

Représentation Parcimonieuse Pondérée pour la Détection d'Anomalies dans des Signaux Multivariés

Barbara PILASTRE¹, Loïc BOUSSOUF², Stéphane D'ESCRIVAN³, Jean-Yves TOURNERET¹

¹IRIT/INP-ENSEEIH/TéSA

7 Boulevard de la gare, 31500 Toulouse, France

²Airbus Defense and Space

31 rue des Cosmonautes, 31400 Toulouse, France

³Centre National d'Etudes Spatiales (CNES)

18 av Edouard Belin, 31401 Toulouse CEDEX 9, France

barbara.pilastre@tesa.prd.fr, loic.boussouf@airbus.com

Stephane.d-Escrivan@cnes.fr, jean-yves.tourneret@enseeiht.fr

Résumé – Cet article présente un modèle de représentation parcimonieuse pondérée pour la détection d'anomalies dans des signaux de télémesure satellite multivariés. La méthode proposée est une extension de l'état de l'art par son adaptation au cadre multivarié et l'intégration d'informations externes par l'intermédiaire d'une pondération appropriée permettant d'améliorer les performances de détection.

Abstract – This paper studies a new weighted sparse representation model for anomaly detection in multivariate telemetry signals. The proposed strategy generalizes an existing state-of-the-art method by an adaptation to the multivariate framework and the integration of external information via appropriate weights that allow detection performance to be improved.

1 Introduction

Dans de nombreux domaines d'application, la surveillance automatique de systèmes et la détection des anomalies présentes dans ces systèmes sont des enjeux majeurs [1]. À titre d'exemple, on peut citer le domaine du spatial dans lequel il est crucial de pouvoir assurer le suivi de l'état de santé des satellites et prévenir les anomalies notamment au travers de la surveillance de la télémesure. Les signaux de télémesure décrivent l'évolution dans le temps de différents paramètres représentant chacun une grandeur physique telle qu'une température, une tension ou encore une pression. Les méthodes récemment proposées pour la détection d'anomalies dans la télémesure satellite reposent sur des algorithmes d'apprentissage statistique semi-supervisé : le fonctionnement de signaux de télémesure associés à un fonctionnement normal du satellite est appris pour construire un modèle de référence auquel sont comparés les signaux de télémesure récemment acquis.

Les méthodes de détection d'anomalies dans la télémesure satellite peuvent être classées en deux catégories : les méthodes univariées et les méthodes multivariées. Les méthodes univariées traitent les paramètres de télémesure indépendamment les uns des autres. Ces approches permettent de détecter un comportement atypique d'un paramètre spécifique, i.e., un comportement inattendu de ce paramètre. On parle d'anomalie univariée car ce comportement est associé à l'évolution d'un unique paramètre. Deux exemples d'anomalies univariées sont visibles dans la figure 1 (haut). Les méthodes classiques de détection d'anomalies univariées dans les signaux de télémesure reposent sur des algorithmes puissants issus de méthodes comme les machines à vecteurs support (One-class support vecteur machine ou One-class SVM) [2], les réseaux de neurones [3] ou la re-

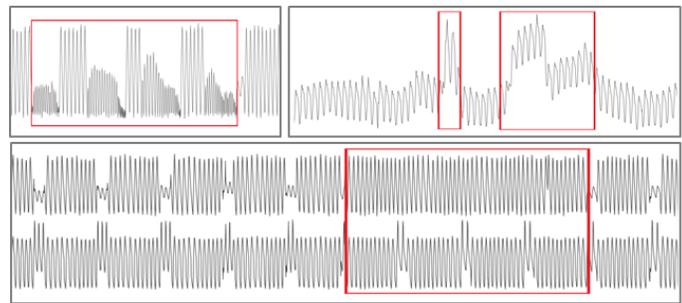


FIGURE 1 – Exemple d'anomalies univariées (haut) et multivariée (bas)

cherche de plus proches voisins [4, 5]. Les méthodes de détection d'anomalies multivariées traitent plusieurs paramètres de manière conjointe afin d'apprendre leurs comportements simultanés et donc les éventuelles relations ou corrélations entre ces paramètres. Ce cadre multivarié permet donc de détecter des anomalies univariées mais aussi des anomalies multivariées ou contextuelles associées à l'observation de changements dans les relations entre paramètres, ce que ne permet pas la détection univariée. La figure 1 (bas) présente un exemple d'anomalie contextuelle. Cette anomalie provient d'un changement de comportement d'un seul paramètre (paramètre du haut dans le rectangle rouge) au regard du second paramètre. Ce fonctionnement est anormal lorsqu'on le compare à un fonctionnement normal (hors rectangle rouge) mais n'est pas détectable facilement à l'aide d'un seul paramètre. Ce problème de détection d'anomalies dans un cadre multivarié a récemment été étudié dans plusieurs articles comme [6], [7] et [8].

Cet article étudie l'intérêt d'un modèle de représentation parcimonieuse introduit dans [9] pour la détection d'anomalies

dans un cadre multivarié et en propose une version pondérée permettant d'intégrer de l'information externe sous forme d'une pondération appropriée et d'améliorer les performances de détection. L'intérêt de la méthode est étudié à travers un exemple simple d'information externe issue d'un coefficient de corrélation et testé sur une base de données composées d'anomalies diverses avec une vérité terrain.

2 Détection d'Anomalies Multivariées

2.1 Prétraitement

Les signaux de télémessure acquis sous forme de séries temporelles sont découpés en fenêtres de données multivariées de taille fixe (ce découpage est souvent appelé "segmentation"). Ce découpage est réalisé de manière glissante et les fenêtres extraites sont chevauchantes. Chaque fenêtre représente un intervalle de temps durant lequel sont enregistrées simultanément les mesures des différents paramètres étudiés. La taille des fenêtres est un paramètre fixé par l'utilisateur. Les fenêtres obtenues après découpage sont des matrices composées d'autant de lignes que de paramètres et d'autant de colonnes que de mesures de chacun d'eux. Chaque matrice de données multivariées est ensuite vectorisée en ligne c'est-à-dire que les séquences de télémessure de chacun des paramètres étudiés (lignes de la matrice) sont concaténées les unes à la suite des autres pour former un vecteur colonne de données de télémessure.

2.2 Représentation parcimonieuse pondérée

Dans notre formalisme, le modèle de décomposition parcimonieuse utilisé pour la détection d'anomalies dans [9] consiste à décomposer un signal test comme la somme de trois signaux : un signal nominal issu d'une décomposition sur un dictionnaire Φ appris au préalable à partir de données sans anomalies, un signal d'anomalie \mathbf{e} décrivant le caractère atypique du signal testé et un signal de bruit \mathbf{b} , i.e.,

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

où $\mathbf{y} \in \mathbb{R}^N$ est le signal observé de taille N , $\Phi \in \mathbb{R}^{N \times L}$ est un dictionnaire connu composé de L atomes (formant les colonnes de Φ), $\mathbf{x} \in \mathbb{R}^L$ est un vecteur parcimonieux de coefficients, $\mathbf{e} \in \mathbb{R}^N$ est un vecteur d'anomalie et $\mathbf{b} \in \mathbb{R}^N$ est un vecteur de bruit contenant les erreurs de modélisation. Etant donné le cadre multivarié dans lequel nous nous plaçons et le prétraitement appliqué, le signal test \mathbf{y} , ainsi que son signal d'anomalie \mathbf{e} sont divisés en P sous-signaux, où P correspond au nombre de paramètres étudiés, i.e., $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_P^T]^T$ et $\mathbf{e} = [\mathbf{e}_1^T, \dots, \mathbf{e}_P^T]^T$, $p = 1, \dots, P$. Le problème de décomposition parcimonieuse peut alors se formaliser mathématiquement de la manière suivante

$$\min_{\mathbf{x}, \mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{x}\|_1 + b \sum_{p=1}^P \|\mathbf{e}_p\|_2 \quad (1)$$

où $\|\mathbf{x}\|_1 = \sum_{i=1}^L |x_i|$ est la norme ℓ_1 du vecteur \mathbf{x} et $\|\mathbf{e}_p\|_2$ est la norme euclidienne du vecteur \mathbf{e}_p . Le modèle (1) impose une double contrainte de parcimonie : la première porte sur le vecteur de coefficients \mathbf{x} et traduit le fait qu'en l'absence d'anomalie le signal \mathbf{y} se décompose de manière parcimonieuse dans

le dictionnaire Φ tandis que la seconde concerne le signal d'anomalie \mathbf{e} et traduit le fait que peu de paramètres sont affectés par une anomalie (ce qui s'exprime à l'aide de la somme des normes ℓ_2 des vecteurs associés aux différents paramètres). La détection d'anomalies est finalement réalisée en comparant la norme de \mathbf{e} à un seuil fixé noté s .

On se place ici dans la phase de détection qui suppose que le dictionnaire Φ a été appris au préalable à partir de données décrivant uniquement des comportements nominaux de la télémessure. L'originalité de cet article est de considérer une version repondérée du problème (1). Le modèle de représentation parcimonieuse pondérée proposé dans cet article pour la détection d'anomalies multivariées s'apparente à un problème de type Group-Lasso [10] qui attribue un poids à chaque groupe de données en considérant ici qu'un groupe réunit l'ensemble des données associées à un paramètre donné. Le problème que nous proposons de résoudre s'écrit alors

$$\arg \min_{\mathbf{x}, \mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{x}\|_1 + b \sum_{p=1}^P \|\mathbf{e}_p\|_2. \quad (2)$$

Ce problème peut être résolu à l'aide d'un l'algorithme découplant de la méthode "Alternating Direction Method of Multipliers" (ADMM) [11] après avoir introduit une variable auxiliaire notée \mathbf{z} et une contrainte d'égalité menant au problème

$$\arg \min_{\mathbf{x}, \mathbf{e}, \mathbf{z}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{z}\|_1 + b \sum_{p=1}^P \|\mathbf{e}_p\|_2 \quad (3)$$

s.c. $\mathbf{z} = \mathbf{x}$

où s.c. signifie "sous la contrainte". Il est intéressant de noter que dans (3) le premier et le second termes sont découplés, ce qui facilite la résolution du problème de minimisation. L'algorithme ADMM est un algorithme itératif qui met à jour les différentes composantes \mathbf{x} , \mathbf{e} , \mathbf{z} de manière alternée par minimisation du Lagrangien augmenté. Plus précisément, chaque composante est actualisée en considérant que les autres composantes sont fixes. Le Lagrangien augmenté de (3) s'écrit

$$\mathcal{L}_A(\mathbf{x}, \mathbf{z}, \mathbf{e}, \mathbf{m}, \mu) = \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{z}\|_1 + b \sum_{p=1}^P \|\mathbf{e}_p\|_2 + \mathbf{m}^T (\mathbf{z} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 \quad (4)$$

où \mathbf{m} est un vecteur contenant les multiplicateurs de Lagrange et μ est un paramètre de régularisation fixant le niveau de pénalité de l'écart entre \mathbf{z} et \mathbf{x} par rapport à la contrainte d'égalité.

Les mises à jour des composantes \mathbf{x} , \mathbf{z} , \mathbf{m} , μ et \mathbf{e} sont détaillées dans [9] et leurs expressions à l'itération $k + 1$ sont rappelées ci-dessous

Mise à jour de \mathbf{x}

$$\hat{\mathbf{x}}^{k+1} = (\Phi \Phi^T + \mu^k I)^{-1} \left[\Phi^T (\mathbf{y} - \mathbf{e}^k) + \mathbf{m}^k + \mu^k \mathbf{z}^k \right]. \quad (5)$$

Mise à jour de \mathbf{z}

$$\hat{\mathbf{z}}^{k+1} = S_\gamma(\mathbf{h}) \text{ avec } \mathbf{h} = \hat{\mathbf{x}}^{k+1} - \frac{1}{\mu^k} \mathbf{m}^k \text{ et } \gamma = \frac{a}{\mu^k}$$

$$[S_\gamma(\mathbf{h})]_i = \begin{cases} \mathbf{h}_i - \gamma & \text{si } \mathbf{h}_i > \gamma \\ 0 & \text{si } |\mathbf{h}_i| \leq \gamma \\ \mathbf{h}_i + \gamma & \text{si } \mathbf{h}_i < -\gamma \end{cases} \quad (6)$$

où $[S_\gamma(\mathbf{h})]_i$ est la $i^{\text{ème}}$ composante de $S_\gamma(\mathbf{h})$ et \mathbf{h}_i est la $i^{\text{ème}}$ composante de \mathbf{h} .

Mise à jour de \mathbf{m}

$$\widehat{\mathbf{m}}^{k+1} = \mathbf{m}^k + \mu^k (\mathbf{z}^{k+1} - \mathbf{x}^{k+1}).$$

Mise à jour de μ

$$\widehat{\mu}^{k+1} = \rho \mu^k; \rho > 1.$$

Mise à jour de \mathbf{e}

La mise à jour du vecteur d'anomalie \mathbf{e} est effectuée à l'aide du problème de minimisation suivant

$$\widehat{\mathbf{e}} = \arg \min_{\mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + b \sum_{p=1}^P \|w_p \mathbf{e}_p\|_2 \quad (7)$$

Le problème (7) peut être découpé en P problèmes qui estiment les sous-vecteurs \mathbf{e}_p séparément

$$\widehat{\mathbf{e}}_p = \arg \min_{\mathbf{e}_p} \frac{1}{2} \|\mathbf{h}_p - \mathbf{e}_p\|_2^2 + b \|w_p \mathbf{e}_p\|_2 \quad (8)$$

où $\mathbf{h} = \mathbf{y} - \Phi \mathbf{x}$ et \mathbf{h}_p est le $p^{\text{ème}}$ signal associé au $p^{\text{ème}}$ paramètre. La solution du problème (8) est classiquement donnée par l'opérateur de seuillage par groupe $Q_b(\mathbf{h}_p)$ suivant

$$\widehat{\mathbf{e}}_p = Q_b(\mathbf{h}_p) = \begin{cases} \frac{\|\mathbf{h}_p\|_2^2 - w_p b}{\|\mathbf{h}_p\|_2^2} \mathbf{h}_p & \text{si } \|\mathbf{h}_p\|_2^2 > b w_p \\ 0 & \text{sinon} \end{cases} \quad (9)$$

L'opérateur Q_b fonde l'estimation des signaux d'anomalies \mathbf{e}_p sur l'erreur de reconstruction et fait intervenir le poids de manière à favoriser la présence d'anomalies lorsque w_p est inférieur à 1 et inversement à empêcher la présence d'anomalies lorsque w_p est supérieur à 1 (quand w_p est supérieur à 1, la première inégalité est difficile à respecter d'où une anomalie estimée égale à 0).

À partir du signal estimé $\widehat{\mathbf{e}}$, on définit un score d'anomalie correspondant à la norme du signal d'anomalie $\|\widehat{\mathbf{e}}\|_2^2$. La détection consiste à comparer ce score d'anomalie à un seuil fixé par l'utilisateur. Une anomalie est déclarée sur le signal testé si le score d'anomalie est supérieur au seuil. Si l'on dispose d'une vérité terrain, ce seuil peut être déterminé à l'aide de caractéristiques opérationnelles du récepteur (courbes COR) qui expriment la probabilité de bonne détection en fonction de la probabilité de fausse alarme. Le seuil choisi permet d'obtenir un compromis entre la probabilité de détection souhaitée et la probabilité de fausses alarmes acceptée relatives au contexte d'application. Dans ces travaux, le seuil choisi est celui associé au couple (P_{FA}, P_D) le plus proche du point $(0,1)$. Le choix du seuil est plus difficile si l'on ne dispose pas d'une vérité terrain. Dans ce cas, c'est l'expérience de l'utilisateur qui permettra de déterminer le seuil le plus approprié au contexte d'application.

2.3 Pondération avec coefficient de corrélation

Les poids w_p sont construits à partir d'informations externes de manière à intégrer de la connaissance métier ou d'utiliser d'autres indicateurs de présence d'anomalies que l'erreur de reconstruction sur laquelle repose notre problème d'estimation. Ils peuvent être fixés par l'utilisateur ou mis à jour à chaque

itération de l'algorithme ADMM. À titre d'exemple, nous nous sommes intéressés au coefficient de corrélation entre le signal testé \mathbf{y} et sa reconstruction $\Phi \mathbf{x}$ obtenue par décomposition sur le dictionnaire. Les poids w_p sont alors mis à jour à chaque itération de l'algorithme ADMM au même titre que le vecteur \mathbf{x} à partir duquel le coefficient de corrélation est calculé. Une faible valeur du coefficient de corrélation indique que le signal reconstruit est peu lié au signal d'origine, ce qui correspond généralement à la présence d'une anomalie.

Comme expliqué précédemment, le poids w_p intervient dans l'estimation du signal d'anomalie \mathbf{e}_p de manière à pénaliser la détection s'il a une forte valeur et à favoriser cette détection si cette valeur est faible. De manière à respecter cette propriété, nous proposons de considérer la fonction suivante

$$f_\alpha : [-1, 1] \rightarrow \mathbb{R}^+ \\ c_p \mapsto w_p = \frac{1}{((1+\alpha)-c_p)^2}$$

qui est une fonction croissante de c_p , où c_p est le coefficient de corrélation entre le paramètre considéré et sa reconstruction et α est un paramètre fixé. Le choix de cette fonction f est motivé par le fait que c'est une fonction simple qui ne dépend que d'un paramètre α . Notons que pour $c_p = \alpha$, on a $w_p = 1$ (ce qui correspond à l'algorithme non pondéré), que pour $c_p < \alpha$, on a $w_p < 1$ (l'algorithme pondéré favorise la présence d'une anomalie par rapport à l'algorithme non pondéré), et enfin que pour $c_p > \alpha$, on a $w_p > 1$ (l'algorithme pondéré limite la présence d'une anomalie par rapport à l'algorithme non pondéré). De plus, nous avons observé que faire varier α dans l'intervalle $] -1, +1[$ était suffisant dans nos applications pratiques.

3 Résultats

La méthode de détection d'anomalies de [9] et sa version pondérée faisant l'objet de cet article ont été testées à l'aide d'une base de données contenant des anomalies connues. Le dictionnaire a été appris à l'aide de l'algorithme K-SVD [12] à partir de deux mois de télémessures associées à un fonctionnement normal du système. Les données de test se composent de 1000 signaux (associés à 18 jours) dont 82 contiennent des anomalies connues et identifiées par des experts. Les anomalies ont été observées dans 5 fenêtres temporelles et la majorité d'entre elles correspondent à de réelles anomalies observées dans des données satellites. Pour ces premiers tests, 7 paramètres de télémessure ont été choisis. Notons qu'une attention particulière a été portée à la construction d'une base d'anomalies hétérogène composée d'anomalies de toutes sortes. Trois des 5 périodes d'anomalies sont représentées dans la figure 1. Les paramètres de régularisation a et b des différents modèles (1) et (2) ont été fixés par validation croisée à l'aide de courbes COR construites grâce à la vérité terrain disponible.

La table I résume les résultats de détection en termes de probabilités de détection et de fausse alarme obtenues avec les signaux tests grâce à la vérité terrain. Ces premiers résultats montrent l'intérêt de la représentation parcimonieuse pour la détection d'anomalies avec 89% de probabilité de bonne détection pour 5.3% de probabilité de fausses alarmes. L'intégration

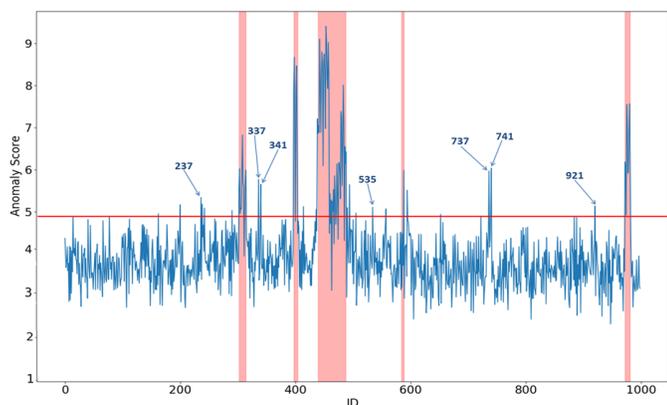


FIGURE 2 – Scores d’anomalie de l’algorithme non pondérée. (Seuil = 4.9 déterminé à l’aide de courbes COR et représenté par une droite horizontale rouge)

d’informations externes par l’intermédiaire d’une pondération adaptée construite à partir du coefficient de corrélation a permis d’améliorer ces résultats avec 89% de probabilité de bonne détection pour 2.2% de probabilité de fausse alarme. Notons que réduire la probabilité de fausse alarme est intéressant pour l’application visée car chaque anomalie détectée est vérifiée par l’expert, ce qui peut représenter un temps important.

TABLE 1 – Résultats de détection en termes de probabilité de détection (P_D), de probabilité de fausse alarme (P_{FA}) et d’aire sous la courbe COR (AUC).

Méthode	Seuil	P_D	P_{FA}	AUC
ADMM [9]	4.9	89%	5.32%	0.93
ADMM Pondéré	4	89%	2.18%	0.96

Les figures 2 et 3 montrent les scores d’anomalie attribués aux signaux tests par la méthode de détection d’anomalies étudiée et sa version pondérée. Les périodes d’anomalie connues à partir de la vérité terrain sont indiquées en rouge. Les seuils fixés donnés dans la table I sont représentés par les droites horizontales rouges. Ces résultats mettent en évidence l’intérêt de la représentation parcimonieuse pour la détection d’anomalies. En effet, on observe que toutes les anomalies sont correctement détectées notamment l’anomalie contextuelle correspondant à la dernière période d’anomalie (Figure 1, en bas). De plus, en comparant les scores retournés par les algorithmes avec et sans pondération, on observe que l’utilisation de la pondération a permis de faire chuter certains scores en dehors des périodes d’anomalies et donc de limiter les fausses alarmes. C’est le cas notamment des signaux tests 237, 535 et 921 qui sont repérés dans les figures 2 et 3 par des flèches bleues. Notons que certaines fausses alarmes subsistent pour les signaux tests 337, 341, 737 et 741.

4 Conclusion

Cet article a montré l’intérêt d’utiliser une représentation parcimonieuse pour la détection d’anomalies dans un cadre multivarié, notamment à l’aide d’une version pondérée qui utilise des poids construits à partir d’informations externes et intervenant dans le processus de détection. À titre d’exemple, nous

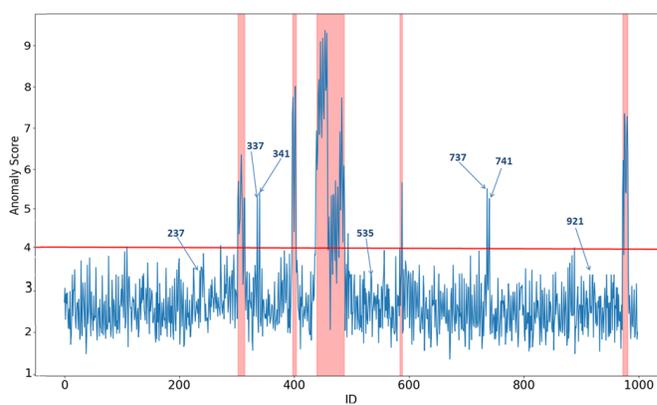


FIGURE 3 – Scores d’anomalie de l’algorithme pondérée. (Seuil = 4 déterminé à l’aide de courbes COR et représenté par une droite horizontale rouge)

avons montré qu’en utilisant un coefficient de corrélation adapté, il est possible d’améliorer les performances de l’algorithme en pénalisant la détection d’anomalies de manière à limiter le nombre de fausses alarmes. Pour aller plus loin, il serait intéressant d’étudier l’intérêt d’ajouter des informations externes issues d’une connaissance métier fournie par les experts travaillant dans le domaine avec la télémétrie.

5 Remerciements

Les auteurs souhaitent remercier Pierre-Baptiste Lambert du CNES et Clémentine Barreyre d’Airbus Defence and Space pour leurs discussions fructueuses sur la détection des anomalies dans la télémétrie des satellites. Ce travail a été soutenu par le CNES et Airbus Defence and Space.

Références

- [1] V. Chandola and *al.*, “Anomaly detection : A survey,” *ACM Computing Survey*, vol. 43, no. 3, pp. 1–58, 2009.
- [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] K. Hundman and *al.*, “Detecting spacecraft anomalies using LSTMs and non-parametric dynamic thresholding,” in *Proc. Int. Conf. Knowledge Data Mining (KDD’18)*, London, United Kingdom, Aug. 2018, pp. 387–395.
- [4] J. Heras and *al.*, “Project sybil : A novelty detection system for human spaceflight operations,” in *Proc. Int. Conf. Space Operations (SpaceOps’2012)*, Stockholm, Sweden, Jun. 2012.
- [5] 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.
- [6] N. Takeishi and *al.*, “Anomaly detection from multivariate times-series with sparse representation,” in *Proc. IEEE Int. Conf. Syst. Man and Cybernetics*, San Diego, CA, USA, Oct. 2014.
- [7] T. Yairi and *al.*, “A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1384–1401, Jun. 2017.
- [8] 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.
- [9] A. Adler and *al.*, “Sparse coding with anomaly detection,” *J. Signal Process. Syst.*, vol. 79, no. 2, pp. 179–188, May 2015.
- [10] 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.
- [11] 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.
- [12] 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.