, 82 of them being affected by anomalies. Note that the 82 anomalies were divided into 5 anomaly periods with various durations, as illustrated in the three examples of anomalies that are displayed in Fig. 1. Note also that a specific attention was devoted to the construction of a heterogeneous database (containing discrete and continuous time-series, with anomalies of different amplitudes and durations, ...). Finally, it is important to note that the majority of anomalies are actual anomalies that have been observed in operated satellites.

Fig. 3 displays the anomaly scores $\|\widehat{e}\|_2^2$ returned by the unweighted AD method (2) (left) and the weighted one (3) (right) for each of the 1000 test signals of the anomaly dataset. Note that the actual anomalies are marked by red backgrounds whereas the red horizontal lines indicate the detection thresholds (determined using cross validation

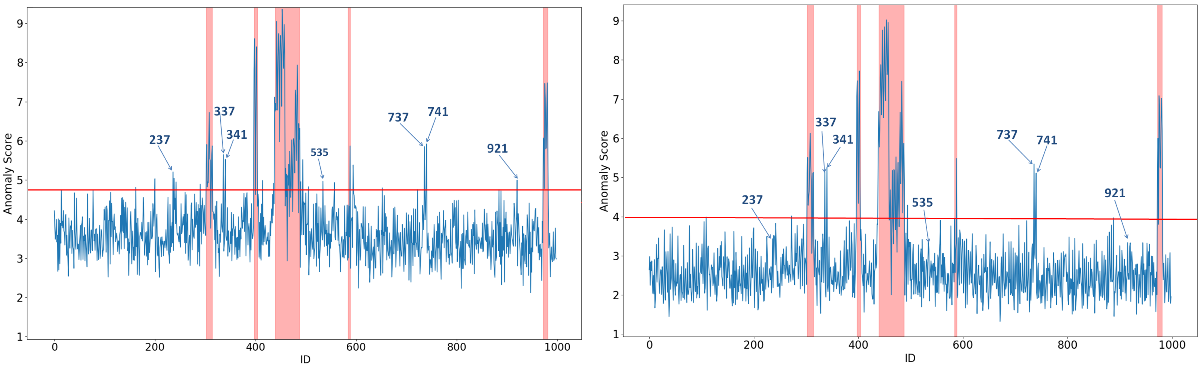

Figure 3. Anomaly scores obtained with unweighted algorithm (2) (left) and with the weighted one (3) (right) with ground-truth marked by red background and the detection threshold indicated by the red horizontal lines

as described in Section 2.3). As can be observed, the score exceeds threshold almost exclusively in presence of anomaly, meaning that anomalies are well detected and that few false alarms are returned. Note that the contextual corresponding to the last anomaly period of the ground-truth is also well detected. These first results show the interest of using sparse decompositions for anomaly detection in spacecraft telemetry. Furthermore, they show that the weighted model allows the number of false alarms to be reduced. For example, the test signals #237, #535 and #921 indicated by blue arrows are false alarms for the non-weighted scheme whereas they are not detected by the weighted algorithm.

Quantitative results in terms of probability of detection, probability of false alarm and area under the curve (AUC) are reported in Table 3. These results confirm that the weighted model reduces the number of false alarms for a fixed probability of detection (for PD = 89%, the PFA decreases from PFA = 5.2% for the unweighted model to PFA = 2.2% for the weighted one). Note that false alarm reduction is very interesting for spacecraft monitoring applications because each detection of the algorithm is examined by experts, which is potentially a very time consuming task. Note also that the detection threshold might be optimized using other criteria, e.g., allowing a smaller value of PD to decrease the number of false alarms.

Table 1. Probability of detection, probability of false alarm and area under curves (AUC) for the unweighted model (2) and the proposed weighted model (3)

Method	Threshold	PD	PFA	AUC
Unweighted model (2)	4.9	89%	5.23%	0.93
Weighted model (3)	4.3	89%	2.18%	0.96

4. Conclusion

This paper showed the interest of sparse decompositions for anomaly detection in multivariate housekeeping telemetry time series. In particular, external information provided by the user can be incorporated in the anomaly detection via appropriate weights. Our first results indicated that weights built from the correlation coefficient between the test signal and its sparse decomposition in a dictionary of normal patterns allowed a significant reduction of the false alarm probability.

For future work, a scaling-up is needed to evaluate the anomaly detection method in an operating context including hundreds to thousands telemetry time series. Concerning the weighted model, it would be interesting to build weights using other information than the correlation coefficient. Finally, we think that the model could be updated sequentially using expert feedback in order to improve detection, e.g., in the case of anomalies evolving with time. This opens the way for many works related to online or sequential anomaly detection, which could be useful for spacecraft health monitoring.

5. Acknowledgements

The authors would like to thank Pierre-Baptiste Lambert from CNES and Clémentine Barreyre from Airbus Defence and Space for fruitful discussions about anomaly detection in spacecraft telemetry. This work was supported by CNES and Airbus Defence and Space.

References

[1] B. Pilastre and *al.*, "Anomaly detection in mixed telemetry data using a sparse representation and dictionary learning," *Signal Processing*, to appear 2019.

[2] S. Fuertes and *al.*, "Improving spacecraft health monitoring with automatic anomaly detection techniques," in *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, Daejeon, South Korea, May 2016.

[3] J.-A. Martínez-Heras, A. Donati, M. Kirksch, and F. Schmidt, "New telemetry monitoring paradigm with novelty detection," in *Proc. Int. Conf. Space Operations (SpaceOps'2012)*, Stockholm, Sweden, Jun. 2012.

[4] C. O'Meara and *al.*, "Athmos: Automated telemetry health monitoring system at GSOC using outlier detection and supervised machine learning," in *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, Daejeon, South Korea, May 2016.

[5] ——, "Applications of deep learning neural networks to satellite telemetry monitoring," in *Proc. Int. Conf. Space Operations (SpaceOps'2018)*, Marseille, France, Jun. 2018.

[6] K. Hundman and *al.*, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. Int. Conf. Knowledge Data Mining (KDD'18)*, London, United Kingdom, Aug. 2018, pp. 387–395.

[7] N. Takeishi and *al.*, "Anomaly detection from multivariate times-series with sparse representation," in *Proc. IEEE Int. Conf. Syst. Man and Cybernetics*, San Siego, CA, USA, Oct. 2014.

[8] T. Yairi and *al.*, "A data-driven health monitoring method for satellite housekeeping data based on pobabilistic clustering and dimensionality reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1384–1401, Jun. 2017.

[9] C. Barreyre, "Statistiques en grande dimension pour la détection d'anomalies dans les données fonctionnelles issues des satellites," Ph.D. dissertation, Université de Toulouse, Toulouse, France, 05 2018.

[10] A. Adler and *al.*, "Sparse coding with anomaly detection," *J. Signal Process. Syst.*, vol. 79, no. 2, pp. 179–188, May 2015.

[11] M. Yuan and *al.*, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. Series B (Methodological)*, vol. 68, no. 1, pp. 49–67, 2006.

[12] S. Boyd and *al.*, "Distributed optimization and statistical learning via alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–222, 2010.

[13] M. Aharon and *al.*, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse decomposition," *IEEE Trans. Signal. Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.