Estimation parcimonieuse et apprentissage de dictionnaires pour la détection d'anomalies multivariées dans des données mixtes de télémesure satellites

Barbara Pilastre

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Mathématiques Appliquées

Présentée et soutenue par :

Mme BARBARA PILASTRE

le vendredi 6 novembre 2020

Titre :

Estimation parcimonieuse et apprentissage de dictionnaires pour la détection d'anomalies multivariées dans des données mixtes de télémesure satellites

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur(s) de Thèse :

M. JEAN-YVES TOURNERET

Rapporteurs :

M. JERÔME MARS, INP GRENOBLE

M. PAULO GONCALVES, ENS SCIENCES LYON

Membre(s) du jury :

Mme GERSENDE FORT, UNIVERSITE TOULOUSE 3, Président

M. GUILLAUME GINOLHAC, POLYTECH ANNECY-CHAMBERY, Membre

M. JEAN-YVES TOURNERET, TOULOUSE INP, Membre

M. LOÏC BOUSSOUF, AIRBUS DEFENCE AND SPACE, Membre

Mme MAGALI FROMONT, UNIVERSITE RENNES 2, Invité

M. STEPHANE D'ESCRIVAN, CENTRE NATIONAL D'ETUDES SPATIALES CNES, Membre

Remerciements

La thèse s'annonce souvent comme une aventure en solitaire. Pourtant, à l'issue de ces trois années je suis convaincue que je n'aurais jamais pu mener à bien ce travail de recherche sans le soutien de nombreuses personnes que je tiens à remercier ici.

En premier lieu, je tiens à remercier le CNES et Airbus Defence and Space, nos partenaires industriels, qui ont rendu ce projet de thèse possible grâce à leur co-financement. En particulier, côté CNES, j'aimerais remercier Sylvain Fuertes pour m'avoir fait confiance et permis de réaliser mon stage de fin d'étude au CNES et pour m'avoir encouragé à me lancer dans la thèse. J'aimerais également remercier Béatrice Deguine, une de nos bienfaitrices, pour son soutien, ses conseils et sa bienveillance. Je remercie également Stéphane D'Escrivan pour son encadrement et son implication tout au long de la thèse et Pierre-Baptiste Lambert pour son temps et son intérêt pour nos travaux. Plus généralement, je remercie toute l'équipe DNO/OP/MSO que j'ai toujours eu grand plaisir à retrouver lors de mes passages au CNES (Sophie, Loïc, Charles, Albert, ...). Côté Airbus Defence & Space je souhaite remercier Romaric Redon pour avoir soutenu ce projet de thèse. Je remercie également Loïc Boussouf, qui en plus de son encadrement et son enthousiasme pour nos travaux à beaucoup travaillé pour intégrer ce projet dans un contexte industriel. Je remercie également Clémentine Barreyre pour nous avoir apporté son expertise tout au long de la thèse. Plus généralement je remercie toute l'équipe TESOE (Anaïs, Linda, Jules, Baptiste, Meguy,...) pour m'avoir accueilli aussi chaleureusement (J'espère intégrer une équipe comme la vôtre !).

Je tiens à remercier tout particulièrement mon directeur de thèse et mentor Jean-Yves Tourneret pour qui j'ai beaucoup d'admiration et d'affection. Beaucoup d'admiration pour ses qualités scientifiques qui en font un chercheur de renom et pour l'humilité et la simplicité qu'il a su garder ; et beaucoup d'affection car il a toujours été présent et disponible pour me guider, m'encourager, me rassurer parfois (souvent) et mettre toute son expérience au service de ma réussite en m'accompagnant aux quatre coins du monde en quête de Best Papers et de collaboration fructueuse. Et la mission est réussie, merci beaucoup Jean-Yves. Tu as eu et tu auras encore de nombreux thésards, moi je n'aurai eu qu'un seul directeur de thèse et je n'en aurai pas voulu un autre.

Je remercie Jérome Mars et Paulo Gonçalves pour avoir rapporté ma thèse et pour l'intérêt démontré pour nos travaux, ainsi que Gersende Fort pour avoir présidé le jury, Guillaume Ginolhac pour sa participation, et Magalie Fromont pour m'avoir suivi tout au long de mon parcours académique et pour avoir accepté l'invitation.

Une partie de ces travaux de thèse a été réalisée en collaboration avec deux chercheurs du Pérou que j'apprécie beaucoup et avec qui j'ai aimé travailler. Je remercie donc Paul Rodriguez pour m'avoir accueilli à Lima, dans son laboratoire, et pour nos échanges scientifiques (ou non) devant un tableau blanc ou sur le chemin de l'université. Je remercie également Jorge et Gustavo pour m'avoir fait découvrir le Pérou de l'intérieur. Je remercie plus particulièrement Gustavo qui a cru en nos idées et m'a beaucoup aidé dans ce travail de recherche.

La réussite de cette thèse est en grande partie due à l'environnement dans lequel elle a été réalisée et je profite donc de ces quelques lignes pour remercier toutes la famille TéSA au sein de laquelle j'ai grandi pendant ces trois années. Il serait impossible de ne pas remercier notre directrice Corinne qui veille sur nous tous au quotidien, Isa dont les jeux de mots et la folie vont me manquer, Raoul que je ne manquerai pas d'appeler pour savoir comment aménager mon nouvel appartement, Philippe que j'aurai peut-être la chance de recroiser lors de ses escapades en terre bretonne, Serge, Patrice (un dernier petit calcul et on s'en va), Jacques dont on ne se lasse pas de ses anecdotes (comme celle de la pâte à colmater fabriquée à partir de noyaux d'abricots), Corentin notre plombier en chef avec qui on aime battre le carton, Julien, surement l'un des plus grands chercheurs de demain, Quentin (désolée j'ai pas pu venir au taekwondo, mon vélo est cassé...), Selma alias cocotte jolie la pétillante (surtout ne perd pas tes bulles), Charles-U (merci pour ces joyeuses pauses déj' à la Sodexo), et bien sûr, mes copains de bureaux et complices des premières heures : Lorenzo la machine à écrire des "papières" et mon plus fidèle camarade du best bureau of TéSA (pas comme ces gars du réseau...), Victor Badoit notre petit stagiaire devenu mamie gâteaux, sculpteur, serrurier et thésard (à ses heures perdues), et Adrien mon compagnon de route (je t'attends à l'arrivée mon pote). Je tiens également à remercier toutes les personnes que j'ai pu croiser et qui représentent autant de belles rencontres qui ont marqué ces trois années : Julian Tachella, Jordi Vila, Yoann Altmann, Olivier et Hélène Besson, Florian Simatos, Nelson Diaz, Adrian et Daiana Basarab, Antoine Auger, Florian Mouret, Oumaima El Mansouri,...

Je remercie également ma famille en commençant par mes parents, qui n'ont pas toujours été d'accords entre eux mais qui ont toujours tout fait pour nous, mon petit frère dont je suis très fière, et ma mamie, une force de la nature au coup de pédale si singulier. Merci de m'avoir toujours soutenu dans mes choix et encouragé à suivre mes idées. J'aurai également un petit mot pour mes amis normands. Je suis fière de ce que nous sommes tous devenus mais surtout très heureuse de voir qu'après toutes ces années nous sommes toujours présents les uns pour les autres et surtout, que la roue des sujets n'a pas pris une ride. Postés derrière votre écran, aucun de vous n'a raté ce rendez-vous. Merci les amis. Merci Hugo (prépare ton baluchon de slips sales on va au marché), Felix l'enfant terrible de la rosette, Nicolas notre dandy parisien qui construit des péages en Espagne (chacun son truc), Paul qui aime se jouer de moi mais sur qui je le sais, je peux toujours compter, Tiffany que je vois malheureusement trop peu mais qui est toujours là dans les moments importants, Théo notre poète des temps modernes, Arthur que j'ai retrouvé au Gato Tulipan après des années mais comme si c'était hier, Océane notre blondinette qui n'a pas froid aux yeux, Justine mon amie de toujours (L'unité B, Melbourne, le Pérou, what's next ?) et sa famille (Sylvie, RV et Aurélien) qui était presque ma deuxième famille et avec qui j'ai beaucoup de bons souvenirs (et seulement un mauvais : la face de Bellevarde).

Enfin, mes derniers mots vont à mon binôme dans la vie, Tristan, dont le soutien indéfectible m'a permis d'atteindre ce que je pensais hors de ma portée. Ces quelques années écoulées témoignent de tout ce qui nous est possible de réaliser ensemble et ce n'est, je nous le promets, qu'un début.

Résumé — La surveillance automatique de systèmes et la prévention des pannes sont des enjeux majeurs dans de nombreux secteurs et l'industrie spatiale ne fait pas exception. Par exemple, le succès des missions des satellites suppose un suivi constant de leur état de santé réalisé à travers la surveillance de la télémesure. Les signaux de télémesure sont des données issues de capteurs embarqués qui sont reçues sous forme de séries temporelles décrivant l'évolution dans le temps de différents paramètres. Chaque paramètre est associé à une grandeur physique telle qu'une température, une tension ou une pression, ou à un équipement dont il reporte le fonctionnement à chaque instant.
Alors que les approches classiques de surveillance atteignent leurs limites, les méthodes d'apprentissage automatique (machine learning en anglais) s'imposent afin d'améliorer la surveillance de la télémesure via un apprentissage semi-supervisé : les signaux de télémesure associés à un fonctionnement normal du système sont appris pour construire un modèle de référence auquel sont comparés les signaux de télémesure récemment acquis. Les méthodes récentes proposées dans la littérature ont permis d'améliorer de manière significative le suivi de l'état de santé des satellites mais elles s'intéressent presque exclusivement à la détection d'anomalies univariées pour des paramètres physiques traités indépendamment. L'objectif de cette thèse est de proposer des algorithmes pour la détection d'anomalies multivariées capables de traiter conjointement plusieurs paramètres de télémesure associés à des données de différentes natures (continues/discrètes), et de prendre en compte les corrélations et les relations qui peuvent exister entre eux.
L'idée motrice de cette thèse est de supposer que la télémesure fraîchement reçue peut être estimée à partir de peu de données décrivant un fonctionnement normal du satellite. Cette hypothèse justifie l'utilisation de méthodes d'estimation parcimonieuse et d'apprentissage de dictionnaires qui seront étudiées tout au long de cette thèse. Une deuxième forme de parcimonie propre aux anomalies satellites a également motivé ce choix, à savoir la rareté des anomalies satellites qui affectent peu de paramètres en même temps. Dans un premier temps, un algorithme de détection d'anomalies multivariées basé sur un modèle d'estimation parcimonieuse est proposé. Une extension pondérée du modèle permettant d'intégrer de l'information externe est également présentée ainsi qu'une méthode d'estimation d'hyperparamètres qui a été developpée pour faciliter la mise en oeuvre de l'algorithme. Dans un deuxième temps, un modèle d'estimation parcimonieuse avec un dictionnaire convolutif est proposé. L'objectif de cette deuxième méthode est de contourner le problème de non-invariance par translation dont souffre le premier algorithme. Les différentes méthodes proposées sont évaluées sur plusieurs cas d'usage industriels associés à de réelles données satellites et sont comparées aux approches de l'état de l'art.

Mots clés : Surveillance de la Télémesure Satellites, Détection d'Anomalies, Estimation Parcimonieuse, Apprentissage de Dictionnaire.

Abstract — Spacecraft health monitoring and failure prevention are major issues in many fields and space industry has not escaped to this trend. Indeed, the proper conduct of satellite missions involves ensuring satellites good health and detect failures as soon as possible. This important task is performed by analyzing housekeeping telemetry data using anomaly detection methods. Housekeeping telemetry consist of sensors data recorded on board and received as time series describing the time evolution of various parameters. Each parameter is associated with physical quantity such as a temperature, a voltage or a pressure, or an equipement status.

As conventional monitoring methods reach their limits, statistical machine learning methods have been studied to improve satellite telemetry monitoring via a semi-supervised learning : telemetry associated with normal operations of the spacecraft is learned to build a reference model. Then, more recent data is compared to this model in order to detect any potential anomalies. Most of the methods recently proposed focus on univariate anomaly detection for continuous parameters and handle telemetry parameters independently remove. The purpose of this thesis is to propose algorithms for multivariate anomaly detection which can handle mixed telemetry parameters jointly and take into account the correlations and relationships that may exist between them in order to detect univariate and multivariate anomalies.

In this work we assume that telemetry signals can be approximated using few telemetry signals associated with normal satellite operations. This first hypothesis of sparsity justifies the use of sparse representation methods that will be studied throughout this thesis. This choice is also motivated by a second form of sparsity which is specific to satellite anomalies and reflect the fact that anomalies are rare and affect few parameters at the same time. In a first time, a multivariate anomaly detection algorithm based on a sparse estimation model is proposed. A weighted extension of the method which integrates external information is presented as well as a hyperparameter estimation method that has been developed to facilitate the operationnal use of the algorithm. In a second step, a sparse estimation model with a convolutional dictionary is proposed. The objective of this second method is to exploit the shift-invariance property of convolutional dictionnaries and improve the detection. The proposed methods are finally evaluated on industrial use cases associated with real telemetry data and are compared to state-of-the-art approches.

Keywords : Spacecraft Telemetry Monitoring, Anomaly Detection, Sparse Representation, Dictionary Learning.

Table des matières

Table des figures

Liste des tableaux

Introduction

Contexte et problématique de la thèse

Cette thèse s'inscrit dans une volonté de la part du CNES et Airbus d'améliorer le suivi de l'état de santé des satellites en exploitation. Les succès récents des techniques d'apprentissage automatique dans de nombreux domaines ont motivé leur application pour la surveillance des satellites et plus précisément pour la détection d'anomalies dans la télémesure. La plupart des méthodes de l'état de l'art assurent une détection d'anomalies univariées et s'intéressent presque exclusivement aux paramètres physiques qu'elles traitent de manière indépendante. Cette thèse étudie l'intérêt des méthodes de détection d'anomalies multivariées capables de traiter conjointement plusieurs signaux de télémesure de différentes natures. L'objectif de ces approches est d'offrir une surveillance plus globale du satellite par la prise en compte des corrélations et relations qui peuvent exister entre les différents paramètres et qui décrivent le contexte opérationnel des satellites.
Une première idée a été de supposer que la nouvelle donnée acquise peut être estimée à l'aide de peu de données associées à un fonctionnement normal du satellite. Cette hypothèse justifie l'application de méthodes d'estimation parcimonieuse et d'apprentissage de dictionnaires qui seront étudiées tout au long de cette thèse. Un premier algorithme, appelé ADDICT (pour Anomaly Detection using a DICTionary), basé sur un modèle d'estimation parcimonieuse a été proposé pour la détection d'anomalies dans des données de télémesure mixtes (discrètes et continues). Par la suite, plusieurs pistes d'amélioration de la méthodes ont été étudiées pour :

- Limiter les fausses alarmes à l'aide d'une pondération adaptée ou d'une généralisation du modèle proposé.
- Assurer une reconstruction invariante par translation via l'introduction d'un dictionnaire convolutif.
- Régler les hyperparamètres du modèle de manière automatique.

Ce travail de thèse a été mené au laboratoire TéSA et a fait l'objet d'un co-financement du CNES et de Airbus Defence and Space.

Organisation du manuscrit

Le premier chapitre pose le contexte de cette thèse. Après avoir présenté les satellites artificiels, leurs applications, et surtout les données de télémesure qu'ils génèrent pour assurer la surveillance du bon déroulement des opérations, ce chapitre passe en revue quelques algorithmes récents de détection d'anomalies satellites qui reposent sur des techniques d'apprentissage semi-supervisé et qui ont été développés par différents acteurs du spatial.

Le deuxième chapitre étudie une nouvelle méthode de détection d'anomalies multivariées

capable de traiter conjointement plusieurs paramètres de télémesure de différentes natures. En particulier, ce chapitre présente l'algorithme ADDICT, un nouvel algorithme qui repose sur un modèle d'estimation parcimonieuse permettant de représenter la nouvelle donnée acquise à partir d'un dictionnaire de données passées associées à un fonctionnement normal (sans anomalies) du satellite, et d'en déduire la présence ou l'absence d'anomalies dans ces nouvelles données. Le modèle mathématique proposé introduit alors une double contrainte de parcimonie afin de 1) limiter la quantité de données utilisées pour l'estimation à l'aide d'une norme ℓ_l, et 2) formaliser le problème de détection d'anomalies à l'aide d'une pénalisation ℓ_2 de type group-LASSO. L'introduction de cette parcimonie structurée est motivé par le fait que les anomalies satellites sont rares et affectent peu de paramètres en même temps. Des variantes de l'algorithme sont également présentées comme par exemple une version pondérée permettant d'améliorer les performances de détection par l'introduction d'une pondération adaptée construite à l'aide d'informations externes. Par ailleurs, une méthode d'estimation des hyperparamètres basée sur la théorie des machines à vecteurs supports est également étudiée afin d'assurer un réglage automatique de ces derniers et faciliter l'utilisation de l'algorithme. Les résultats obtenus sur une base de données de test composées d'anomalies diverses et munie de sa vérité terrain sont encourageants et compétitifs vis à vis des méthodes de l'état de l'art. Cependant, les méthodes classiques d'estimation parcimonieuse souffrent d'un manque d'invariance par translation et retournent de nombreuses fausses alarmes. Ce problème a pu être partiellement réglé dans ce chapitre en introduisant une option d'invariance par translation. L'objectif du prochain chapitre est d'étudier l'intérêt d'un dictionnaire convolutif pour résoudre ce problème et assurer une détection d'anomalies multivariées plus efficace.

Le troisième chapitre présente une extension de l'algorithme ADDICT permettant d'assurer une invariance par translation et de traiter conjointement, et selon une unique stratégie, plusieurs signaux de télémesure mixtes. Plus précisément, ce chapitre s'intéresse aux méthodes d'estimation parcimonieuse convolutive et d'apprentissage de dictionnaire convolutif. Le modèle convolutif proposé a permis d'améliorer de manière significative les performances de détection de l'algorithme. Par ailleurs, il a pu être étendu au problème d'apprentissage de dictionnaires. Ces travaux ont été menés en collaboration avec Gustavo Silva et Paul Rodriguez de la Pontificia Universidad Catolica del Peru (PUCP).

Le dernier chapitre, plus applicatif, traite plusieurs cas d'usage industriels associés à des données réelles de télémesure satellite. D'une part, l'objectif de ce chapitre est de tester et d'évaluer les performances des méthodes proposées dans des contextes opérationnels variés décrivant un réel besoin métier. D'autre part, une stratégie d'apprentissage séquentiel ou "en ligne" est étudiée pour l'algorithme ADDICT et ses extensions afin de prendre en compte le retour d'expérience des utilisateurs et leur expertise métier et de limiter les fausses alarmes.

Principales contributions

Les principales contributions de cette thèse sont exposées ci-dessous.

Chapitre 1

Ce chapitre présente les satellites artificiels ainsi que les données de télémesure permettant d'assurer le bon déroulement de leur mission et la prévention des anomalies. Dans un second temps, plusieurs méthodes récentes de détection d'anomalies basées sur des techniques d'apprentissage semi-supervisé sont détaillées.

Chapitre 2

Ce chapitre formule le problème de la détection d'anomalies multivariées sous la forme d'un problème d'estimation parcimonieuse qui impose une double parcimonie : une parcimonie assurée par une pénalisation ℓ_1 et une parcimonie de groupe assurée par une pénalisation de type ℓ_2. Ces travaux ont abouti au développement de l'algorithme ADDICT (Anomaly Detection using a DICTionary) qui a fait l'objet d'une publication dans un revue scientifique Pilastre et al. (2020a). D'autre part, plusieurs variantes de l'algorithme proposé ont été étudiées telles qu'une version pondérée permettant d'intégrer de l'information externe et d'améliorer les performances de détection à l'aide d'une pondération adaptée. Cette dernière a fait l'objet d'une communication dans une conférences francophone Pilastre et al. (2019b) et une conférence internationale Pilastre et al. (2019c) lors de laquelle elle a été récompensée (paper award). Par ailleurs, une généralisation du problème mathématique et une méthode d'estimation des hyperparamètres du modèle sont étudiées.

Chapitre 3

Ce chapitre étudie l'intérêt des méthodes d'estimation parcimonieuse convolutive pour la détection d'anomalies multivariées et présente une extension invariante par translation de l'algorithme ADDICT. Le modèle proposé à pu être étendu au problème d'apprentissage de dictionnaire convolutif. Ces travaux, menés en collaboration avec Gustavo Silva et Paul Rodriguez de la Pontificia Universidad Catolica del Peru (PUCP) ont fait l'object d'une publication dans une conférence internationale Pilastre et al. (2020b).

Chapitre 4

Ce chapitre présente quelques résultats expérimentaux obtenus pour les algorithmes proposés dans cette thèse appliqués à différents cas industriels de surveillance de la télémesure

satellite. Ces cas d'usage métier, associés à des données réelles satellite, sont autant d'occasions d'évaluer l'intérêt et les performances de ces méthodes dans des contextes qui décrivent les besoins industriels de suivi de l'état de santé des satellites. En particulier, ce chapitre aborde le problème d'apprentissage séquentiel ou "en ligne" pour la détection d'anomalies et tente d'y répondre à l'aide de l'algorithme ADDICT et de ses extensions.

Publications

Article de revue

- Pilastre, Barbara, Boussouf, Loïc, D'Escrivan, Stéphane, Tourneret, Jean-Yves, "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning". *Signal Processing*, vol. 168, pp. 107320, March 2020.

Conférences

- Pilastre, Barbara, Boussouf, Loïc, D'Escrivan, Stéphane, Tourneret, Jean-Yves, "Représentation Parcimonieuse Pondérée pour la Détection d'Anomalies dans des Signaux Multivariés". Dans *Proc. du Groupe d'Etude du Traitement du Signal et des Images (GRETSI'2019)*, Lille, France, Aout, 2019.
- Pilastre, Barbara, Boussouf, Loïc, D'Escrivan, Stéphane, Tourneret, Jean-Yves, "Spacecraft Health Monitoring using a Weighted Sparse Decomposition". Dans *World Congress on Condition Monitoring (WCCM'2019)*, Singapore, December 2019. (Best paper)
- Pilastre, Barbara, Boussouf, Loïc, D'Escrivan, Stéphane, Tourneret, Jean-Yves, "Multivariate Anomaly Detection in Mixed Telemetry Time-Series Using a Sparse Decomposition", dans *Proc. of 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'19)*, pages 430-434, Le gosier, Guadeloupe, Dec. 2019.
- Pilastre, Barbara, Silva, Gustavo, Boussouf, Loïc, D'Escrivan, Stéphane, Rodriguez, Paul, Tourneret, Jean-Yves, "Anomaly Detection in Mixed Time-Series using a Convolutional Sparse Representation with Application to Spacecraft Health Monitoring. Dans *Proc. of IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP'2020)*, Barcelona, Spain, May 2020.

Chapitre 1

Suivi de l'état de santé des satellites en exploitation

Sommaire

1.1 Introduction

Ce chapitre a pour objectif de décrire le contexte d'application de cette thèse, à savoir le suivi de l'état de santé des satellites en exploitation à travers la surveillance de la télémesure par l'application de méthodes d'apprentissage automatique (ou machine learning). Après une présentation générale des satellites et de leurs applications, nous nous intéresserons aux données qu'ils génèrent afin de permettre un suivi permanent du bon déroulement de leurs missions. Enfin, nous passerons en revue quelques méthodes récentes basées sur des techniques d'apprentissage semi-supervisé et développées par différents acteurs du spatial pour la surveillance de la télémesure satellite.

1.2 Les satellites artificiels

1.2.1 Historique

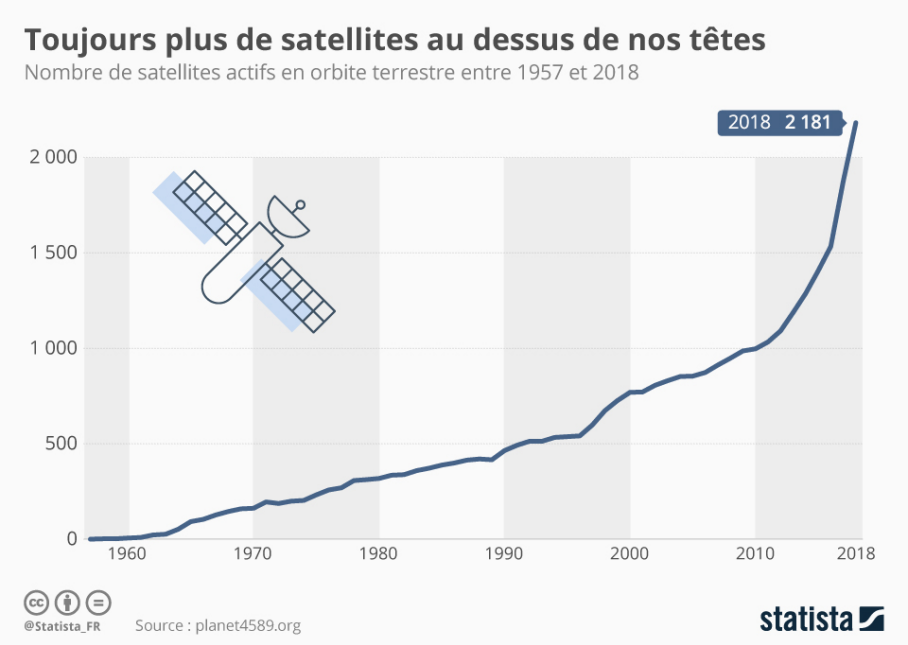

FIGURE 1.1 – Illustration de l'évolution du nombre de satellites actifs en orbite terrestre entre 1957 et 2018.

Par définition, un satellite est un corps qui gravite autour d'une étoile ou d'une planète. Il en existe deux types : les satellites naturels, telle que la Lune qui est un satellite de la Terre, et les satellites artificiels construits par l'homme et porteurs d'équipements à des fins scientifiques, industrielles, militaires etc. Nous nous intéresserons dans cette thèse aux satellites artificiels de la Terre.
L'idée de placer un satellite en orbite émerge dès la fin de la seconde guerre mondiale. Les applications imaginées sont alors militaires. En effet, un rapport de faisabilité de 1946 Clauser

et al. (1946), resté longtemps classifié, évoquait l'intérêt d'un satellite pour la reconnaissance des territoires ennemis et l'acquisition de données météorologiques utiles à la planification des opérations militaires. Il faut attendre le 4 octobre 1957 pour que le premier satellite artificiel, Spoutnik 1, soit lancé par l'URSS qui devient alors première puissance spatiale. Cet évènement historique marque le début de l'ère spatiale et de la course entre l'URSS et les États-Unis pour la conquête de l'espace (voir Lehot and Clervoy (2015) pour plus d'information sur le contexte historique).
Aujourd'hui, l'utilisation des satellites s'est banalisée et le nombre de satellites en orbite autour de la Terre ne cesse de croître, notamment ces dernières années durant lesquelles l'évolution du nombre de satellites actifs autour de la Terre a pris une allure exponentielle. En effet, comme présenté dans la figure 1.1, le nombre de satellites actifs en orbite terrestre était de 2181 en 2018 contre 997 en 2010. Ce chiffre devrait encore augmenter avec l'arrivée de gigantesques flottes de satellites telle que la méga-constellation du projet Starlink de SpaceX qui envisage l'envoi de 42 000 micro-satellites afin de créer un accès internet mondial depuis l'espace.

1.2.2 Les orbites

L'orbite d'un satellite artificiel gravitant autour de la terre correspond à sa trajectoire autour de cette dernière. Elle définit son altitude, sa vitesse ou encore sa période de révolution et est donc choisie afin de répondre au mieux aux besoins de sa mission. Le choix de l'orbite dicte la façon dont va pouvoir être assuré le suivi, notamment au sol, de l'état de santé du satellite car il fixe par exemple les conditions de réception par les stations au sol des données acquises à bord. Les stratégies de surveillance peuvent alors varier d'un satellite à l'autre. Par ailleurs, outre les spécificités liées à la mission du satellite qui peuvent requérir une orbite particulière, les données produites peuvent dépendre de l'environnement du satellite, ce dernier étant caractéristique de l'orbite choisie. Á titre d'exemple nous pouvons citer les mesures de température fortement influencées par le niveau d'ensoleillement du satellite qui dépend lui-même de l'orbite choisie. Parmi les orbites les plus utilisées nous citerons les orbites basses et l'orbite géostationnaire :

- Les orbites basses (aussi appelées LEO pour Low Earth Orbit) sont des orbites situées entre 180 et 2 000km d'altitude. Ce sont des orbites privilégiées par les satellites d'observation car il permettent de bénéficier d'une haute résolution optique. Enfin, ces orbites permettent de profiter des orbites héliosynchrones qui assurent un éclairement identique de la scène lorsque le satellite passe au-dessus d'un point donné de la Terre. Cette propriété est particulièrement intéressante pour la surveillance de zones précises. On trouve sur des orbites basses des satellites d'observation, certains satellites de télécommunication, mais également des satellites scientifiques, météorologiques ou encore la station spatiale internationale (ISS). Quelques orbites basses actuellement utilisées sont présentées dans la figure 1.2. La communication avec un satellite en orbite basse ne peut être établie que lorsque que celui-ci survole une station au sol. Cependant, la communication est rarement établie à chaque passage car seuls quelques passages par jour suffisent pour mener à bien la mission du satellite et récupérer les données

acquises à bord. De plus, la liaison nécessite de réserver les créneaux d'utilisation des antennes et leur télégestion, ce qui a un coût. Le nombre de passages nécessaires à la récupération des données dépend du volume de données et du débit de la voie descendante.

- Les orbites géostationnaires (aussi appelés GEO) sont des orbites circulaires autour de la terre, d'inclinaison orbitale nulle (dans le plan de l'équateur) et de vitesse de rotation égale à celle de la Terre (soit 23H 54min 4s). Un satellite placé sur l'orbite géostationnaire Terrestre est situé à exactement 35 786km d'altitude (on parlera couramment de satellite à 36 000 km) et reste en permanence au-dessus du même point de l'équateur. Cette propriété est intéressante pour couvrir une zone fixe, ce qui en fait l'orbite privilégiée des satellites de télécommunication et de certains satellites d'observation tels que les satellites météo. Du fait de son altitude, un satellite geostationnaire couvre un tiers de la surface du globe ce qui peut être intéressant pour étudier de larges zones. Cependant, la mise à poste d'un satellite sur une orbite géostationnaire nécessite des lanceurs puissants et la résolution optique est logiquement moins bonne que celle des orbites basses. L'orbite géostationnaire est présentée dans la figure 1.2. Les satellites geostationnaires, fixes par rapport à la surface de la terre sont en visibilité constante. En théorie, la communication peut donc être établie à tout moment et la récupération des données acquises est moins contraignante.

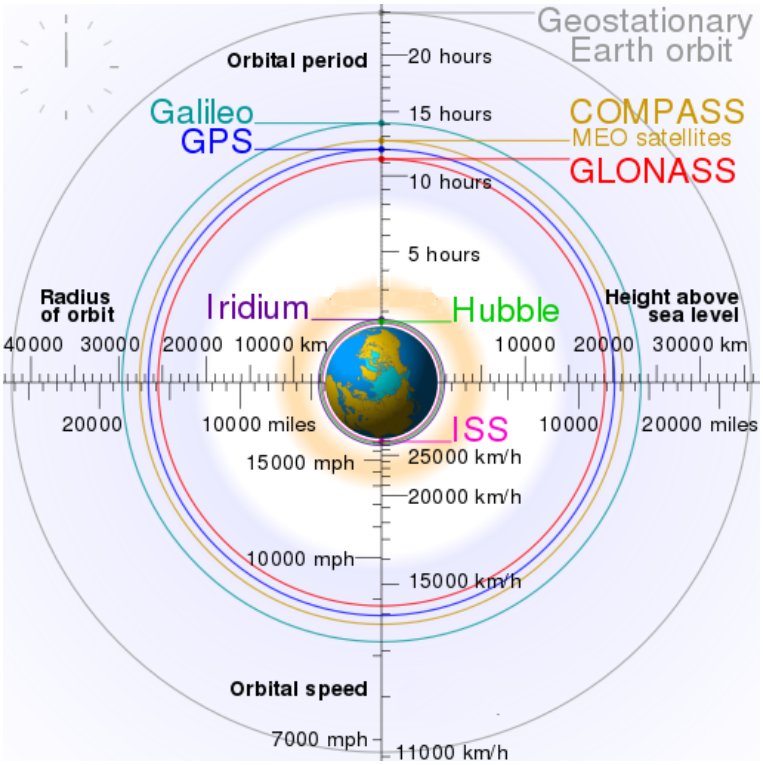

FIGURE 1.2 – Illustration de quelques orbites satellite (Image : Cmglee, Geo Swan/Wikimedia Commons, Licence CC).

1.2.3 Les applications satellite

Les applications des satellites artificiels ne cessent de se multiplier avec les avancées technologiques qui transforment notre société. Ils convient de distinguer deux types de satellites artificiels : les satellites scientifiques destinés à la recherche scientifique et les satellites d'applications, devenus indispensables à notre quotidien. Parmi les différents satellites d'applications nous nous intéresserons dans cette thèse aux satellites d'observation de la Terre et aux satellites de télécommunications.

1.2.3.1 Satellite d'observation de la Terre

Comme évoqué précédemment, l'observation de la Terre représente la première application des satellites artificiels. D'abord imaginés à des fins militaires, les satellites d'observation font aujourd'hui l'objet d'une exploitation commerciale pour la météorologie, l'inventaire et la gestion des ressources naturelles, l'étude et la modélisation du climat, la prévention et le suivi des catastrophes naturelles, etc. (voir Dubois et al. (2014) pour plus d'informations concernant les enjeux liés à l'observation de la terre). Ces satellites peuvent être très différents les uns des autres selon les instruments qu'ils embarquent (instrument optique, radar, multi-spectral, etc.) et leur résolution. La plupart de ces satellites circulent en orbite basse pour une meilleure résolution optique et ont une période de révolution d'environ 100 minutes. Certains satellites d'observation, notamment météorologique, circulent cependant sur l'orbite géostationnaire, fixe, afin d'assurer une surveillance permanente en temps réel d'une même zone.
Dans cette thèse, des données issues d'un satellite d'observation de la Terre opéré par la CNES ont pu être étudiées. La liaison avec ce satellite est établie en général 4 fois par jour pendant environ 10 minutes. C'est durant ce laps de temps que lui sont transmis sous forme de télécommandes les ordres concernant sa mission et/ou la récupération des données acquises à bord. Dans ce contexte, il est évident qu'une surveillance au sol en quasi temps réel n'est pas envisageable.

1.2.3.2 Satellite de télécommunication

Avec le lancement du premier satellite de télécommunication Intelsat I en 1965, la télécommunication par satellite représente la première application commerciale de l'ère spatiale. Initialement développés pour les télécommunications téléphoniques longue distance, les satellites de télécommunication ont aujourd'hui pour mission d'assurer les liaisons téléphoniques et satellitaires, la transmission de données et des programmes télévisés, l'accès à internet ou encore le positionnement GPS. La majorité de ces satellites circulent sur l'orbite géostationnaire et sont donc en visibilité permanente des stations de réception. Les données peuvent alors être reçues dès leur acquisition et une surveillance au sol en quasi-temps réel pourrait être imaginée. Dans ces travaux, certaines expérimentations ont été réalisées sur des données issues de satellites de télécommunication conçus par Airbus.

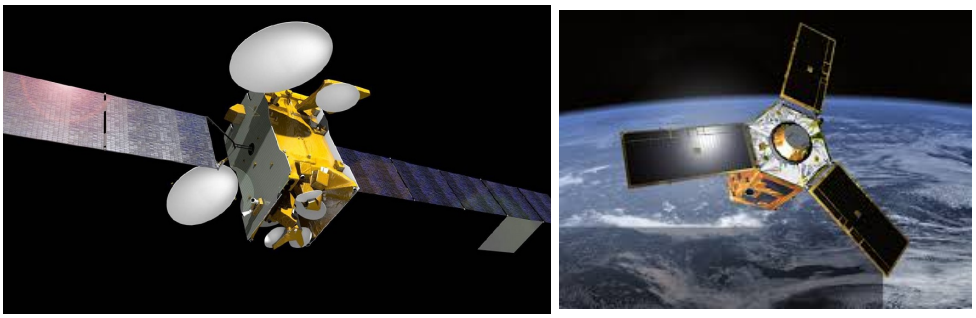

FIGURE 1.3 – SES-10, un satellite de télécommunication conçu par Airbus (droite), Pléiade 1, un satellite d'observation de la Terre conçu par le CNES.

1.3 La télémesure satellite

1.3.1 Définition

La télémesure satellite représente un ensemble de signaux acquis à bord par des capteurs embarqués. Reçue sous forme de séries temporelles, la télémesure satellite décrit l'évolution dans le temps de différents paramètres associés à une grandeur physique (une température, une tension, une pression, etc.), ou au fonctionnement d'un équipement (ex : statut ON/OFF ou position d'antenne). La télémesure satellite décrit le fonctionnement général du satellite, de ses systèmes, de ses sous-systèmes et de ses équipements. Il existe deux types de télémesures :

- La télémesure de servitude. Elle est constituée de signaux décrivant l'état global du satellite. Elle est systématiquement transmise, enregistrée en station puis reçue en temps différé par le centre de commande contrôle. Deux types de télémesure de servitude sont à distinguer : la télémesure temps réel (aussi appelée HKTMP pour HouseKeeping TeleMetry Pass) qui n'est acquise que lorsque l'émetteur du satellite est allumé, c'est à dire pendant un passage au-dessus d'une station, et la télémesure temps différé (aussi appelée HKTMR pour HouseKeeping TeleMetry Recorded) qui est enregistrée en permanence. Les données HKTMR sont stockées à bord dans une mémoire interne pour un téléchargement ultérieur. De ce fait, la télémesure HKTMR représente un volume de données beaucoup plus important que la télémesure HKTMP. C'est pourquoi les méthodes d'apprentissage automatique proposées pour la détection d'anomalies satellites sont appliquées à la télémesure HKTMR.
- La télémesure charge utile. Cette télémesure contient les données scientifiques, les données images (pour les satellites d'observation) et les données de fonctionnement de la charge utile, c'est à dire les équipements du satellite qui remplissent sa mission (les antennes et systèmes d'amplification du signal pour un satellite de télécommunications par exemple). Elle est enregistrée à bord hors et pendant les passages.

Un comportement atypique inattendu d'un ou plusieurs paramètres de télémesure reflète bien souvent un problème. C'est donc à partir de ces données qu'est réalisé le suivi et la surveillance de l'état de santé des satellites. Des extraits de télémesures satellites sont représentés dans la figure 1.4.

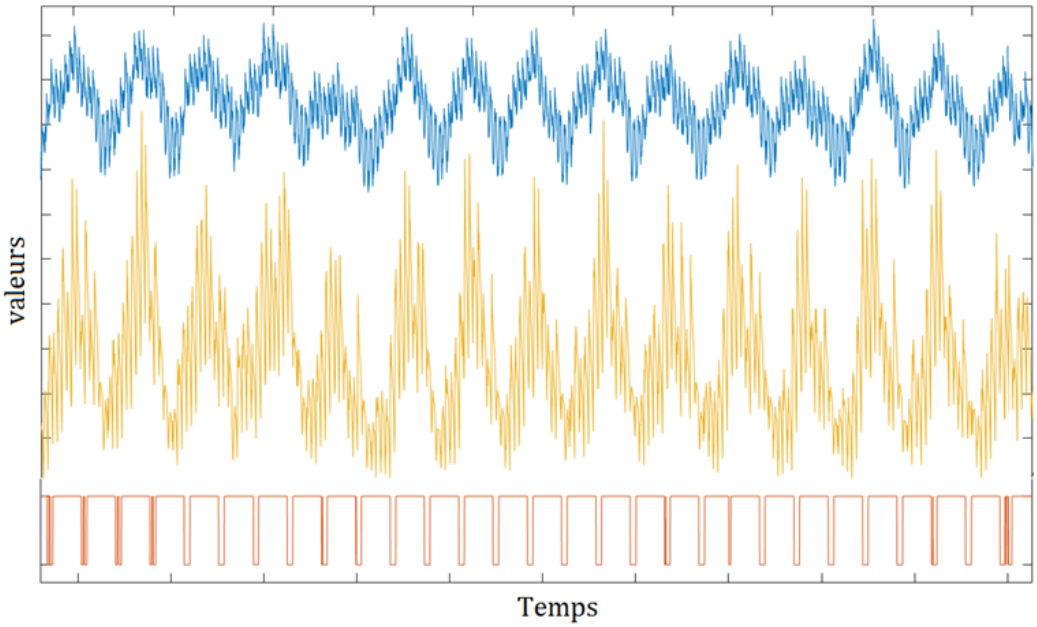

FIGURE 1.4 – Exemple de trois signaux télémesure satellite.

1.3.2 Les caractéristiques de la télémesure

Les données de télémesure satellite sont bien souvent numérisées de manière a optimiser leur stockage à bord (on parle alors de données brutes). Après réception, les données brutes sont converties au sol à l'aide d'une fonction de transfert correspondant à une grandeur physique (Ampère, mode d'équipements, etc) afin de pouvoir être traiter. Comme dans de nombreux domaines d'application, un traitement efficace des données de télémesure nécessite de cerner leurs caractéristiques afin d'appliquer les méthodes les mieux adaptées. Les principales caractéristiques des données de télémesures que nous avons étudiées dans cette thèse sont les suivantes :

1. De gros volumes de données.
   Selon la complexité et la taille du satellite, la télémesure peut compter plusieurs milliers de paramètres dont les mesures sont acquises à une fréquence élevée. En effet, au moins une mesure par minute est acquise pour chaque paramètre.
2. Des données quasi-périodiques.
   Les signaux de télémesure, très influencés par l'environnement du satellite, sont souvent caractérisés par une forte répétitivité avec une période correspondant à la période de révolution du satellite. Cette caractéristique peut être intéressante car elle rend possible l'application de méthodes adaptées aux données périodiques avec présence d'un bruit de mesure.
3. Des données hétérogènes.
   Comme évoqué précédemment, les paramètres de télémesure peuvent être associés à une grandeur physique (on parlera de paramètres "analogiques"), ou à un fonctionne-

ment d'équipement (on parlera de paramètres "symboliques"). Les paramètres symboliques prennent peu de valeurs différentes et peuvent être considérés comme des variables aléatoires discrètes tandis que les paramètres analogiques peuvent être considérés comme des variables aléatoires continues. Un exemple de signal discret (signal rouge) et deux exemples de signaux continus (signaux bleu et orange) sont représentés dans la figure 1.4. Les signaux continus sont également très hétérogènes car ils représentent l'évolution dans le temps de grandeurs physiques différentes, mesurées dans des unités et des intervalles de mesure différents.

4. Des données liées et/ou corrélées.
   Certains paramètres de télémesure sont très corrélés du fait notamment de la redondance des capteurs. Par ailleurs, il peut exister des relations de dépendance entre plusieurs paramètres de télémesure. C'est le cas par exemple de deux équipements qui fonctionnent de pair ou d'un capteur de température qui enregistre des valeurs plus ou moins élevées selon qu'un équipement proche est allumé ou éteint.

5. Des données numérisées.
   Certains signaux de télémesure associés à un paramètre continu ne sont pas convertis et sont alors reçus sous forme de signaux numérisés dont le nombre de valeurs possibles a été réduit à un ensemble fini (par une opération de quantification) afin de permettre le codage en binaire du signal et son stockage à bord. La figure 1.5 représente la télémesure d'un paramètre analogique reçu sous forme de signal numérisé. Ces signaux, à mi-chemin entre discrets et continus, peuvent faire l'objet d'un pré-traitement spécifique selon le traitement à réaliser.

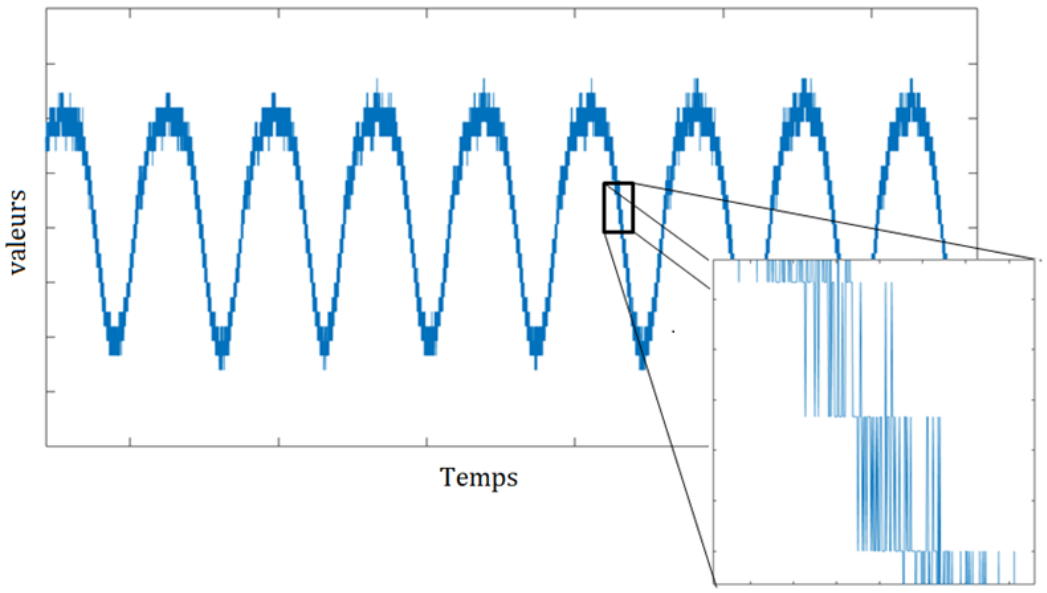

FIGURE 1.5 – Exemple d'un signal de télémesure numérisé associé à une grandeur physique.

6. Des données asynchrones.
   Á priori, les mesures d'un paramètre de télémesure sont acquises à une fréquence

donnée. Cependant, selon les besoins des opérations, la période d'échantillonage peut varier ponctuellement comme en témoigne la figure 1.6 qui représente la télémesure d'un paramètre dont la fréquence d'échantillonnage à été multipliée par deux pendant une courte période (figure 1.6, fond rouge). Par ailleurs, les mesures des différents paramètres ne sont pas toutes acquises aux mêmes instants et à la même fréquence. Un traitement conjoint de plusieurs paramètres de télémesure suppose donc un prétraitement adapté.

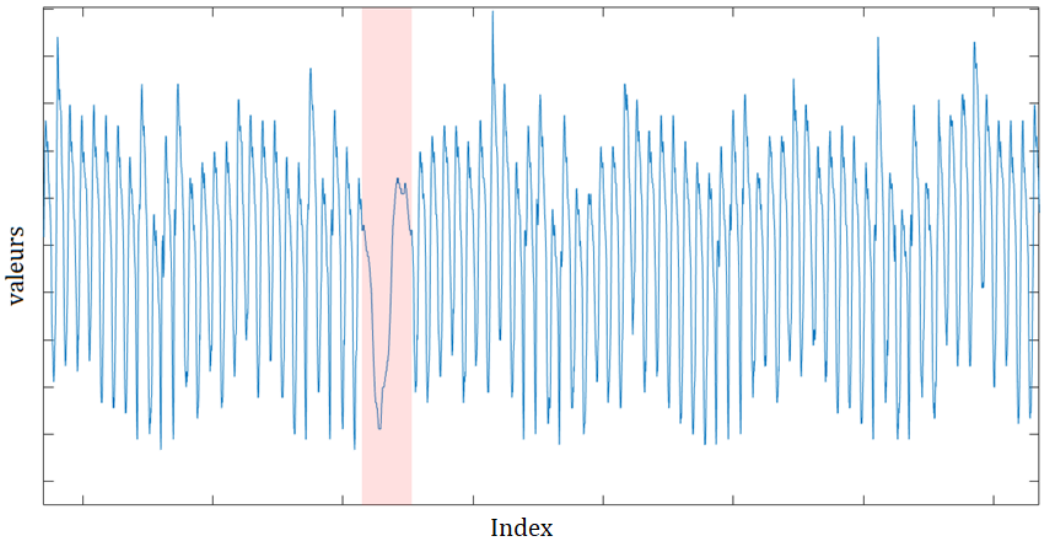

FIGURE 1.6 – Exemple de changement ponctuel de la fréquence d'échantillonnage d'un paramètre de télémesure. Durant la période marquée par un fond rouge, la fréquence d'échantillonage a été multipliée par deux afin de créer un zoom.

7. Des données aberrantes.
   Il se peut que quelques mesures très aberrantes apparaissent dans les données mais ne correspondent pas à une réelle anomalie. Il s'agit bien souvent d'erreurs lors de la conversion et/ou de la transmission des données.
8. Des données bruitées.
   De nombreux signaux de télémesure satellite sont bruités, comme en témoigne la figure 1.4. Des étapes de débruitage peuvent alors être intéressantes à effectuer selon la détection réalisée.

1.3.3 Les anomalies satellites

Ce qui caractérise les anomalies satellites est qu'elles sont rares et difficiles à répertorier étant donné la complexité des systèmes. Cependant, comme présenté dans la figure 1.7, nous pouvons distinguer deux catégories d'anomalies : les anomalies univariées et les anomalies multivariées. Les anomalies univariées affectent un unique paramètre pour lequel un comportement atypique est observé. Notons que lorsqu'une anomalie satellite se produit, plusieurs

anomalies univariées peuvent être observées au même moment sur différents paramètres de télémesure. Les anomalies univariées peuvent être classées en trois catégories Chandola et al. (2009) :

- Les anomalies locales : une anomalie locale correspond à une valeur aberrante ou extrême d'un paramètre de télémesure.
- Les anomalies contextuelles : une anomalie contextuelle correspond à une mesure ou une séquence de mesures déjà observée dans ce même signal mais anormale étant donné le contexte (décrit par le reste du signal) dans lequel elle est observée.
- Les anomalies collectives : une anomalie collective correspond à une séquence de mesures jamais observée pour le paramètre étudié.

Des exemples d'anomalies univariées locale, contextuelle et collective sont représentées dans la figure 1.7 (encadrés bleus). Les anomalies univariées peuvent être détectées par des méthodes de détections d'anomalies univariées qui traitent les paramètres de télémesure indépendamment les uns des autres.

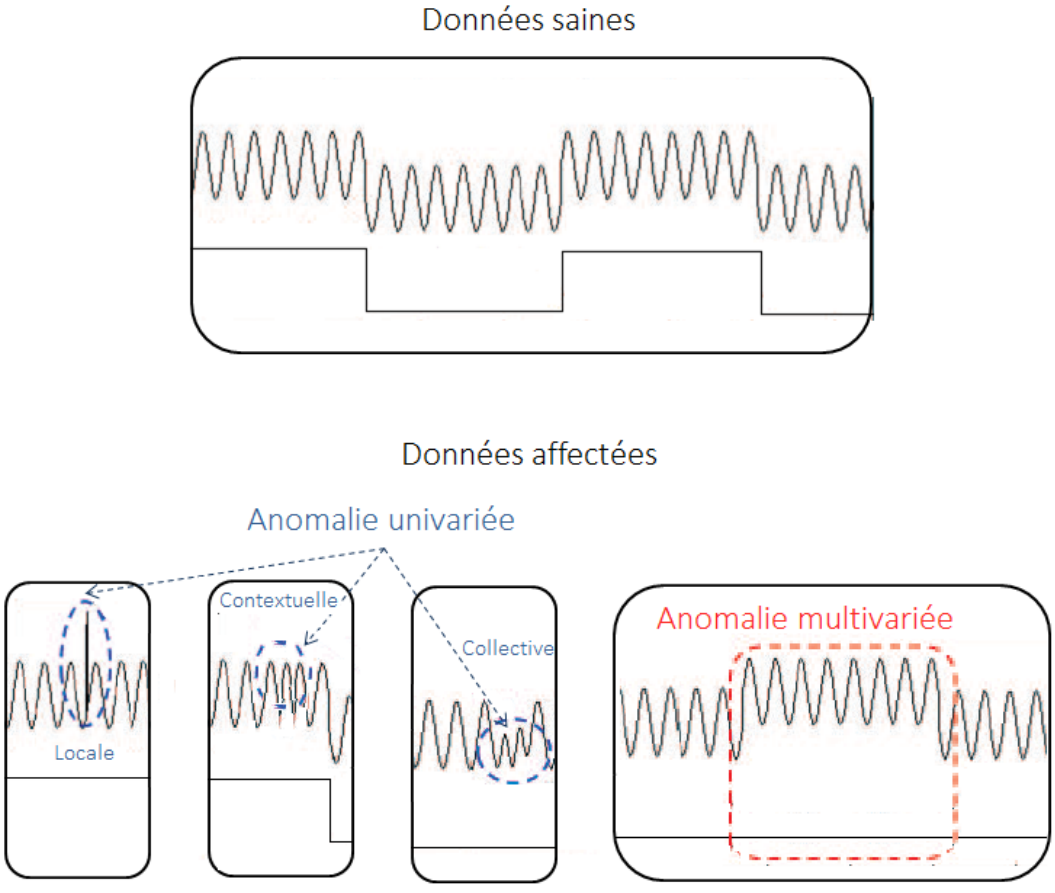

FIGURE 1.7 – Les anomalies de la télémesure satellite

Les anomalies multivariées correspondent à l'observation d'un changement dans les relations entre plusieurs paramètres. Dans l'exemple de la figure 1.7 les deux paramètres étudiés sont clairement liés. En effet, il apparaît qu'en fonctionnement normal (données saines) la moyenne du signal continu diminue lorsque l'équipement, dont le fonctionnement est décrit par le signal discret, est éteint. Ce phénomène n'est pas observé dans les données affectées (encadré rouge), ce qui représente une anomalie multivariée. La détection des anomalies multivariées suppose l'application de méthodes capables de traiter conjointement plusieurs paramètres de télémesure de manière à prendre en compte les relations qu'ils entretiennent.

1.3.4 Surveillance classique

1.3.4.1 Surveillance Bord

De nombreuses protections matérielles (hardware) existent à bord du satellite afin de le protéger des courts-circuits, des pics de courant ou encore des radiations. Une surveillance logicielle, appelée fonction FDIR (Failure Detection Isolation and Recovery), est également embarquée et directement intégrée au logiciel de vol . Cette dernière permet la détection temps réel des mesures aberrantes des paramètres et la gestion autonome des éventuelles anomalies par le satellite. Plus précisément, cette surveillance repose généralement sur une méthode de détection d'anomalies par seuils, aussi appelée méthode OOL pour Out Of Limit, qui consiste à définir pour chaque paramètre de télémesure continu un intervalle de valeurs attendues du paramètre. Les bornes de cet intervalle jouent alors un rôle de seuils de détection et une anomalie est détectée lorsque qu'une ou plusieurs valeurs successives hors de l'intervalle sont enregistrées. En cas de détection par la FDIR et selon la gravité de l'anomalie détectée sur le ou les paramètres impactés, le satellite est capable d'agir de manière autonome. Il peut alors reconfigurer ou isoler un de ses équipements ou de ses sous-systèmes, ou, dans les cas les plus critiques, arrêter sa mission en attendant l'intervention des équipes au sol.

1.3.4.2 Surveillance Sol

La surveillance au sol est également réalisée en grande partie à l'aide de la méthode OOL. Cependant, il s'agit d'une surveillance plus fine qu'à bord avec des intervalles de valeurs attendues réduits. La détection au sol, réalisée en temps différé, n'entraîne pas de réaction autonome du satellite. Elle permet aux équipes d'ingénieurs bord d'être informés de l'état de santé du satellite et de prévenir la détection des anomalies à bord en intervenant en amont. Des actions correctives peuvent alors être réalisées afin d'empêcher l'anomalie et assurer le bon déroulement de la mission du satellite. La méthode OOL représente un standard pour la détection d'anomalies dans la télémesure satellite mais elle connaît des limites. En effet, elle ne permet de détecter que des anomalies univariées locales, c'est à dire des valeurs extrêmes qui sortent des seuils, et ne détecte pas les comportements atypiques observés à l'intérieur de l'intervalle. La figure 1.8 représente un exemple d'une anomalie satellite réelle qui n'a pas été détectée par la méthode OOL. Peu de temps après, une anomalie plus sévère s'est

produite avec des valeurs sortant des seuils provoquant l'arrêt de la mission du satellite pendant plusieurs jours. Ce type d'évènement a motivé de nombreux acteurs du spatial à rechercher de nouvelles méthodes de détection d'anomalies pour la télémesure satellite. La majorité des études menées jusqu'à maintenant concerne la surveillance au sol car elle permet l'application de méthodes plus gourmandes en capacités de calcul et/ou en mémoire, des ressources limitées à bord.

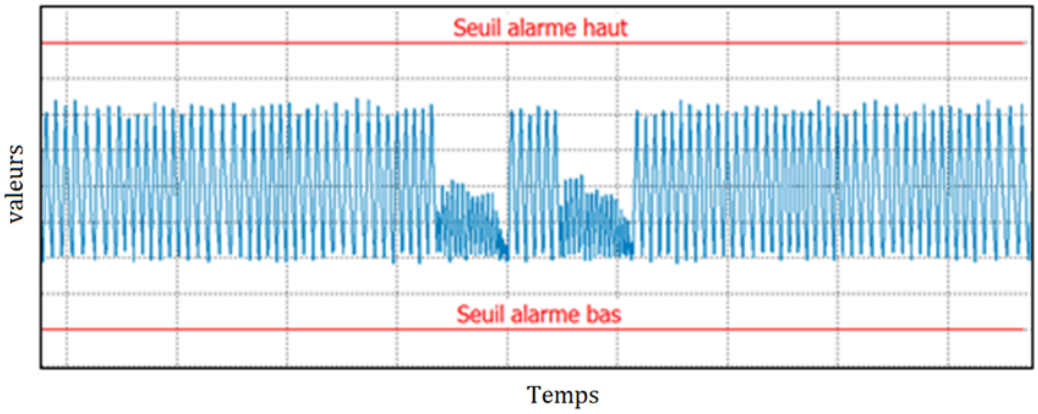

FIGURE 1.8 – Exemple d'anomalie non détectée par la surveillance par seuil OOL.

1.4 État de l'art

La détection d'anomalies est un vaste domaine de recherche tant ses applications sont nombreuses Chandola et al. (2009, 2012); Pimentel et al. (2014); Markou and Singh (2003b,a) Dans le domaine du spatial par exemple il est primordial d'assurer le suivi et le bon déroulement de la mission des satellites en exploitation à travers la surveillance de la télémesure. Ces dernières années ont vu naître un réel intérêt pour les méthodes de machine learning qui ont été appliquées avec succès pour la détection d'anomalies dans la télémesure satellite Martínez-Heras et al. (2012); Verzola et al. (2016); O'Meara et al. (2016); Fuertes et al. (2016); O'Meara et al. (2018); Hundman et al. (2018); Barreyre (2018); Takeishi and Yairi (2014). Cette section passe en revue quelques algorithmes récents proposés par différents acteurs du spatial.

1.4.1 Apprentissage semi-supervisé

Les méthodes de détection d'anomalies récemment proposées pour la surveillance de la télémesure satellite reposent, pour la grande majorité d'entre elles, sur des techniques d'apprentissage automatique semi-supervisé Zhu (2008); Engelen and Hoos (2019). L'aspect semi-supervisé est privilégié car il est particulièrement adapté au contexte de surveillance de la télémesure pour lequel : 1) la télémesure associée à un fonctionnement normal et attendu du satellite est connue et peut être fournie par les experts pour construire un modèle de référence,

et 2) les anomalies ne sont pas toutes répertoriées et ne peuvent donc pas être apprises afin de détecter les nouvelles occurrences, ce qui exclut un apprentissage supervisé. La détection d'anomalies satellite à l'aide de méthodes d'apprentissage semi-supervisé peut être résumée en deux étapes : l'apprentissage et la détection. La première étape consiste à construire un modèle de référence à partir de données saines, sans anomalie, et la seconde étape compare au modèle appris les nouvelles données reçues afin de détecter tout comportement atypique.

1.4.2 Prétraitement

Au-delà du choix de la méthode de détection d'anomalies, c'est le prétraitement appliqué aux données de télémesure qui caractérise les différents algorithmes développés et qui permet d'optimiser l'efficacité de la détection. Des étapes classiques de normalisation des données par exemple sont alors largement réalisées et ce, quelque soit la détection choisie. Par ailleurs, les séries temporelles sont souvent découpées en fenêtres de taille fixe W (ce découpage est souvent appelé "segmentation" ou "fenêtrage"). La taille des fenêtres est un paramètre fixé par l'utilisateur et dépend de la stratégie de surveillance adoptée (surveillance orbitale, quotidienne, hebdomadaire, etc.).

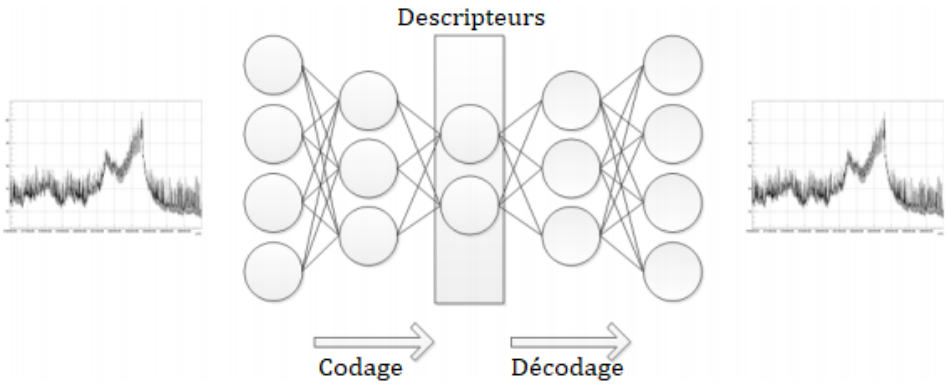

FIGURE 1.9 – Illustration du principe d'extraction de descripteurs à l'aide d'un auto-encoder.

Compte tenu du volume important de données que représente la télémesure, il peut être intéressant de chercher à résumer l'information pour faciliter le traitement. Il s'agit alors de réduire la dimension des fenêtres obtenues suite à la segmentation. Ces problématiques de réduction de dimension sont récurrentes en statistique et comptent de nombreuses solutions. Parmi celles appliquées à des données de télémesure satellite, la plus simple à mettre en oeuvre consiste à calculer sur chaque fenêtre des statistiques classiques tels que la moyenne, l'écart-type ou encore le minimum Fuertes et al. (2016); Verzola et al. (2016); Martínez-Heras et al. (2012); O'Meara et al. (2016). Une autre idée est d'utiliser une projection sur une base adaptée et d'extraire les premiers coefficients. De nombreuses bases et méthodes de détection d'anomalies ont été étudiées et comparées dans et une méthode de sélection automatique de coefficients pertinents pour la détection d'anomalies a été développée Barreyre et al. (2019).

Par ailleurs l'intérêt des réseaux de neurones artificiels et plus précisément des auto-encodeurs à également été étudié. Comme illustré dans la figure 1.9, tiré de O'Meara et al. (2018), l'idée est de construire un réseaux de neurones artificiel dont la couche centrale renvoie une représentation du signal dans un espace de dimension réduit. Outre quelques étapes de pré-traitement des données telles que celles présentées précédemment, l'application de méthodes d'apprentissage semi-supervisé suppose de construire un ensemble de données d'apprentissage à partir duquel sera appris le modèle de référence. Dans notre contexte, il s'agit de fournir une quantité suffisante de télémesure associée à un fonctionnement normal du satellite. Cela suppose une lourde étape de nettoyage de la télémesure afin de supprimer les données liées à des évènements ponctuels ou à des opérations particulières (une manoeuvre de correction d'orbite ou d'évitement par exemple) qui ne doivent pas être appris. Ce travail est souvent réalisé manuellement par les experts. Pourtant, certains algorithmes proposent de construire le jeux de données d'apprentissage de manière automatique. C'est le cas par exemple de l'algorithme Sybil Verzola et al. (2016), utilisé pour la surveillance du module Columbus de la station spatiale internationale (aussi appelée ISS pour International Space Station), ou de l'algorithme ATHMoS O'Meara et al. (2016, 2018) développé par le DLR (l'agence spatiale allemande), qui utilisent une méthode de classification non supervisée : l'algorithme DBSCAN Ester et al. (1996). L'algorithme DBSCAN repose sur la recherche du ϵ-voisinage de chaque point (fenêtre), c'est-à-dire qu'il détermine pour chaque point l'ensemble de ses voisins situés à une distance inférieure à ϵ, où ϵ est un paramètre fixé par l'utilisateur. Les points ayant moins de M voisins sont considérés comme comportements atypiques ou bruit et ne sont affectés à aucun groupe. Les points ayant un ϵ–voisinage assez grand et donc atteignable de proche en proche sont regroupés dans un même groupe. Le fonctionnement de l'algorithme DBSCAN est représenté de manière schématique dans la figure 1.10. Dans le cadre de la détection d'anomalies Sybil ou ATHMoS, seuls les données classées dans des groupes qui représentent plus de 5% des données d'apprentissage sont utilisées pour l'apprentissage.

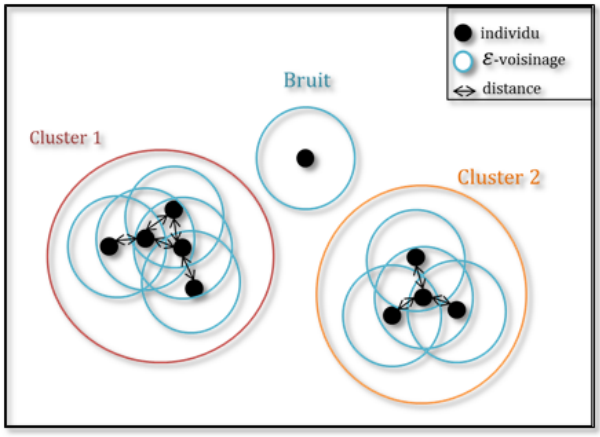

FIGURE 1.10 – Représentation schématique du fonctionnement de l'algorithme de classification DBSCAN.

1.4.3 Détection d'anomalies univariées

1.4.3.1 Facteur d'anormalité locale

Le facteur d'anormalité locale ou facteur LOF (Local Outlier Factor) Breunig et al. (2000) est un score d'anomalie qui repose sur l'étude des densités locales des observations, estimées à l'aide de calculs de distance entre plus proches voisins. L'hypothèse clé de cette approche est que les anomalies sont localisées dans les régions de faible densité. Dans un contexte de détection d'anomalies semi-supervisée il s'agit de comparer la densité locale d'une nouvelle observation à celles de ses k plus proches voisins issus d'un ensemble d'apprentissage composé d'observations saines (sans anomalie). L'étape d'apprentissage consiste alors en la construction et le stockage de l'ensemble de données associées à un fonctionnement normal du satellite. Le calcul du facteur LOF d'une nouvelle observation $\boldsymbol{x} \in \mathbb{R}^N$ associée à la télémesure fraîchement acquise suppose de définir les notions suivantes

- La k-distance de $\boldsymbol{x}$, notée $\mathrm{Dist}_k(\boldsymbol{x})$ est la distance entre $\boldsymbol{x}$ et son $k^{\text{ème}}$ plus proche voisin. Nous considérerons ici la distance euclidienne, notée $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_i (\boldsymbol{x}_i - \boldsymbol{y}_i)^2}$.

- Le k-voisinage de $\boldsymbol{x}$, noté $\mathrm{V}_k(\boldsymbol{x})$, est l'ensemble de ses k plus proches voisins. Il peut être défini comme suit

$$\mathrm{V}_k(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathcal{D} | d(\boldsymbol{x}, \boldsymbol{y}) \leq \mathrm{Dist}_k(\boldsymbol{x})\} \tag{1.1}$$

  où $\mathcal{D}$ désigne l'ensemble d'apprentissage.

- La distance atteignable entre deux observations $\boldsymbol{x}$ et $\boldsymbol{y}$, notée $\mathrm{RDist}_k(\boldsymbol{x}, \boldsymbol{y})$ est définie comme suit

$$\mathrm{RDist}_k(\boldsymbol{x}, \boldsymbol{y}) = \max(\mathrm{Dist}_k(\boldsymbol{y}), d(\boldsymbol{x}, \boldsymbol{y})) \tag{1.2}$$

La densité locale de $\boldsymbol{x}$, notée $\mathrm{lr}_k(\boldsymbol{x})$, est définie par l'inverse de la moyenne des distances atteignables entre $\boldsymbol{x}$ et ses k plus proches voisins, i.e,

$$\mathrm{lr}_k(x) = \frac{\#\mathrm{V}_k(\boldsymbol{x})}{\sum_{\boldsymbol{y} \in V_k(\boldsymbol{x})} \mathrm{RDist}_k(\boldsymbol{x}, \boldsymbol{y})} \tag{1.3}$$

où $\#\mathrm{V}_k(\boldsymbol{x})$ est le cardinal du k-voisinage de $\boldsymbol{x}$. Le facteur LOF de $\boldsymbol{x}$ est obtenu en comparant sa densité locale à celle de ses plus proches voisins tel que

$$\mathrm{LOF}_k(\boldsymbol{x}) = \frac{1}{\#V_k(\boldsymbol{x})} \sum_{\boldsymbol{y} \in V_k(\boldsymbol{x})} \frac{\mathrm{lr}_k(\boldsymbol{y})}{\mathrm{lr}_k(\boldsymbol{x})}. \tag{1.4}$$

Plus $\boldsymbol{x}$ est éloigné de ses voisins au sens de la densité locale, plus le facteur LOF est élevé. La détection d'anomalies à l'aide du facteur LOF consiste à considérer comme atypique toute observation dont le LOF est supérieur à un seuil, noté S_{PFA} fixé par l'utilisateur, i.e

$$\text{Une anomalie est détectée si } \mathrm{LOF}_k(\boldsymbol{x}) > \mathrm{S}_{\mathrm{PFA}}. \tag{1.5}$$

Le facteur LOF n'est pas borné ce qui le rend difficile à interprété et peut rendre difficile la détermination du seuil de détection S_{PFA}. Par ailleurs il peut être sensible au choix du nombre k de plus proches voisins. Afin de contourner ces limites, plusieurs variantes ont été proposées tel que le score LoOP (pour Local Outlier Probability), une version probabiliste et normalisé du score LOF Kriegel et al. (2009). En effet, le score LoOP est compris entre 0 et 1 et peut être directement interprété comme la probabilité qu'un point soit atypique. La figure 1.11 tiré de Kriegel et al. (2009) compare les indicateurs LOF et LoOP d'un nuage de points 2D. Les résultats présentés dans la figure 1.11 illustrent l'efficacité de ces scores (représentés par des cercles rouges de taille proportionnelle au score) pour discriminer les points isolés et témoignent de leur intérêt pour la détection d'anomalies.

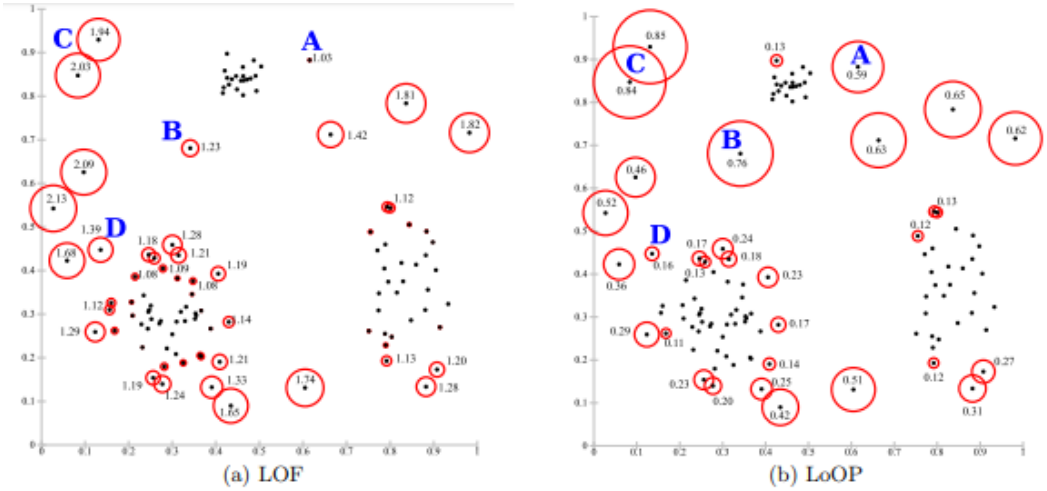

FIGURE 1.11 – Comparaison des indicateurs LOF et LoOP d'un nuage de point 2D. $k = 20$.

L'algorithme NOVELTY Martínez-Heras et al. (2012) developpé par l'agence spatiale européenne (aussi appelée ESA pour European Space Agency) fonde sa détection d'anomalies dans la télémesure continue sur le score LoOP. Après avoir normalisé les données, NOVELTY réalise un pré-traitement en deux étapes. La première étape est une étape de segmentation lors de laquelle les signaux de télémesure satellite sont découpés en fenêtres temporelles de taille fixe correspondant à une durée de 90 minutes. La seconde étape est une étape de réduction de dimension lors de laquelle des descripteurs classiques (moyenne, écart-type, minimum et maximum) sont calculés pour chaque fenêtre temporelle. En ce qui concerne la détection, l'idée est de calculer le score LoOP pour différentes valeurs de k plus proches voisins, à savoir $k \in \{10, 20, 30\}$. C'est alors le score minimal parmi les scores obtenus à la suite des différentes exécutions qui est retenu et comparé à un seuil de détection fixé à $S_{\text{PFA}} = 0.9$, i.e

$$\text{Une anomalies est detectée si } \min(\text{LoOP}_{10}(\boldsymbol{x}), \text{LoOP}_{20}(\boldsymbol{x}), \text{LoOP}_{30}(\boldsymbol{x})) > 0.9. \tag{1.6}$$

Largement inspiré de NOVELTY, l'algorithme SYBIL Verzola et al. (2016) développé par le GSOC (le centre des opérations de l'agence spatiale allemande), repose également sur l'étude du score LoOP et adopte un pré-traitement et une détection similaires à celle de NOVELTY.

Cependant il réalise la construction de l'ensemble d'apprentissage automatiquement à l'aide de l'algorithme de classification non supervisé DBSCAN (voir la section précédente) tandis que la construction de l'ensemble d'apprentissage NOVELTY est réalisée manuellement. Ces deux algorithmes sont résumés dans la tableau 1.1 où $\#g_k$ signifie le cardinal du $k^{ème}$ groupe, $\mathcal{D}$ désigne l'ensemble d'apprentissage et mean, std, min et max sont respectivement la moyenne, l'écart-type, le minimum et le maximum calculés sur chaque fenêtre de données. La détection d'anomalies NOVELTY ou SYBIL est facile à mettre en oeuvre et est humainement interprétable. Cependant, elle suppose de stocker les données d'apprentissage qui représentent souvent un volume important de données. Par ailleurs, la détermination des k plus proches voisins d'une nouvelle observation suppose de nombreux calculs, ce qui peut être très coûteux si aucun moyen de parallélisation n'est disponible.

TABLE 1.1 – Les algorithmes NOVELTY (ESA) et SYBIL (GSOC/DLR)

Paramètres	Continus	Continus
Pré-traitement	1- Normalisation 2- Ségmentation : $W = 90'$ 2- Descripteurs :{mean, std, min, max}	Normalisation 1- Ségmentation : $W = 90'$ 2- Descripteurs :{mean, std, min, max}
Apprentissage	Manuel	Automatique : DBSCAN $\mathcal{D} = \{g_k \vert \#g_k / \sum_k \#g_k > 0.05\}$
Méthode	LoOP $k = \{10, 20, 30\}$	LoOP $k = \{10, 20, 30\}$
Score	$\min(\text{LoOP}_{10}(\boldsymbol{x}), \text{LoOP}_{20}(\boldsymbol{x}), \text{LoOP}_{30}(\boldsymbol{x}))$	$\min(\text{LoOP}_{10}(\boldsymbol{x}), \text{LoOP}_{20}(\boldsymbol{x}), \text{LoOP}_{30}(\boldsymbol{x}))$
Détection	$\text{LoOP}(\boldsymbol{x}) > 0.9$	$\text{LoOP}(\boldsymbol{x}) > 0.9$

1.4.3.2 Dimension intrinsèque locale

La dimension intrinsèque d'un ensemble de données correspond à la dimension de l'espace dans lequel vivent les données. Elle est bien souvent inférieure à la dimension d'acquisition des données et est étroitement liée au nombre de degrés de liberté du processus qui les a généré. Le problème d'estimation de la dimension intrinsèque est souvent lié à des applications de réduction de dimension. Cependant, dans des travaux récents Houle (2013), Houle introduit la notion de dimension intrinsèque locale, très intéressante pour des applications de classification ou de détection d'anomalies par exemple. Il démontre que si l'on considère les distances entre les points d'un sous-ensemble donné comme des réalisations d'une variable aléatoire $\boldsymbol{X}$ sur $[0, \infty)$, alors la dimension intrinsèque d'un sous-ensemble de rayon x, appelée dimension intrinsèque locale (DIL) et notée $\text{ID}_{\boldsymbol{X}}(x)$, est donnée par

$$\text{ID}_{\boldsymbol{X}}(x) = \frac{\boldsymbol{x} f_{\boldsymbol{X}}(x)}{F_{\boldsymbol{X}}(x)} \tag{1.7}$$

où $f_{\boldsymbol{X}}(\boldsymbol{x})$ est la densité de probabilité et $F_{\boldsymbol{X}}(\boldsymbol{x})$ la fonction de répartition associées à la variable aléatoire $\boldsymbol{X}$. Á partir de cette définition, Amsaleg et al. (2015) proposent plusieurs estimateurs de la DIL tel que l'estimateur du maximum de vraissemblance donné par

$$\widehat{\text{ID}}_{\boldsymbol{X}} = -\left[\frac{1}{k}\sum_{i=1}^{k-1} \ln\left(\frac{\boldsymbol{x}_i}{\boldsymbol{x}_k}\right)\right]^{-1} \tag{1.8}$$

où les $\boldsymbol{x}_i$ sont des observations ordonnées de $\boldsymbol{X}$ telles que $\boldsymbol{x}_1 \leq \boldsymbol{x}_2 \leq ... \leq \boldsymbol{x}_k$.
Ces résultats ont été utilisés pour la détection d'anomalies dans la télémesure satellite. En effet, la détection de l'algorithme ATHMoS O'Meara et al. (2016), développé par le DLR (l'agence spatiale allemande), repose sur l'étude de cet indicateur calculé à partir des mesures de distance entre une observation de référence $\boldsymbol{q}$ associée à la nouvelle télémesure acquises et ses k plus proches voisins issus de l'ensemble d'apprentissage (sans anomalie). Un score d'anomalie associé au point $\boldsymbol{q}$, noté PIDOS($\boldsymbol{q}$) a alors été défini comme suit

$$\text{PIDOS}_k(\boldsymbol{q}) = \widehat{\text{ID}}_{\boldsymbol{X}}(\boldsymbol{q}) - \frac{1}{k} \sum_{\boldsymbol{y} \in \text{V}_k(q)} \widehat{\text{ID}}_{\boldsymbol{X}}(\boldsymbol{y}) \tag{1.9}$$

où $\text{V}_k(\boldsymbol{q})$ est le k-voisinage de $\boldsymbol{q}$ défini à la section 1.4.3.1. Á l'image du facteur LOF, le score PIDOS peut être normalisé pour appartenir à l'intervalle [0,1]. La normalisation proposée du score PIDOS mène au score OPVID défini comme suit

$$\text{OPVID}_k(\boldsymbol{q}) = \max \left\{ 0, erf \left(\frac{\text{PIDOS}(\boldsymbol{q})}{\lambda.\sqrt{2}\text{PIDOS}_{\text{STD}}(\boldsymbol{q})} \right) \right\} \tag{1.10}$$

où $\text{PIDOS}_{\text{STD}}(\boldsymbol{q})$ correspond à l'écart-type des scores PIDOS des k plus proches voisins de $\boldsymbol{q}$, erf est la fonction d'erreur de Gauss, et λ est un paramètre d'échelle fixé par l'utilisateur. La détection ATHMoS est réalisée à partir de ce score et une anomalie est détectée si le score excède un seuil de détection fixé par l'utilisateur, i.e

$$\text{Une anomalie est détectée si OPVID}(\boldsymbol{q}) > S_{\text{PFA}}. \tag{1.11}$$

Le fonctionnement de l'algorithme ATHMoS est décrit dans le tableau 1.2 où $\mathcal{D}$, $\mathcal{D}_{\text{Smooth}}$ et $\mathcal{D}_{\text{Noise}}$ désignent respectivement l'ensemble d'apprentissage d'origine, l'ensemble d'apprentissage dont les signaux ont subi un lissage et le bruit correspondant à la différence des deux premiers ensembles.

TABLE 1.2 – L'algorithme ATHMoS (DLR)

Paramètres	Continus
Pré-traitement	1- Segmentation sur 3 ensembles : $\mathcal{D}, \mathcal{D}_{\text{Smooth}}, \mathcal{D}_{\text{Noise}}$, $W = 90'$ 2- Descripteurs : classiques (mean, std, min, max, etc.) + auto-enccodeur 3- Concaténation
Apprentissage	Automatique : DBSCAN $\mathcal{D} = \{g_k \vert \frac{\#g_k}{\sum_k \#g_k} > 0.05\}$
Méthode	OPVID $k = 10$
Score	$\text{OPVID}_{10}(\boldsymbol{x}) \in [0, 1]$
Détection	$\text{OPVID}(\boldsymbol{x}) > S_{PFA}$

De même que pour la détection d'anomalies à l'aide du score LOF ou LoOP, la détection à partir du score OPVID est simple et facile à mettre en oeuvre mais demande de réaliser beaucoup de calculs et suppose de disposer d'une capacité de stockage suffisante.

1.4.3.3 One-Class Support Vector Machine

Appliqué pour la détection d'anomalies dans la télémesure satellite, l'algorithme One-Class Support Vector Machine (OC-SVM) Schölkopf et al. (2001) consiste à plonger les données de l'ensemble d'apprentissage décrivant un fonctionnement normal du satellite (sans anomalies) dans un espace de plus grande dimension $\mathcal{H}$ à l'aide d'une transformation φ, et à trouver dans cet espace un hyperplan qui sépare les données de l'origine avec une marge maximale notée γ. Cet hyperplan peut être vu comme une frontière délimitant l'espace occupé par les données d'apprentissage du reste de l'espace. La détermination de cette frontière revient à résoudre le problème suivant

$$\min_{\mathbf{w},\rho,\mathcal{E}_i} \frac{1}{2} \parallel \mathbf{w} \parallel_2 + \frac{1}{\nu N} \sum_{i=1}^{N} \mathcal{E}_i - \rho \quad \text{s.t.} \quad \langle \mathbf{w}, \varphi(\boldsymbol{x}_i) \rangle_{\mathcal{H}} \geq \rho - \mathcal{E}_i \forall i = 1, ..., n, \ \mathcal{E}_i \geq 0 \ i = 1, ..., n \tag{1.12}$$

où $\mathbf{w}$ est le vecteur normal à la frontière, ρ est le biais, $\mathcal{E}_i, i = 1, ..., N$ sont des variables ressort (slack variables en anglais), qui sont égales à 0 lorsque $\boldsymbol{y}$ satisfait la contrainte (i.e, $\boldsymbol{y}$ est à l'intérieur de la frontière) et sont strictement positives lorsque la contrainte n'est pas satisfaite (i.e, $\boldsymbol{y}$ est à l'extérieur de la frontière), ν est un paramètre de relaxation fixé par l'utilisateur qui correspond à la proportion des données autorisées à être en dehors de la frontière et $< ., . >$ est un produit scalaire. Ce problème peut être résolu par la méthode des multiplicateurs de Lagrange qui mène au problème dual suivant

$$\min_{\alpha} \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \langle \varphi(\boldsymbol{x}_i), \varphi(\boldsymbol{x}_j) \rangle \text{ s.t. } 0 \leq \alpha_i \leq \frac{1}{\nu n}, \sum_{i=1}^{n} \alpha_i = 1 \tag{1.13}$$

où les paramètres α_i sont les multiplicateurs de Lagrange. Le calcul des produits scalaires dans un espace de grande dimension peut être coûteux et la transformation φ est parfois difficile à déterminer. L'astuce pour résoudre la problème (1.13) est de définir un noyau $\mathrm{K}(\boldsymbol{x}, \boldsymbol{y}) = \langle \varphi(\boldsymbol{x}), \varphi(\boldsymbol{y}) \rangle$. Nous considérerons ici le noyau gaussien dont la définition est donnée ci-dessous

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\sigma \parallel \boldsymbol{x} - \boldsymbol{y} \parallel^2\right) \tag{1.14}$$

avec σ un paramètre fixé par l'utilisateur qui contrôle la régularité de la frontière. Pour la détection, un nouvel individu $\boldsymbol{x}$ associé à la télémesure fraîchement acquise est jugé atypique s'il est situé à l'extérieur de la frontière, i.e.,

$$\text{Une anomalie est détectée si } \mathrm{sign}[\mathrm{K}(\mathbf{w}, \boldsymbol{x}) - \rho] = -1 \tag{1.15}$$

où sign(.) est une fonction définie par

$$\text{sign}(\boldsymbol{y}) = \begin{cases} -1 & \text{si } \boldsymbol{y} < 0 \\ 0 & \text{si } \boldsymbol{y} = 0 \\ 1 & \text{si } \boldsymbol{y} > 0 \end{cases} \tag{1.16}$$

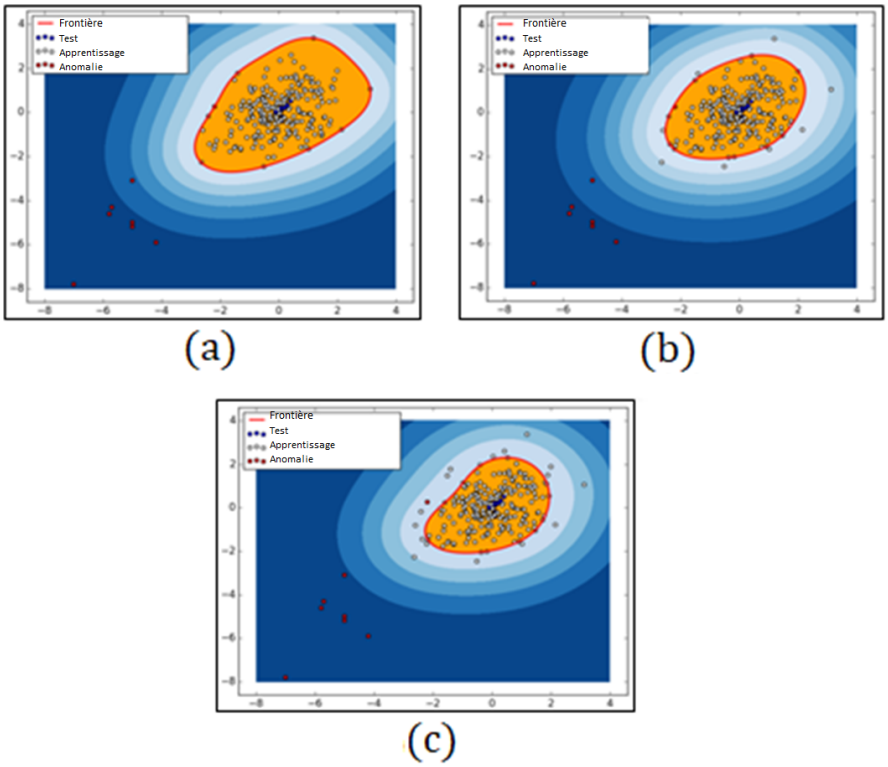

FIGURE 1.12 – Comparaison de la frontière séparatrice OC-SVM en faisant varier la valeur du paramètre ν : $\nu = 0.01$ (a), $\nu = 0.05$ (b) et $\nu = 0.1$ (c) avec $\sigma = 0.18$.

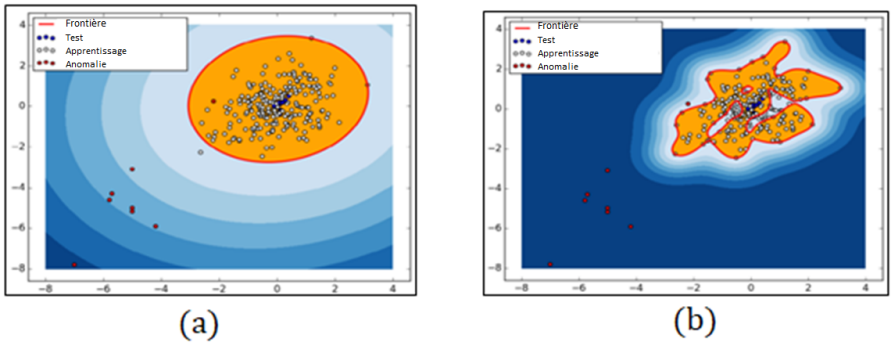

FIGURE 1.13 – Comparaison de la frontière séparatrice OC-SVM en faisant varier la valeur du paramètre σ : $\sigma = 0.01$ (a) et $\sigma = 1$ (b) avec $\nu = 0.01$.

L'algorithme OC-SVM impose de régler deux hyperparamètres : ν et σ qui contrôlent la proportion de points d'apprentissage autorisés à être à l'extérieur de la frontière et la régularité de la frontière. Le premier permet de se prémunir contre la présence d'éventuels comportements atypiques dans les données d'apprentissage. Les figures 1.12 et 1.13 illustrent l'influence de ces paramètres sur la construction de la frontière séparatrice. On observe que plus la valeur de ν est élevée, plus le nombre de points d'apprentissage situés à l'extérieur de la frontière est élevé. Par ailleurs, plus la valeur de σ est élevée, plus la forme de la frontière est irrégulière et le risque de sur-apprentissage est important.
L'algorithme de détection d'anomalies univariées NOSTRADAMUS Fuertes et al. (2016) proposé par le CNES repose sur la méthode OC-SVM. Les signaux de télémesure sont découpés en fenêtres temporelles de taille fixe correspondant à une durée de 24h. Des descripteurs (moyenne, minimum, écart-type etc.) sont ensuite calculés sur chaque fenêtre. La méthode NOSTRADAMUS est capable de traiter des paramètres continus et des paramètres discrets en utilisant des descripteurs adaptés. Le paramètre ν à été fixé à $\nu = 0.01$ et le paramètre σ est donné par l'heuristique de Jaakola Luus and Jaakola (1973). Par ailleurs, afin de nuancer la détection OC-SVM, un score d'anomalie normalisé a été défini tel que

$$\text{NOSTRADAMUS}(\boldsymbol{x}) = \begin{cases} \frac{(\text{K}(\mathbf{w},\boldsymbol{x})-\rho)+\gamma}{\gamma} & \text{si } (\text{K}(\mathbf{w},\boldsymbol{x})-\rho) < 0 \\ 0 & \text{sinon} \end{cases}. \tag{1.17}$$

L'algorithme NOSTRADAMUS est résumé dans le tableau 1.3 ci-dessous.

TABLE 1.3 – L'algorithme NOSTRADAMUS (CNES)

Paramètres	Continus & Discrets
Pré-traitement	1- Segmentation , $W = 24\text{h}H$
	2- Calcul de Descripteurs
	3- Réduction de dimension ACP (option)
Apprentissage	Manuel
Méthode	OC-SVM
	$\nu = 0.01$
	$\sigma =$ Jaakola
Score	NOSTRADAMUS$(\boldsymbol{x}) \in [0, 1]$
Détection	NOSTRADAMUS$(\boldsymbol{x}) > 0.4$

1.4.3.4 Distance à noyaux

Une autre approche proposée par Airbus Barreyre et al. (2019); Barreyre (2018) pour la détection d'anomalies dans la télémesure satellite fonde sa détection sur l'étude d'un indice de dissimilarité appelé TNND pour Temporal Nearest Neighbours Dissimilarity. L'indice TNND

d'un point $\boldsymbol{x} \in \mathbb{R}^N$ est défini par

$$\text{TNND}_{\boldsymbol{x}} = \sum_{j \leq n-1} d_{\text{K}}(\boldsymbol{x}, \boldsymbol{y}_j) \tag{1.18}$$

où $d_{\text{K}}(\boldsymbol{x}, \boldsymbol{y}_j)$ est une distance à noyau telle que

$$d_{\text{K}}(\boldsymbol{x}, \boldsymbol{y}_j) = \text{K}(n, j) \times d(\boldsymbol{x}, \boldsymbol{y}_j), \tag{1.19}$$

$j = 1, ..., n-1$, $\text{K}(n, j)$ est un noyau tel que $\forall i, j, h \in \mathbb{R}, \text{K}(i, j) = \text{K}(i + h, j + h)$, et $d(\boldsymbol{x}, \boldsymbol{y}_j)$ est la distance euclidienne telle que $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_i (\boldsymbol{x}_i - \boldsymbol{y}_i)^2}$.
L'indice TNND permet de mesurer la dissimilarité entre une nouvelle entrée $\boldsymbol{x}$ et des données de référence $\boldsymbol{y}_j, j = 1, ..., n-1$ en accordant plus de poids aux derniers éléments. La détection d'anomalies à partir de cet indice consiste à définir un seuil S_{PFA} et la règle de décision suivante :

$$\text{Anomalie détectée si } \text{TNND}_{\boldsymbol{x}} > S_{\text{PFA}}. \tag{1.20}$$

Dans un contexte général de détection d'anomalies semi-supervisée, nous pouvons imaginer calculer cet indice pour une nouvelle observation $\boldsymbol{x}$ à partir de ses $n-1$ plus proches voisins $\boldsymbol{y_j}, j = 1, ..., n-1$ issus de l'ensemble d'apprentissage. Afin de donner plus de poids aux premiers plus proches voisins, on notera $\boldsymbol{y}_j, j = 1, ..., n-1$ tel que $d(\boldsymbol{x}, \boldsymbol{y}_1) > ... > d(\boldsymbol{x}, \boldsymbol{y}_{n-1})$.
Pour la détection d'anomalies dans la télémesure satellite Barreyre et al. (2019); Barreyre (2018), les auteurs proposent de tirer partie de la périodicité de la télémesure et de comparer plusieurs périodes consécutives d'un même signal de télémesure afin de détecter les changements brutaux. L'indice TNND associé à la n^{ème} période du signal mesure alors la dissimilarité entre ce dernier et les $n-1^{ème}$ dernières périodes du signal en accordant plus de poids aux périodes les plus récentes.
Cette méthode de détection d'anomalies a été appliquée a des données réelles de satellites géostationnaires en utilisant un noyau Laplacien. Les signaux de télémesure de ces satellites sont périodiques et leur période correspond à peu près à une journée de télémesure (soit la période de révolution du satellite). Le prétraitement proposé se décompose en deux étapes : 1) Les signaux de télémesures sont découpés en fenêtres de 24h, et 2) les fenêtres sont projetées sur une base connue afin d'extraire les premiers coefficients. Chaque fenêtre est alors représentée par un vecteur de coefficients à partir duquel est réalisée la détection. Plusieurs bases ont été étudiées telles que des bases fixes de Fourier ou d'ondelettes ou des bases construites à partir de données d'apprentissage (ACP, KACP). L'algorithme de détection d'anomalies univariées à l'aide l'indice TNND est résumé dans le tableau 1.4.

TABLE 1.4 – L'algorithme TNND (Airbus)

Paramètres	Continus
Pré-traitement	1- Segmentation , $W = 24$h 2- Projection 3- Sélection automatique des coefficients
Apprentissage	Manuel
Méthode	Distance à noyau Noyau Laplacien
Score	TNND($\boldsymbol{x}$)
Détection	TNND($\boldsymbol{x}$) $> S_{\text{PFA}}$

1.4.4 Détection d'anomalies multivariées

1.4.4.1 MPPCAD

L'algorithme MPPCAD Tipping and Bishop (1999) (Mixture of Probabilistic principal component Analyzers) est un algorithme de classification qui a été appliqué à la détection d'anomalies multivariées dans la télémesure satellite par la JAXA Yairi et al. (2017), l'agence spatiale japonaise. Il repose sur un apprentissage semi-supervisé de modèles de mélanges de lois de probabilité et a l'avantage de pouvoir traiter conjointement des données continues et discrètes.
Considérons deux vecteurs $\boldsymbol{x}_{\text{C},t} \in \mathbb{R}^{P_\text{C}}$ et $\boldsymbol{x}_{\text{D},t} \in \mathbb{R}^{P_\text{D}}$ composés des mesures de P_C paramètres continus et P_D paramètres discrets acquis à un même instant t. Chaque paramètre discret $\boldsymbol{x}_\text{D}(j)(j = 1, ..., P_\text{D})$ prend ses valeurs dans l'ensemble $\{1, ..., M_j\}$ contenant M_j valeurs différentes. L'algorithme MPPCAD proposé pour la détection d'anomalies satellite repose sur l'hypothèse que les données de $\boldsymbol{x}_{\text{D},t}$ associées aux paramètres discrets sont distribuées selon un mélanges de lois discrètes ou lois catégorielles, et que les données de $\boldsymbol{x}_{\text{C},t}$ associées aux paramètres continus sont distribuées selon un mélange de lois gaussiennes. De plus, la dimension intrinsèque L des données continues est supposée inférieure à la dimension d'acquisition correspondant au nombre de paramètres continus étudiés. La distribution associée à chaque groupe peut alors être contrainte dans un espace de dimension réduit $L < P_\text{C}$. La densité de probabilité des données continues est donnée par

$$p(\boldsymbol{x}_{\text{C},t}|\Theta_\text{C}) = \sum_{g=1}^{G} \pi_g \mathcal{N}(\boldsymbol{x}_{\text{C},t}|\mu_g, \mathbf{C}_g) \tag{1.21}$$

où $\Theta_\text{C} = \{\pi_g, \mu_g, \mathbf{W}_g, \sigma_g^2, g = 1, ..., G\}$ est l'ensemble des hyperparamètres du modèle continu, μ_g est la moyenne et $\mathbf{C}_g = \mathbf{W}_g\mathbf{W}_g^T + \sigma_g^2\mathbf{I}_{P_\text{C}}$ est la matrice de covariance associée à la $g^{\text{ème}}$ distribution gaussienne avec $\mathbf{W}_g$ la matrice de facteurs estimés à l'aide des L premières valeurs propres et vecteurs propres de la matrice de covariance empirique, σ_g^2 la variance du bruit, $\mathbf{I}_{P_\text{C}}$ la matrice identité de taille $P_\text{C} \times P_\text{C}$ et π_g la probabilité a priori d'appartenir au $g^{\text{ème}}$

groupe. La densité de distribution des données discrètes est donnée par

$$p(\boldsymbol{x}_{\mathrm{D}}|\Theta_{\mathrm{D}}) = \sum_{g=1}^{G} \pi_g \prod_{j=1}^{K_{\mathrm{D}}} \mathrm{Cat}(\boldsymbol{y}_{\mathrm{D}}(j)|\boldsymbol{\theta}_{g,j}) \tag{1.22}$$

où $\Theta_{\mathrm{D}} = \{\theta_{g,j}, g = 1, ..., G, j = 1, ..., M_j\}$ est l'ensemble des hyperparamètres du modèle discret, Cat(.) est une distribution discrète, $\boldsymbol{x}_{\mathrm{D}}(j)$ est la $j^{\text{ème}}$ composante de $\boldsymbol{x}_{\mathrm{D}}$ et $\boldsymbol{\theta}_{g,j} = [\theta_{g,j,1}, ..., \theta_{g,j,M_j}]$ est le vecteur composé des hyperparamètres des distributions discrètes, i.e., $P(\boldsymbol{x}_{\mathrm{D}}(j) = l|g) = \theta_{g,j,l}$. Enfin, la distribution conjointe des données mixtes est obtenue en supposant l'indépendance des vecteurs $\boldsymbol{y}_C$ et $\boldsymbol{y}_D$

$$p(\boldsymbol{y}_{\mathrm{C}}, \boldsymbol{y}_{\mathrm{D}}|\Theta) = p(\boldsymbol{y}_{\mathrm{C}}|\Theta_{\mathrm{C}})p(\boldsymbol{y}_{\mathrm{D}}|\Theta_{\mathrm{D}}) \tag{1.23}$$

où $\Theta = \{\Theta_{\mathrm{C}}^T, \Theta_{\mathrm{D}}^T\}^T$. La phase d'apprentissage de l'algorithme MMPCAD consiste à déterminer les hyperparamètres de Θ à partir de données de télémesure saines, notées $\boldsymbol{x}_t, t = 1, ..., T$, et revient à maximiser la fonction de log-vraisemblance suivante

$$\mathcal{L}(\Theta) = \sum_t \ln\, p(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}|\Theta). \tag{1.24}$$

La maximisation de la log-vraisemblance peut être réalisée à l'aide de l'algorithme Espérance-Maximisation (EM) Dempster et al. (1997) qui à chaque itération alterne entre estimation de l'espérance de la vraisemblance à l'aide des observations et en considérant les hyperparamètres de Θ fixes (E-step), et estimation du maximum de vraisemblance des hyperparamètres en maximisant l'espérance déterminée à l'étape précédente. Notons que pour la première itération de l'algorithme EM, une initialisation des hyperparamètres de Θ est nécessaire.
Pour la détection d'anomalies, un score MPPCAD$(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}|\widehat{\Theta})$ est défini par

$$\mathrm{MPPCAD}(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}|\widehat{\Theta}) = -\ln\, p(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}|\widehat{\Theta}). \tag{1.25}$$

Ce score est comparé à un seuil de détection fixé par l'utilisateur et une anomalie est détectée si le score est supérieur au seuil, i.e

$$\text{Anomalie détectée si } \mathrm{MPPCAD}(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}|\widehat{\Theta}) > S\mathrm{PFA}. \tag{1.26}$$

L'algorithme MMPCAD a été appliqué à des données réelles de télémesure satellite. Les résultats obtenus après deux ans d'étude sur 89 paramètres continus et 356 paramètres discrets du satellite SDS-4 sont présentés dans Yairi et al. (2017). L'algorithme MPPAD impose de régler deux hyperparamètres, à savoir K le nombre de groupes et L la dimension intrinsèque des données continues. Après avoir réalisé une analyse en composantes principales (ACP) sur les données d'apprentissage, la dimension intrinsèque des données L est déterminée en recherchant "le coude" dans les valeurs propres de l'ACP appliqué au données d'apprentissage. Le nombre de groupes K est fixé manuellement en étudiant les nuages de points des composantes principales de l'ACP. Pour la phase d'apprentissage de l'algorithme, les auteurs préconisent d'initialiser les hyperparamètres de Θ afin d'éviter les minimums locaux. Pour se faire, il proposent une initialisation de l'algorithme EM Dempster et al. (1997) qui peut être décrite

en deux étapes : 1) application de l'algorithme de classification k-means pour déterminer les groupes, 2) exécution du M-step de l'algorithme EM pour l'estimation des hyperparamètres initiaux. L'algorithme MPPCAD pour la détection d'anomalies satellite est résumé dans le tableau 1.5. Les avantages et limites de cette méthodes seront étudiés plus en détail dans les chapitres suivants.

TABLE 1.5 – L'algorithme MPPCAD (JAXA)

Paramètres	Continus & Discrets
Pré-traitement	1- Suppression données aberrantes (OOL) + normalisation 2- Segmentation , $W = 1$
Apprentissage	Manuel
Méthode	MPPCAD K nombre de groupes : "règle du coude" ACP L dimension intrinsèque : étude des composantes principales de l'ACP Θ^0 initialisation : $k - means$ + M-step
Score	MPPCAD$(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}\|\widehat{\Theta})$
Détection	MPPCAD$(\boldsymbol{x}_{\mathrm{C},t}, \boldsymbol{x}_{\mathrm{D},t}\|\widehat{\Theta}) > S$PFA

1.4.4.2 Tests d'égalité de matrices de covariance

L'aspect multivarié de la télémesure satellite a également été étudié à travers l'étude des covariances entre les paramètres. Un test statistique d'égalité de deux matrices de covariance a été proposé par Airbus pour la détection d'anomalies multivariées Barreyre et al. (2017); Barreyre (2018). Ce test considère deux matrices de covariance $\mathbf{\Gamma}^{(1)}, \mathbf{\Gamma}^{(2)} \in \mathbb{R}^{m\times m}$ et cherche à tester les hypothèses suivantes

$$\mathrm{H}_0 : \mathbf{\Gamma}^{(1)} = \mathbf{\Gamma}^{(2)} = \mathbf{\Gamma} \tag{1.27}$$

$$\mathrm{H}_1 : \mathbf{\Gamma}^{(1)} \neq \mathbf{\Gamma}^{(2)} \tag{1.28}$$

où $\mathbf{\Gamma}$ est une matrice de référence. En pratique, les matrices $\mathbf{\Gamma}^{(1)}$ et $\mathbf{\Gamma}^{(2)}$ sont estimées empiriquement à partir des données de télémesure satellite. On notera $\widehat{\mathbf{\Gamma}}^{(1)}$ et $\widehat{\mathbf{\Gamma}}^{(2)}$ les matrices de covariance empirique associée à $\mathbf{\Gamma}^{(1)}$ et $\mathbf{\Gamma}^{(2)}$ et calculée à partir des n mesures de m paramètres de télémesure différents acquises sur une même période de temps donnée. La matrice $\mathbf{\Gamma}$ est alors estimée par la moyenne des matrices $\widehat{\mathbf{\Gamma}}^{(1)}$ et $\widehat{\mathbf{\Gamma}}^{(2)}$. La matrice $\mathbf{\Gamma}$ est exprimée par sa décomposition en valeurs singulières telle que

$$\mathbf{\Gamma} = \varphi \Lambda \varphi^T \tag{1.29}$$

où les d première valeurs propres, $\Lambda_{i,i} = \lambda_i, i = 1, ..., d$, sont supposées strictement décroissantes. la matrice φ issue de la décomposition de $\mathbf{\Gamma}$ est utilisée pour construire les matrices

$\Delta^{(1)}$ et $\Delta^{(2)}$ telles que

$$\Delta^{(1)} = \varphi \mathbf{\Gamma}^{(1)} \varphi \tag{1.30}$$

$$\Delta^{(2)} = \varphi \mathbf{\Gamma}^{(2)} \varphi \tag{1.31}$$

dont les termes diagonaux seront notés $\delta_i^{(1)} = \Delta_{i,i}^{(1)}$ et $\delta_i^{(2)} = \Delta_{i,i}^{(2)}$. La statistique de test proposée est donnée par

$$\mathrm{T} = \frac{n}{4} \left(\sum_{i=1}^{d} \left[ln \left(\frac{\delta_i^{(1)}}{\delta_i^{(2)}} \right) \right] \right). \tag{1.32}$$

Comme démontré dans Barreyre et al. (2017); Barreyre (2018), sous l'hypothèse nulle H_0, la statistique de test T converge en loi vers une loi du χ^2 lorsque le nombre de mesures n de chaque paramètre tend vers l'infini, i.e

$$\mathrm{T} \xrightarrow[n \to +\infty]{\mathcal{L}} \chi_d^2. \tag{1.33}$$

La règle de détection d'anomalies associée est définie par

$$\text{Rejet de } \mathrm{H}_0 \text{ si } \mathrm{T} \geq \chi_{d,1-\alpha}^2 \tag{1.34}$$

où α désigne le risque, consenti par l'utilisateur, de rejeter à tort l'hypothèse nulle H_0.
Ce test statistique a été testé sur des données réelles de télémesure de satellites géostationnaires pour détecter les changements brutaux entre deux fenêtres de télémesure consécutives de 24h. Cette stratégie de détection s'appuie sur le fait que les données de télémesure sont périodiques et repose sur l'hypothèse que deux fenêtres consécutives décrivent des comportements similaires des signaux. Ce test a été comparé à plusieurs autres tests de l'état de l'art et s'est montré très compétitif. Il est aujourd'hui intégré aux outils internes de Airbus et est utilisé pour diverses applications liées à la détection d'anomalies dans la télémesure satellite. Son exécution est peu coûteuse et il a l'avantage de pouvoir traiter conjointement plusieurs centaines de paramètres de télémesure. Cependant, il ne considère que des paramètres de télémesure continus et n'est pas facilement adaptable à lla détection d'anomalies dans des données discrètes ou des données mixtes. La détection d'anomalies à l'aide de ce test est résumée dans le tableau 1.6

TABLE 1.6 – Détection d'anomalies multivariées (Airbus)

Paramètres	Continus
Pré-traitement	1- Segmentation , $W = 24\mathrm{H}$ 2- Calcul de matrices de covariance empiriques
Méthode	Test statistique d'égalité de matrice de covariance
Statistique de Test	$\mathrm{T} = \frac{n}{4} \left(\sum_{i=1}^{d} \left[ln \left(\frac{\delta_i^{(1)}}{\delta_i^{(2)}} \right) \right] \right)$
Détection	$\mathrm{T} \geq \chi_{d,1-\alpha}^2$

1.5 Conclusion

Cette section nous plonge au coeur des opérations spatiales et plus précisément de la surveillance des satellites en exploitation à l'aide de méthodes de détection d'anomalies appliquées à la télémesure satellite. Conscients des limites des méthodes classiques, de nombreux acteurs du spatial tels que le CNES, Airbus, la JAXA ou encore le DLR se sont intéressés à de nouvelles approches basées sur un apprentissage semi-supervisé afin d'améliorer le suivi de l'état de santé de leurs satellites. Nous l'avons vu, plusieurs techniques ont été étudiées mais la majorité des algorithmes développés repose sur des méthodes de détection d'anomalies univariées qui traitent les paramètres de télémesure indépendamment les uns des autres. La détection des anomalies multivariées est un problème plus compliqué qui fait l'objet de cette thèse. Par ailleurs peu de méthodes ont considéré ce problème de détection d'anomalies pour des paramètres de télémesure discrets. Enfin, quelques solutions seulement ont été proposées pour prendre en compte le caractère multivarié de la télémesure satellite comme c'est le cas de l'algorithme MPPCAD utilisé par la JAXA ou le test statistique d'égalité de deux matrices de covariance proposé par Airbus. La détection d'anomalies multivariées dans la télémesure satellite reste donc un sujet de recherche largement ouvert. Cette thèse s'inscrit dans une volonté du CNES et de Airbus d'explorer ce thème et a pour objectif de proposer des algorithmes de détection d'anomalies multivariées capables de traiter de manière conjointe plusieurs signaux de télémesure de différentes natures (continu/discret) et de prendre en compte les corrélations et les relations qui peuvent exister entre eux afin de détecter les anomalies univariées et multivariées.

Chapitre 2

Estimation parcimonieuse

Sommaire

2.1 Introduction

À l'heure où l'acquisition de données n'est plus un problème, la question se pose de pouvoir traiter efficacement la masse de données accumulée pour en extraire le maximum d'information : bienvenue dans l'ère du Big Data ! Ce problème peut être simplifié en cherchant à prendre en compte non pas toute l'information disponible mais l'information la plus pertinente. Les méthodes d'estimation parcimonieuse et d'apprentissage de dictionnaires s'inscrivent dans cette logique. Elles consistent à décrire la donnée en contraignant la quantité d'information à utiliser. Plus concrètement, l'idée est de construire un dictionnaire uniquement composé des éléments les plus représentatifs à partir desquels il est possible de décrire toute nouvelle entrée par une combinaison linéaire de quelques éléments. Ces méthodes ont été appliquées avec succès pour de nombreuses applications telles que le débruitage Elad and Aharon (2006); Dong and Li (2010); Xu et al. (2017), la classification Huang and Aviyente (2006); Mei and Ling (2011); Oktar and Turkan (2018), la reconnaissance faciale Wright et al. (2009); Zhang and Li. (2011); Cheng et al. (2013) et plus récemment la détection d'anomalies univariées Adler et al. (2015). Cette thèse étudie l'intérêt de ces méthodes pour la détection d'anomalies multivariées dans des données mixtes de télémesure satellites.
Les sections qui suivent abordent quelques points théoriques relatifs à ces méthodes et présentent l'algorithme ADDICT (Anomaly Detection using a DICTionnary), une nouvelle méthode de détection d'anomalies multivariées qui repose sur un modèle d'estimation parcimonieuse inspiré de travaux menés dans Adler et al. (2015). Développée dans le but d'améliorer la surveillance de la télémesure satellite, la méthode ADDICT a l'avantage de pouvoir traiter conjointement plusieurs signaux de télémesure mixtes afin de prendre en compte les éventuelles corrélations et relations que peuvent entretenir les différents paramètres entre eux Pilastre et al. (2020a, 2019a). L'algorithme proposé nécessite de construire au préalable un dictionnaire composé de télémesures saines décrivant le fonctionnement normal du satellite. La nouvelle télémesure fraîchement acquise est ensuite décomposée sur ce dictionnaire et la détection d'anomalies est réalisée en analysant l'erreur d'estimation issue de la décomposition.
Ce chapitre est organisé comme suit

- la section 2.2 rappelle les principes de base de l'estimation parcimonieuse et de l'apprentissage de dictionnaire.
- la section 2.3 détaille la méthode ADDICT qui a été évaluée et comparée à d'autres méthodes concurrentes de l'état de l'art sur une base de données composée d'anomalies connues et identifiées par des experts.
- la section 2.4 présente une version généralisée et une version pondérée de l'algorithme ADDICT permettant d'adapter la méthode au cadre multivarié et d'y intégrer de l'information externe afin d'en améliorer ses performances de détection.
- la section 2.5 s'intéresse au réglage des hyperparamètres ADDICT à l'aide de données d'apprentissage associées à un fonctionnement normal du satellite.

2.2 Estimation parcimonieuse et apprentissage de dictionnaires

2.2.1 Estimation parcimonieuse

L'estimation parcimonieuse consiste à exprimer un signal observé $\boldsymbol{y} \in \mathbb{R}^N$ comme une combinaison linéaire de quelques éléments, appelés atomes, d'une base connue appelée dictionnaire $\boldsymbol{\Phi} \in \mathbb{R}^{N \times L}$, *i.e.*,

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{2.1}$$

où $\boldsymbol{x} \in \mathbb{R}^L$ est un vecteur parcimonieux composé de peu de valeurs non nulles. On suppose ici que le dictionnaire $\boldsymbol{\Phi}$ est connu et l'on cherche à estimer le vecteur $\boldsymbol{x}$ (l'apprentissage du dictionnaire sera abordé dans un second temps).
De manière classique, le problème d'estimation parcimonieuse consiste à trouver le vecteur $\boldsymbol{x}$ satisfaisant (2.1) et composé d'un minimum de valeurs non nulles, ce qui revient à résoudre le problème de minimisation suivant :

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \ \text{ s.c. } \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} \tag{2.2}$$

où s.c signifie sous contrainte et $\|\boldsymbol{x}\|_0 = \#\{x_i, i = 1, ..., L : x_i \neq 0\}$ (# désignant le cardinal d'un ensemble) est la norme ℓ_0 renvoyant le nombre de valeurs non nulles de $\boldsymbol{x}$. Dans la réalité, les données sont souvent affectées par un bruit additif. L'équation (2.1) peut alors être reformulée :

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{b} \tag{2.3}$$

où $\boldsymbol{b} \in \mathbb{R}^N$ est un bruit additif. Le problème d'estimation parcimonieuse (2.2) devient Soubies et al. (2015) :

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0 \ \text{ s.c. } \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \leq \epsilon. \tag{2.4}$$

où $\|\mathbf{u}\|_2 = \sqrt{\sum_i u_i^2}$ correspond à la norme euclidienne d'un vecteur $\mathbf{u}$. L'objectif de ce nouveau problème est d'estimer le vecteur $\boldsymbol{x}$ parcimonieux pour lequel l'erreur d'estimation associée, donnée par $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2$, n'excède pas le seuil $\epsilon > 0$ fixé par l'utilisateur. La contrainte permet ici d'assurer la qualité de l'estimation tout en acceptant la présence d'un bruit additif.
Il est également possible de contraindre explicitement le nombre de valeurs non nulles de $\boldsymbol{x}$ en fixant un nombre maximal k de valeurs non nulles autorisées, ce qui revient à résoudre

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \ \text{ s.c. } \|\boldsymbol{x}\|_0 \leq k. \tag{2.5}$$

Enfin, il peut être intéressant de travailler avec un forme non pénalisée de ce problème

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_0 \tag{2.6}$$

où λ est un paramètre de régularisation qui contrôle la parcimonie de $\boldsymbol{x}$. La mécanique est telle que, plus la valeur de ce paramètre est élevée, plus l'effort de minimisation est porté sur le vecteur $\boldsymbol{x}$, jusqu'à l'annuler ($\widehat{\boldsymbol{x}} = 0$) pour une valeur trop élevée de λ. À l'inverse, plus la valeur de λ est petite, plus l'effort de minimisation est porté sur le terme d'attache aux données, à savoir $\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2$, qui contrôle la qualité de l'estimation.

2.2.2 Algorithmes d'estimation parcimonieuse

La présence de la norme ℓ_0 dans les problèmes de minimisation présentés à la section précédente mène à des problèmes appelés NP-complets ou NP-difficiles car ils correspondent à des problèmes combinatoires qui nécessitent de tester toutes les configurations possibles de $\boldsymbol{x}$. Ceci est lié au fait qu'il s'agit de problèmes non-convexes pour lesquels l'unicité de la solution n'est pas garantie. Il existe alors deux grandes familles de solutions pour résoudre ces problèmes : les algorithmes gloutons et les algorithmes de résolution de problèmes convexes applicables après relaxation de la norme ℓ_0 par la norme ℓ_1 par exemple.

2.2.2.1 Les algorithmes gloutons

Les algorithmes gloutons tels que l'algorithme *Matching Poursuit* (MP) Mallat and Zhang (1993) ou sa variante orthogonale, l'algorithme *Orthogonal Matching Poursuit* (OMP) Pati and Rezaiifar (1993) permettent de résoudre les problèmes classiques d'estimation parcimonieuse présentés ci-dessus. L'algorithme MP consiste à trouver à chaque itération l'atome du dictionnaire le plus corrélé au signal d'entrée $\boldsymbol{y}$. La projection de $\boldsymbol{y}$ sur l'espace vectoriel engendré par les atomes du dictionnaire sélectionnés à chaque itération lui est ensuite soustraite. L'opération est répétée jusqu'à satisfaire un critère d'arrêt associé à la contrainte imposée. L'algorithme OMP est détaillé dans l'annexe A.1.

2.2.2.2 Relaxation de la norme ℓ_0

Une autre idée pour résoudre les problèmes (2.4),(2.5) et (2.6) est de relaxer la parcimonie en remplaçant la norme ℓ_0 par la norme ℓ_1 définie par $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$. Cette astuce permet de se ramener à un problème convexe dont l'unicité de la solution est assurée. Si l'on reprend les problèmes énumérés plus haut, nous obtenons pour (2.4)

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \ \text{ s.c. } \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \leq \epsilon \tag{2.7}$$

et pour (2.5)

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \ \text{ s.c. } \|\boldsymbol{x}\|_1 \leq k. \tag{2.8}$$

Ces deux problèmes sont connus sous le nom de Basis Poursuit Chen et al. (1998) et Basis Poursuit Denoising et peuvent être résolus par l'algorithme du simplex Dantzig (1990); Chen et al. (1996) ou par un algorithme du premier ordre si la dimension considérée est trop grande Nesterov and Nemirovski (2013). En remplaçant la norme ℓ_0 par la norme ℓ_1 dans le problème non pénalisé (2.6), on se ramène au problème bien connu du *Least Absolute Shrinkage Operator* (LASSO) Tibshirani (1996)

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1. \tag{2.9}$$

Le problème LASSO peut être résolu par un algorithme de descente de gradient Beck and Teboulle (2009a); Figueiredo and Nowak (2003); Fu (1998); Pendse (2011) ou par l'algorithme ADMM (*Alternative Direction Method of Multipliers*) Boyd et al. (2011). L'algorithme ADMM est itératif et met à jour les différentes composantes à estimer de manière alternée par minimisation du Lagrangien augmenté. Plus précisément il permet de résoudre des problèmes de minimisation convexe de la forme

$$\min_{\mathbf{X},\boldsymbol{z}} f(\mathbf{X},\boldsymbol{z}) \quad \text{s.c} \quad \mathbf{A}\mathbf{X}+\mathbf{B}\boldsymbol{z}=\mathbf{C} \tag{2.10}$$

où $\mathbf{X},\boldsymbol{z},\mathbf{A},\mathbf{B},\mathbf{C}$ sont des matrices de tailles appropriées et $f(\mathbf{X},\boldsymbol{z})=g(\mathbf{X})+h(\boldsymbol{z})$ est une fonction de coût séparable par rapport à $\mathbf{X}$ et $\boldsymbol{z}$. Le Lagrangien augmenté du problème (2.10) est défini par

$$\mathcal{L}_A(\mathbf{X},\boldsymbol{z},\mathbf{M},\mu)=f(\boldsymbol{y},\boldsymbol{z})+\langle\mathbf{M},\mathbf{A}\mathbf{X}+\mathbf{B}\mathbf{Z}-\mathbf{C}\rangle+\frac{\mu}{2}\parallel\mathbf{A}\mathbf{X}+\mathbf{B}\mathbf{Z}-\mathbf{C}\parallel_F^2. \tag{2.11}$$

À partir de ces considérations, en introduisant une variable auxiliaire notée $\boldsymbol{z}$ et une contrainte d'égalité nous pouvons reformuler le problème LASSO (2.9) sous la forme

$$\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2}\parallel\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\parallel_2^2+\lambda\parallel\boldsymbol{z}\parallel_1 \ \text{ s.c } \boldsymbol{z}=\boldsymbol{x}. \tag{2.12}$$

Cette reformulation nous permet de simplifier l'optimisation de $\boldsymbol{x}$, rendue difficile dans (2.9) par le facteur $\boldsymbol{\Phi}$. L'introduction de la variable $\boldsymbol{z}$ permet de découpler le terme d'attache aux données, à savoir $\parallel\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\parallel_2^2$, du terme en norme ℓ_1 et ainsi de séparer le problème initial en deux problèmes distincts plus simples à résoudre. Le Lagrangien augmenté associé au problème (B.1) est donnée par

$$\mathcal{L}_A(\boldsymbol{x},\boldsymbol{z},\mathbf{m},\mu)=\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{z}\|_1+\mathbf{m}^T(\boldsymbol{z}-\boldsymbol{x})+\frac{\mu}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2 \tag{2.13}$$

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 $\boldsymbol{z}$ et $\boldsymbol{x}$ par rapport à la contrainte d'égalité. La minimisation du Lagrangien augmenté par l'algorithme ADMM impose d'initialiser les variables $\boldsymbol{z}$ et $\mathbf{m}$ afin d'assurer la mise à jour itérative et alternée des différentes variables $\boldsymbol{x},\boldsymbol{z}$ et $\mathbf{m}$ dont l'expression à l'itération $i+1$ sont détaillées ci-dessous

- Mise à jour de $\boldsymbol{x}$:

La mise à jour de $\boldsymbol{x}$ consiste à résoudre le problème de minimisation suivant

$$\boldsymbol{x}^{i+1}=\arg\min_{\boldsymbol{x}}\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}\|_2^2+\mathbf{m}^i(\boldsymbol{z}^i-\boldsymbol{x})+\frac{\mu^i}{2}\|\boldsymbol{z}^i-\boldsymbol{x}\|_2^2, \tag{2.14}$$

ce qui conduit à l'aide de calculs algébriques simples à la solution suivante

$$\boldsymbol{x}^{i+1}=(\boldsymbol{\Phi}\boldsymbol{\Phi}^T+\mu^i I)^{-1}(\boldsymbol{\Phi}^T\boldsymbol{y}^i+\mathbf{m}^i+\mu^i\boldsymbol{z}^k). \tag{2.15}$$

- Mise à jour de $\boldsymbol{z}$:

La mise à jour de $\boldsymbol{z}$ consite à résoudre le problème de minimisation suivant

$$\boldsymbol{z}^{i+1} = \arg\min_{\boldsymbol{z}} \lambda\|\boldsymbol{z}\|_1 + (\mathbf{m}^i)^T(\boldsymbol{z} - \boldsymbol{x}^{i+1}) + \frac{\mu^i}{2}\|\boldsymbol{z} - \boldsymbol{x}^{i+1}\|_2^2. \tag{2.16}$$

Le problème (2.16) peut être reformulé pour se ramener à un problème de la forme

$$\boldsymbol{z}^{i+1} = \arg\min_{\boldsymbol{z}} \frac{1}{2}\|\mathbf{g} - \boldsymbol{z}\|_2^2 + \gamma^{i+1}\|\boldsymbol{z}\|_1. \tag{2.17}$$

dont la solution est donnée par l'opérateur de seuillage par élément tel que $\boldsymbol{z}_k^{i+1} = S_{\gamma^{i+1}}(\mathbf{g}_k^{i+1})$, avec $\boldsymbol{z}_k^{i+1}$ $\mathbf{g}_k^{i+1}$ respectivement la $k^{ème}$ composante de $\boldsymbol{z}^{i+1}$ et $\mathbf{g}^{i+1}$, $\mathbf{g}^{i+1} = \boldsymbol{x}^{i+1} - \frac{1}{\mu^i}\mathbf{m}^i$, $\gamma^{i+1} = \frac{\lambda}{\mu^i}$ et

$$S_{\gamma^{i+1}}(\mathbf{g}_k^{i+1}) = \begin{cases} \mathbf{g}_k^{i+1} - \gamma^{i+1} & \text{si } \mathbf{g}_k^{i+1} > \gamma^{i+1} \\ 0 & \text{si } |\mathbf{g}_k^{i+1}| \leq \gamma^{i+1} \\ \mathbf{g}_k^{i+1} + \gamma^{i+1} & \text{si } \mathbf{g}_k^{i+1} < -\gamma^{i+1} \end{cases} \tag{2.18}$$

La reformulation du problème (2.16) est détaillée dans l'annexe A.2 et une représentation graphique de l'opérateur de seuillage par éléments $S_\gamma(\mathbf{g}_k)$ est donnée dans la figure 2.1

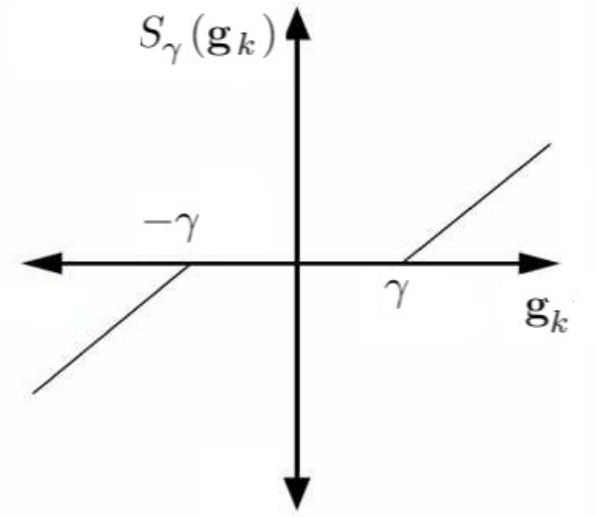

FIGURE 2.1 – Représentation graphique de l'opérateur de seuillage par groupe

- Mise à jour de $\mathbf{m}$:

La mise à jour de $\mathbf{m}$ est donnée par

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

La convergence de l'algorithme ADMM est assurée si le Lagrangien augmenté à minimiser

est convexe. Cette condition est satisfaite étant donné que (2.13) est une somme de fonctions convexes positives Boyd et al. (2011); Eckstein and Bertsekas (1992). L'algorithme ADMM sera appliqué à d'autres problèmes de minimisation posés dans ces travaux.

2.2.3 Parcimonie structurée : Group-LASSO

Dans certains cas, il peut être judicieux d'imposer une parcimonie de groupe si une information a priori est disponible quant à la structure du vecteur parcimonieux solution $\widehat{\boldsymbol{x}}$. Ce problème a été défini dans Yuan and Lin (2006) sous le nom de Group-LASSO

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda \sum_{j=1}^{J} w_j \|\boldsymbol{x}_{G_j}\|_2 \tag{2.19}$$

où w_j est le poids associé au groupe G_j et $\boldsymbol{x}_{G_j}$ est un sous-vecteur de $\boldsymbol{x}$ composé des coefficients du groupe G_j. Le problème (2.19) peut être résolu par les mêmes méthodes que celles utilisées pour résoudre le problème LASSO. Nous verrons par la suite que cette parcimonie structurée est très intéressante pour les applications qui nous occupent, à savoir la détection d'anomalies multivariées dans les données de télémesure satellite.

2.2.4 Relaxation pondérée (reweighted ℓ_1)

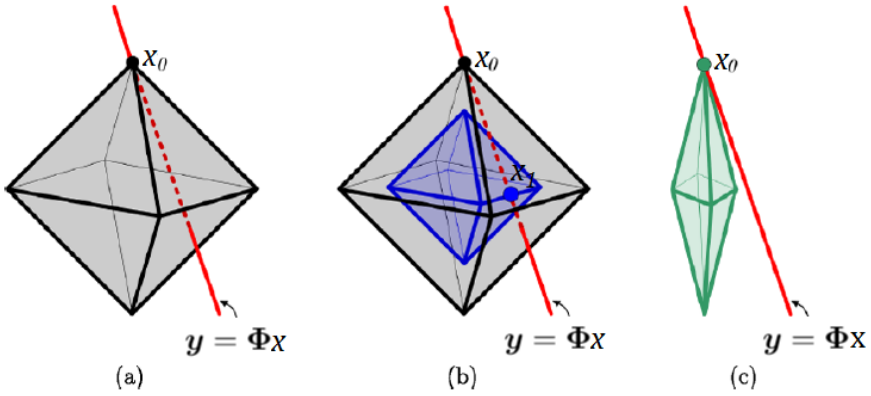

FIGURE 2.2 – Illustration en dimension 3 : (a) vecteur parcimonieux $\boldsymbol{x}_0$, contrainte $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ et boule ℓ_1 de rayon $\|\boldsymbol{x}_0\|_1$, (b) il existe un vecteur $\boldsymbol{x}_1 \neq \boldsymbol{x}_0$ tel que $\|\boldsymbol{x}_1\|_1 < \|\boldsymbol{x}_0\|_1$ et satisfaisant la contrainte, (c) boule $\|\boldsymbol{x}_0\|_1$ pondérée de sorte qu'il n'existe pas de vecteur $\boldsymbol{x}_1 \neq \boldsymbol{x}_0$ tel que $\|\mathbf{W}\boldsymbol{x}_1\|_1 < \|\mathbf{W}\boldsymbol{x}_0\|_1$ avec $\mathbf{W}$ une matrice de pondération.

Nous l'avons vu, la relaxation de la norme ℓ_0 par la norme ℓ_1 est une astuce classique pour résoudre un problème simple d'estimation parcimonieuse. Elle permet de se ramener à un problème convexe plus facile à résoudre et pour lequel l'unicité de la solution est assurée. Cependant, il n'est pas garanti que la solution retournée corresponde à la solution la plus parcimonieuse au sens strict. En effet, la norme ℓ_1 a tendance à privilégier la présence

de nombreux petits coefficients au détriment de la parcimonie associée à la norme ℓ_0. Par exemple, il est possible de trouver pour solution du problème LASSO, un vecteur de coefficients $\boldsymbol{x}_1 \neq \boldsymbol{x}_0$ tel que $\|\boldsymbol{x}_1\|_1 < \|\boldsymbol{x}_0\|_1$ et $\|\boldsymbol{x}_1\|_0 > \|\boldsymbol{x}_0\|_0$. La figure 2.2 tiré de Candès et al. (2008) permet d'illustrer ce phénomène. Pour pallier ce problème, Candès dans Candès et al. (2008) propose une version pondérée du problème LASSO (2.9) pouvant s'écrire

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{W}\boldsymbol{x}\|_1 \tag{2.20}$$

où $\mathbf{W} = \text{diag}(w_i)_{i=1,\dots,n}$ est une matrice diagonale de pondération. Les poids de la matrice $\mathbf{W}$ sont actualisés à chaque itération à partir des valeurs des coefficients comme détaillé dans l'algorithme 1. L'introduction de ces poids permet d'ajouter de l'information au modèle et d'améliorer l'estimation du vecteur $\boldsymbol{x}$. Cette idée sera étudiée pour la détection d'anomalies dans la section 2.3. L'algorithme de relaxation pondérée (reweighted ℓ_1) est rappelé ci-dessous

Algorithm 1 Algorithme *Reweighted-ℓ_1*

Initialiser $\mathbf{W}$
for $\ell = 0, \dots, \ell_{\max}$ **do**
  Résoudre $\boldsymbol{x}^{(\ell)} = \text{argmin}_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{W}^{(\ell)}\boldsymbol{x}\|_1$ (voir section 2.2.2)
  Mettre à jour les poids
  for $i = 1, \dots, N$ **do**
    $w_i^{(\ell+1)} = \frac{1}{\boldsymbol{x}_i^{(\ell)} + \epsilon}$
  end for
end for

2.2.5 Apprentissage de dictionnaire

Le dictionnaire constitue une base sur laquelle tout nouveau vecteur de données peut être décomposé. La qualité de l'estimation des nouvelles données repose donc en grande partie la qualité de cette base. Il existe deux types de dictionnaires : les dictionnaires paramétriques (base de Fourier, d'ondelettes etc.) et les dictionnaires appris à partir des données elles-mêmes (data-driven en anglais) qui se sont révélés très intéressants pour de nombreuses applications Tosic and Frossard (2011). Ces derniers peuvent être appris par une méthode d'analyse de données telle que l'analyse en composantes principales (ACP) ou par un algorithmes d'apprentissage de dictionnaires. Cette section s'intéresse à l'apprentissage de dictionnaires à partir des données d'apprentissage.

2.2.6 Formulation du problème

L'apprentissage de dictionnaire cherche à résoudre l'un des problèmes suivants

$$\widehat{\boldsymbol{x}}_i, \widehat{\boldsymbol{\Phi}} = \arg\min_{\boldsymbol{x}_i, \boldsymbol{\Phi}\in\mathcal{C}} \sum_i \|\boldsymbol{x}_i\|_0 \ \text{ s.c. } \|\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{x}_i\|_2^2 \leq \epsilon, \ \forall\ i \tag{2.21}$$

$$\widehat{\boldsymbol{x}}_i, \widehat{\boldsymbol{\Phi}} = \arg \min_{\boldsymbol{x}_i, \boldsymbol{\Phi} \in \mathcal{C}} \sum_i \|\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{x}_i\|_2^2 \text{ s.c. } \|\boldsymbol{x}_i\|_0 \leq k, \ \forall \ i \tag{2.22}$$

$$\widehat{\boldsymbol{x}}_i, \widehat{\boldsymbol{\Phi}} = \arg \min_{\boldsymbol{x}_i, \boldsymbol{\Phi} \in \mathcal{C}} \sum_i \|\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{x}_i\|_2^2 + \lambda \sum_i \|\boldsymbol{x}_i\|_0. \tag{2.23}$$

où

$$\mathcal{C} = \{\boldsymbol{\Phi} \in \mathbb{R}^{N \times L}, \|\boldsymbol{\phi_l}\|_2 \leq 1, \ \forall \ l\}$$

et $\boldsymbol{y}_i$ et $\boldsymbol{x}_i$ correspondent respectivement au $i^{ème}$ signal d'apprentissage et au $i^{ème}$ vecteur de coefficient qui lui est associé. La contrainte sur $\boldsymbol{\Phi}$ permet de lever l'ambiguïté d'échelle entre les atomes et le vecteur $\boldsymbol{x}$. Cette contrainte est importante pour que les atomes n'atteignent pas des valeurs arbitrairement élevées, ce qui permet d'obtenir des valeurs faibles (mais non nulles) de $\boldsymbol{x}$. Notons que la normalisation des atomes du dictionnaire peut être réalisée, de manière sous-optimale, à posteriori après la mise à jour du dictionnaire. Les problèmes (2.21), (2.22) et (2.23) peuvent être découplés en deux sous-problèmes plus simples à résoudre de manière itéartive et alternée. À chaque itération, une première étape estime le vecteur $\boldsymbol{x}$ en considérant les atomes $\boldsymbol{\phi}_l$ fixes, et une seconde étape met à jour les atomes avec le vecteur $\boldsymbol{x}$ fixe, estimé à l'étape précédente. La première étape revient à résoudre l'un des problèmes d'estimation parcimonieuse présenté à la section précédente. La seconde étape cherche à résoudre le problème suivant

$$\widehat{\boldsymbol{\Phi}} = \arg \min_{\boldsymbol{\Phi} \in \mathcal{C}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \text{ s.c } \|\phi_l\|_2 \leq 1, \ , \forall \ l. \tag{2.24}$$

qui peut être reformulé de manière non pénalisée

$$\widehat{\boldsymbol{\Phi}} = \arg \min_{\boldsymbol{\Phi} \in \mathcal{C}} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \mathcal{I}_{\mathcal{C}}(\boldsymbol{\Phi}) \tag{2.25}$$

en introduisant la fonction indicatrice $\mathcal{I}_{\mathcal{C}}$ telle que

$$\mathcal{I}_{\mathcal{C}}(\boldsymbol{\Phi}) = \begin{cases} 0 & \text{si } \boldsymbol{\Phi} \in \mathcal{C} \\ \infty & \text{sinon} \end{cases} \tag{2.26}$$

La mise à jour des atomes du dictionnaire peut être réalisée à l'aide d'une décomposition en valeur singulière (SVD)Aharon et al. (2006) ou par un algorithme de descente de gradient Mairal et al. (2009); Barthélemy et al. (2013).

2.2.7 Quelques algorithmes d'apprentissage de dictionnaires

Plusieurs algorithmes ont été proposés pour répondre au problème d'apprentissage de dictionnaire Engan et al. (1999); Aharon et al. (2006); Mairal et al. (2009); Barthélemy et al. (2013); W. Dong and Shi (2011); Zhang and Li. (2011). Dans ces travaux nous nous sommes principalement intéressés à deux algorithmes d'apprentissage de dictionnaires proposés dans la littérature : l'algorithme K-SVD et l'algorithme ODL (Online Dictionary Learning). Cette section présente de manière succincte la stratégie adoptée par chacun d'eux.

2.2.7.1 L'algorithme K-SVD

L'algorithme K-SVD, proposé par Aharon et al. (2006), est un algorithme d'apprentissage de dictionnaires inspiré du célèbre algorithme de classification K-Means. L'algorithme K-Means est un algorithme de classification non-supervisé qui permet de classer tous les éléments d'un nuage de points en K groupes les plus homogènes possibles. Une fois affecté à un groupe, chaque élément est représenté par l'élément moyen du groupe auquel il appartient. L'algorithme K-SVD peut être vu comme une généralisation de l'algorithme K-Means dans le sens où l'on cherche ici à représenter des signaux, non pas par un élément (l'élément moyen), mais par plusieurs éléments (quelques atomes du dictionnaire). À chaque itération de l'algorithme K-Means, les éléments moyens qui représentent chaque groupe sont actualisés selon les réaffectations. Par analogie, à chaque itération de l'algorithme K-SVD, les atomes sont mis à jour selon leur intervention dans la décomposition sur le dictionnaire des signaux d'apprentissage. Pour résoudre le problème d'estimation parcimonieuse (première étape), l'algorithme K-SVD utilise l'algorithme glouton OMP. Dans la deuxième étape, le dictionnaire $\boldsymbol{\Phi}$ ainsi que les coefficients de la combinaison linéaire $\boldsymbol{x}$ sont mis à jour à l'aide d'une décomposition en valeurs singulières (SVD). L'algorithme K-SVD est détaillé dans l'annexe A.1

2.2.7.2 L'algorithme ODL

L'algorithme ODL (Online Dictionary Learning) proposé par Mairal et al. (2009) est un algorithme d'apprentissage dit "en ligne" car à chaque itération le dictionnaire est actualisé en prenant en compte les informations des itérations précédentes (grâce à une reformulation astucieuse du problème de minimisation associé à la mise à jour du dictionnaire). L'algorithme ODL a été conçu pour apprendre des dictionnaires à partir de gros volumes de données d'apprentissage. Il utilise l'algorithme LARS-LASSO Efron et al. (2004); Osborne et al. (2000) pour la première étape de résolution du problème d'estimation parcimonieuse. La mise à jour du dictionnaire à la deuxième étape est effectuée par un algorithme de descente de gradient stochastique. L'algorithme ODL est détaillé dans l'annexe A.2.

2.3 L'algorithme ADDICT

Cette section présente l'algorithme ADDICT (Anomaly Detection using a DICTionary). Inspiré de travaux menés dans Adler et al. (2015), cet algorithme repose sur une méthode d'estimation parcimonieuse développée pour la détection d'anomalies dans des données mixtes de télémesure satellite. Ces travaux ont été publiés dans Pilastre et al. (2020a, 2019a).

2.3.1 Pré-traitement

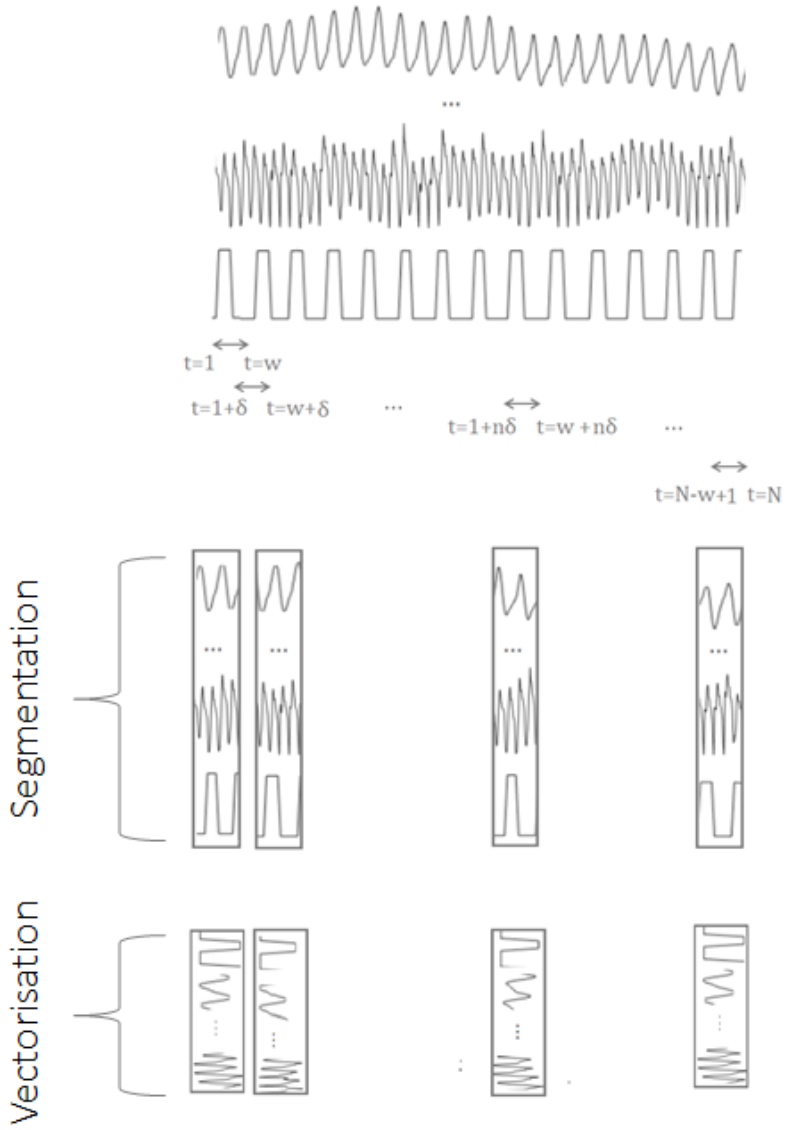

FIGURE 2.3 – Pré-traitement ADDICT.

Comme nous avons pu le voir précédemment, l'exploitation de données de télémesure satellite implique souvent en amont quelques étapes de pré-traitement (voir section 1.4.2). L'utilisation de l'algorithme ADDICT impose un pré-traitement en deux étapes : une étape de segmentation et une étape de vectorisation.
Dans un premier temps, les signaux de télémesure acquis sous forme de séries temporelles sont découpés en fenêtres de données multivariées de taille fixe W (ce découpage est souvent appelé "segmentation"). Ce découpage peut être réalisé de manière glissante afin d'extraire des fenêtres chevauchantes en segmentant les fenêtres avec un décalage fixe Δ. La période de chevauchement de deux fenêtres est alors de $W - \Delta$. Chaque fenêtre représente un intervalle

de temps durant lequel les mesures des différents paramètres considérés ont été acquises simultanément. La taille des fenêtres est un paramètre fixé par l'utilisateur et dépend du contexte opérationnel et /ou de la stratégie de surveillance adoptée (surveillance orbite par exemple).
Les fenêtres obtenues à l'issue de la segmentation 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 télémesure 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 multivariées. Le pré-traitement de l'algorithme ADDICT est résumé dans la figure 2.3.

2.3.2 Formulation du problème

À partir d'un dictionnaire $\boldsymbol{\Phi} \in \mathbb{R}^{N \times L}$ composé de L atomes (formant les colonnes de $\boldsymbol{\Phi}$) décrivant le fonctionnement normal du satellite, la stratégie ADDICT consiste à décomposer le signal multivarié associé à une nouvelle fenêtre de télémesure acquise, noté $\boldsymbol{y} \in \mathbb{R}^N$, comme une somme de trois signaux : un signal nominal issu de la décomposition sur le dictionnaire, un signal d'anomalie $\mathbf{e} \in \mathbb{R}^N$ décrivant le caractère atypique du signal d'intérêt et un bruit additif $\mathbf{b} \in \mathbb{R}^N$, i.e.,

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \mathbf{e} + \mathbf{b} \tag{2.27}$$

où $\boldsymbol{\Phi x} \in \mathbb{R}^N$ représente le signal nominal associé à $\boldsymbol{y}$ avec $\boldsymbol{x} \in \mathbb{R}^L$ un vecteur parcimonieux. Cette décomposition suppose qu'un signal $\boldsymbol{y}$ sain (sans anomalie) peut être correctement estimé par une combinaison linéaire de quelques atomes du dictionnaire. L'erreur d'estimation d'un signal sain ou résidu de la décomposition parcimonieuse sur le dictionnaire s'apparente alors à un bruit additif $\boldsymbol{b}$ et le signal d'anomalie $\boldsymbol{e}$ est nul. Lorsqu'une anomalie affecte le signal étudié, le modèle (2.27) suppose que le résidu peut être divisé en deux parties : un signal d'anomalie $\boldsymbol{e} \neq 0$ et un signal de bruit additif $\boldsymbol{b}$.
Étant donné le cadre multivarié dans lequel nous nous plaçons et le pré-traitement appliqué, chaque signal peut être divisé en K sous-signaux, où K est le nombre de paramètres de télémesure étudiés, i.e., $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_K^T]^T$ et $\mathbf{e} = [\mathbf{e}_1^T, ..., \mathbf{e}_K^T]^T, k = 1, ..., K$. De plus, dans le cas de données mixtes contenant des paramètres continus et discrets, il convient de noter que les N_D premières composantes sont associées à des paramètres discrets tandis que les N_C dernières composantes sont associées à des paramètres continues (avec $N = N_D + N_C$). Ainsi par exemple, le signal $\boldsymbol{y}$ est composé d'une partie discrète, notée $\mathbf{y}_\mathrm{D}$, et d'une partie continue, notée $\mathbf{y}_\mathrm{C}$, tel que $\mathbf{y} = (\mathbf{y}_\mathrm{D}^T, \mathbf{y}_\mathrm{C}^T)^T$. Enfin, ces deux signaux se divisent respectivement en K_D et K_C sous-signaux tels que $\mathbf{y}_\mathrm{D} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_{K_\mathrm{D}}^T]^T \in \mathbb{R}^{N_\mathrm{D}}$ et $\mathbf{y}_\mathrm{C} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_{K_\mathrm{C}}^T]^T \in \mathbb{R}^{N_\mathrm{C}}$ avec K_D et K_C les nombres de paramètres discrets et continus.
L'originalité de l'algorithme ADDICT réside dans sa capacité à traiter conjointement des séries temporelles discrètes et continues et à détecter à la fois des anomalies univariées et multivariées. Pour ce faire, il considère un dictionnaire de données mixtes $\boldsymbol{\Phi} \in \mathbb{R}^{N \times 2L}$ défini comme suit

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_\mathrm{D} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_\mathrm{C} \end{bmatrix}$$

où $\boldsymbol{\Phi}_{\mathrm{D}}$ et $\boldsymbol{\Phi}_{\mathrm{C}}$ sont respectivement composés d'atomes discrets et d'atomes continus. Les deux dictionnaires $\boldsymbol{\Phi}_{\mathrm{D}} \in \mathbb{R}^{N_{\mathrm{D}} \times L}$ et $\boldsymbol{\Phi}_{\mathrm{C}} \in \mathbb{R}^{N_{\mathrm{C}} \times L}$ ont été extraits d'un dictionnaire composé de L atomes mixtes. En d'autres termes, le $l^{\text{ème}}$ atome discret de $\boldsymbol{\Phi}_{\mathrm{D}}$ et le $l^{\text{ème}}$ atome continu de $\boldsymbol{\Phi}_{\mathrm{C}}$ sont issus d'un même atome mixte composé de mesures des différents paramètres acquises aux mêmes instants. L'algorithme ADDICT propose deux stratégies distinctes selon la nature des données et cherche à résoudre les problèmes suivants

$$\widehat{\boldsymbol{x}}_{\mathrm{D}}, \widehat{\boldsymbol{e}}_{\mathrm{D}} = \arg \min_{\boldsymbol{x}_{\mathrm{D}} \in \mathcal{B}, \boldsymbol{e}_{\mathrm{D}} \in \mathbb{R}^{N_{\mathrm{D}}}} \| \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\Phi}_{\mathrm{D}} \boldsymbol{x}_{\mathrm{D}} - \boldsymbol{e}_{\mathrm{D}} \|_2^2 + b_{\mathrm{D}} \sum_{k=1}^{K_{\mathrm{D}}} \|\mathbf{e}_{\mathrm{D},k}\|_2 \tag{2.28}$$

Estimation parcimonieuse discrète

$$\widehat{\boldsymbol{x}}_{\mathrm{C}}, \widehat{\boldsymbol{e}}_{\mathrm{C}} = \arg \min_{\boldsymbol{x}_{\mathrm{C}}, \mathbf{e}_{\mathrm{C}}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}} \boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\boldsymbol{x}_{\mathrm{C}}\|_1 + b_{\mathrm{C}} \sum_{k=1}^{K_{\mathrm{C}}} \|\mathbf{e}_{\mathrm{C},k}\|_2 \tag{2.29}$$

Estimation parcimonieuse continue

où $\boldsymbol{\Phi}_{\mathcal{M}}$ est un sous-dictionnaire de $\boldsymbol{\Phi}_{\mathrm{C}}$ utilisé pour représenter le signal continu $\boldsymbol{y}_{\mathrm{C}}$ de manière à prendre en compte les relations entre les données discrètes et continues (la construction et l'intérêt de $\boldsymbol{\Phi}_{\mathcal{M}}$ seront explicités dans les sections qui suivent), $b_{\mathrm{D}}, a_{\mathrm{C}}$ et b_{C} sont les hyperparamètres du modèle permettant de contrôler respectivement la parcimonie de $\boldsymbol{e}_{\mathrm{D}}, \boldsymbol{x}_{\mathrm{C}}$ et $\boldsymbol{e}_{\mathrm{C}}$, $\mathcal{B}$ représente l'espace canonique ou base naturelle de $\mathbb{R}^L$, c'est-à-dire que $\mathcal{B} = \{\boldsymbol{\epsilon}_l, l = 1, \cdots, L\}$, avec $\boldsymbol{\epsilon}_l$ un vecteur dont la $l^{\text{ème}}$ composante est égale à 1 et dont les autres composantes sont égales à 0. Le vecteur $\boldsymbol{e}_{\mathrm{D},k}, k = 1, ..., K_{\mathrm{D}}$ (resp. $\boldsymbol{e}_{\mathrm{C},k}, k = 1, ..., K_{\mathrm{C}}$) correspond au $k^{\text{ème}}$ sous-signal de $\boldsymbol{e}_{\mathrm{D}}$ (resp. de $\boldsymbol{e}_{\mathrm{C}}$) associé au $k^{\text{ème}}$ paramètre discret (resp. continu).
Chacun de ces deux problèmes (2.28) et (2.29) impose une double parcimonie : une sur les vecteurs de coefficients $\boldsymbol{x}_{\mathrm{D}}$ et $\boldsymbol{x}_{\mathrm{C}}$ et une sur les signaux d'anomalie $\boldsymbol{e}_{\mathrm{D}}$ et $\boldsymbol{e}_{\mathrm{C}}$. La première correspond à notre hypothèse de départ selon laquelle un signal qui n'est pas affecté par une anomalie peut être correctement estimé par une combinaison linéaire de quelques atomes du dictionnaire associés au fonctionnement normal du satellite. On notera que cette contrainte de parcimonie ne se formalise pas de la même manière selon que les données soient discrètes ou continues. Pour les données continues, le problème s'apparente à un problème classique d'estimation parcimonieuse et la parcimonie est assurée par la norme ℓ_1, tandis que pour les données discrètes une parcimonie spécifique est introduite. En effet, l'estimation parcimonieuse discrète proposée ici consiste à estimer le signal discret d'intérêt $\boldsymbol{y}_{\mathrm{D}}$ par un seul atome du dictionnaire. Cette contrainte est assurée par l'appartenance du vecteur $\boldsymbol{x}_{\mathrm{D}}$ à l'espace canonique de $\mathbb{R}^L$. Cette stratégie s'est montrée plus efficace que les méthodes d'estimation parcimonieuse classiques pour les données discrètes. La deuxième contrainte de parcimonie concerne les signaux d'anomalies $\boldsymbol{e}_{\mathrm{D}}$ et $\boldsymbol{e}_{\mathrm{C}}$ et traduit le fait que les anomalies satellite sont rares et affectent peu de paramètres en même temps. En théorie et sous réserve d'hyperparamètres bien choisis, l'estimation des signaux d'anomalies mène donc souvent à des signaux nuls, ou à des signaux parcimonieux par groupe. Cette parcimonie est assurée par le dernier terme de (2.28) et (2.29) qui s'apparente à une parcimonie de type Group-LASSO dans laquelle les poids associés à chaque groupe sont tous égaux à 1 (voir section 2.2.3). La stratégie de décomposition de l'algorithme ADDICT ainsi décrite est résumé dans la figure 2.4.

La stratégie d'estimation parcimonieuse ADDICT de données mixtes consiste à découpler le problème en deux sous-problèmes plus adaptés à la nature des données. Cependant, afin d'assurer la détection des anomalies multivariées, notamment entre les paramètres discrets et les paramètres continus, il est important de ne pas totalement dissocier l'estimation discrète de l'estimation continue et de pouvoir prendre en compte les relations qui peuvent exister entre les paramètres. Dans la méthode ADDICT, la conservation des liens entre les paramètres discrets et continus et donc de la détection des anomalies multivariées est assurée par une sélection d'atomes. Cette dernière est réalisée lors de la première étape de l'algorithme, à savoir la résolution du problème d'estimation parcimonieuse discrète et permet de réaliser l'estimation des données continues uniquement à partir des atomes les plus représentatifs au regard du signal discret $\boldsymbol{y}_{\mathrm{D}}$. La résolution des problèmes (2.28) et (2.29) et la sélection des atomes sont détaillées dans la section suivante.

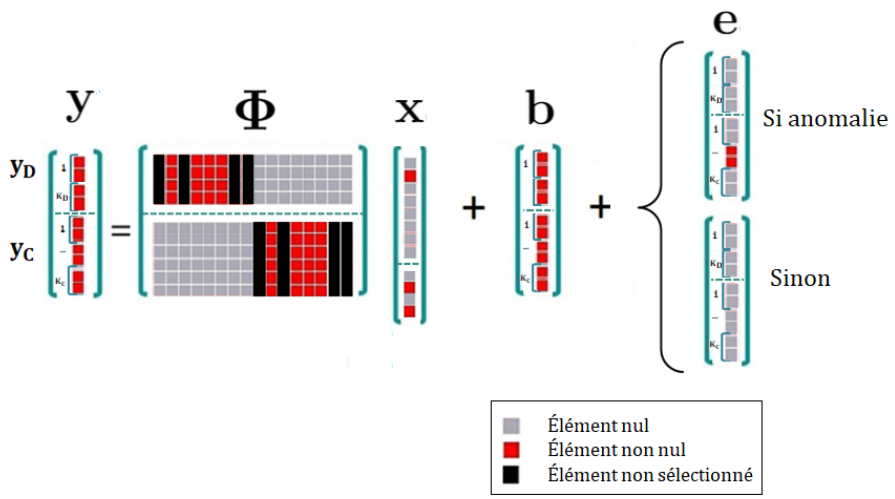

FIGURE 2.4 – Illustration de la stratégie de décomposition ADDICT.

2.3.3 Résolution du problème

2.3.3.1 Estimation parcimonieuse discrète

Étant donné la contrainte de parcimonie imposée au vecteur $\boldsymbol{x}_{\mathrm{D}}$, le problème de minimisation (2.28) associé à l'estimation parcimonieuse discrète est un problème combinatoire qui nécessite de résoudre pour tous les atomes discrets, notés $\boldsymbol{\phi}_{\mathrm{D},l}, l = 1, ..., L$, le problème suivant

$$\widehat{\mathbf{e}}_{\mathrm{D},l} = \arg\min_{\mathbf{e}_{\mathrm{D},l}} \|\boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D},l} - \boldsymbol{e}_{\mathrm{D},l}\|_2^2 + b_{\mathrm{D}} \sum_{k=1}^{K_{\mathrm{D}}} \|\mathbf{e}_{\mathrm{D},k}\|_2 \tag{2.30}$$

où $\boldsymbol{\phi}_{\mathrm{D},l}$ est la $l^{\text{ème}}$ colonne de $\boldsymbol{\Phi}_{\mathrm{D}}$. La solution du problème (2.30) est classiquement donnée par l'opérateur de seuillage par groupe tel que $\widehat{\mathbf{e}}_{\mathrm{D},l} = T_{b_{\mathrm{D}}}(\mathbf{h}_{\mathrm{D},l})$, avec $\mathbf{h} = \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D},l}$ et

$$[T_{b_{\mathrm{D}}}(\mathbf{h}_{\mathrm{D},l})]_{\mathrm{k}} = \begin{cases} \frac{\|\mathbf{h}_{\mathrm{D},l,k}\|_2 - b_{\mathrm{D}}}{\|\mathbf{h}_{\mathrm{D},l,k}\|_2} \mathbf{h}_{\mathrm{D},l,k} & \text{si } \|\mathbf{h}_{\mathrm{D},l,k}\|_2 > b_{\mathrm{D}} \\ 0 & \text{sinon} \end{cases} \tag{2.31}$$

où $\mathbf{h}_{\mathrm{D},l,k}$ est le $k^{\text{ème}}$ sous-signal de $\mathbf{h}_{\mathrm{D},l}$ associé au $k^{\text{ème}}$ paramètre discret avec $k = 1, ..., K_{\mathrm{D}}$. À l'issue de la résolution du problème (2.30), l'atome $\boldsymbol{\phi}_{\mathrm{D},l}$ est sélectionné si le signal d'anomalie $\widehat{\mathbf{e}}_{\mathrm{D},l}$ qui lui est associé est nul [1]. Cela revient à construire un ensemble $\mathcal{M}$ défini par

$$\mathcal{M} = \{l \in \{1, \cdots, L\} |\ \|\widehat{\mathbf{e}}_{\mathrm{D},l}\|_2 = 0\}. \tag{2.32}$$

Il convient de noter que $\mathcal{M}$ contient les valeurs de $l, l = 1, ..., L$, associées aux atomes discrets de $\boldsymbol{\Phi}_{\mathrm{D}}$ les plus proches de $\boldsymbol{y}_{\mathrm{D}}$ au sens de l'erreur d'estimation $\|\mathbf{h}_{\mathrm{D},l}\|_2$ et pour l'hyperparamètre b_{D} fixé. L'ensemble $\mathcal{M}$ permettra dans un second temps de construire le dictionnaire $\Phi_{\mathcal{M}}$, composé des atomes $\boldsymbol{\phi}_{\mathrm{C},l}$, qui sera utilisé pour résoudre le problème d'estimation des données continues. Le paramètre de régularisation b_{D} joue un rôle important dans la détection des anomalies. En effet, plus la valeur de b_{D} est faible, plus il y aura de détections sur les données discrètes (au risque de retourner de nombreuses fausses alarmes), et plus le nombre d'atomes sélectionnés pour estimer le signal continu $\boldsymbol{y}_{\mathrm{C}}$ sera faible. Cependant, un nombre insuffisant d'atomes sélectionnés rend plus difficile l'estimation du signal continu et peut être source de fausses alarmes. Inversement, plus la valeur de b_{D} est élevée, plus un nombre important d'atomes est sélectionné et meilleure est l'estimation de $\boldsymbol{y}_{\mathrm{C}}$. En revanche, le risque de rater certaines anomalies augmente.

2.3.3.2 Estimation parcimonieuse continue

Le problème d'estimation parcimonieuse continue (2.29) s'apparente à un problème classique d'estimation parcimonieuse de type LASSO dans lequel le signal d'intérêt est estimé par une combinaison linéaire de quelques atomes du dictionnaire. Le dictionnaire $\boldsymbol{\Phi}_{\mathcal{M}}$ considéré pour ce problème, ne contient que les atomes continus $\phi_{\mathrm{C},l}, l = 1, ..., L$ dont la partie discrète $\boldsymbol{\phi}_{\mathrm{D},l}$ a été sélectionnée à l'étape précédente (c'est-à-dire, $l \in \mathcal{M}$) de façon à prendre en compte les liens entre les paramètres discrets et continus. Ce problème peut être résolu par l'algorithme ADMM (voir section 2.2.2.2) après avoir introduit une variable auxiliaire notée $\boldsymbol{z}$ et une contrainte d'égalité menant au problème suivant :

1. On pourrait généraliser cette approche en conservant les atomes associés à un signal d'anomalie de norme inférieure à un seuil fixé par l'utilisateur

$$\arg\min_{\boldsymbol{x},\boldsymbol{e},\boldsymbol{z}} \frac{1}{2} \parallel \boldsymbol{y} - \boldsymbol{\Phi}_{\mathcal{M}}\boldsymbol{x}_{\mathrm{C}} - \boldsymbol{e} \parallel_2^2 + a_{\mathrm{C}} \parallel \boldsymbol{z} \parallel_1 + b_C \sum_{k=1}^{K_{\mathrm{C}}} \parallel \boldsymbol{e}_k \parallel_2$$

$$\text{s.c. } \boldsymbol{z} = \boldsymbol{x}_{\mathrm{C}}. \tag{2.33}$$

Le Lagrangien augmenté associé au problème (2.33) est donné par

$$\mathcal{L}_A(\boldsymbol{x}_{\mathrm{C}}, \boldsymbol{z}, \mathbf{e}_{\mathrm{C}}, \mathbf{m}, \mu) = \frac{1}{2}\|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}}\boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}}\|\boldsymbol{z}\|_1$$

$$+b_{\mathrm{C}} \sum_{k=1}^{K_{\mathrm{C}}} \|\mathbf{e}_{\mathrm{C,k}}\|_2 + \mathbf{m}_{\mathrm{C}}^T(\boldsymbol{z} - \boldsymbol{x}_{\mathrm{C}}) + \frac{\mu_{\mathrm{C}}}{2}\|\boldsymbol{z} - \boldsymbol{x}_{\mathrm{C}}\|_2^2 \tag{2.34}$$

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 $\boldsymbol{z}$ et $\boldsymbol{x}_{\mathrm{C}}$ par rapport à la contrainte d'égalité. Les mises à jour des composantes $\boldsymbol{x}_{\mathrm{C}}, \boldsymbol{z}$ et $\mathbf{m}$ à l'itération $i+1$ sont détaillées ci-dessous

- Mise à jour de $\boldsymbol{x}_{\mathrm{C}}$:

La mise à jour de $\boldsymbol{x}_{\mathrm{C}}$ découle du problème de minimisation suivant

$$\widehat{\boldsymbol{x}}_{\mathrm{C}}^{i+1} = \arg\min_{\boldsymbol{x}_{\mathrm{C}}} \frac{1}{2}\|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}}\boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}^i\|_2^2 + \mathbf{m}_{\mathrm{C}}^i(\boldsymbol{z}^i - \boldsymbol{x}_{\mathrm{C}}) + \frac{\mu_{\mathrm{C}}^i}{2}\|\boldsymbol{z}^i - \boldsymbol{x}_{\mathrm{C}}\|_2^2. \tag{2.35}$$

De simples calculs algébriques mènent à la solution suivante

$$\widehat{\boldsymbol{x}}_{\mathrm{C}}^{i+1} = (\boldsymbol{\Phi}_{\mathcal{M}}\boldsymbol{\Phi}_{\mathcal{M}}^T + \mu_{\mathrm{C}}^i I)^{-1}(\boldsymbol{\Phi}_{\mathcal{M}}^T(\boldsymbol{y}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}^i) + \mathbf{m}_{\mathrm{C}}^i + \mu_{\mathrm{C}}^i \boldsymbol{z}^k) \tag{2.36}$$

- Mise à jour de $\boldsymbol{z}$:

La mise à jour de $\boldsymbol{z}$ est issue du problème de minimisation suivant

$$\widehat{\boldsymbol{z}}^{i+1} = \arg\min_{\boldsymbol{z}} a_{\mathrm{C}}\|\boldsymbol{z}\|_1 + (\mathbf{m}_{\mathrm{C}}^i)^T(\boldsymbol{z} - \boldsymbol{x}_{\mathrm{C}}^{i+1}) + \frac{\mu_{\mathrm{C}}^i}{2}\|\boldsymbol{z} - \boldsymbol{x}_{\mathrm{C}}^{i+1}\|_2^2. \tag{2.37}$$

La solution du problème (2.37) est donnée par l'opérateur de seuillage par élément tel que $\boldsymbol{z}_k^{i+1} = S_{\gamma^{i+1}}(\mathbf{g}_k^{i+1})$, avec $\boldsymbol{z}_k^{i+1}$ et $\mathbf{g}_k^{i+1}$ respectivement la $k^{ème}$ composante de $\boldsymbol{z}^{i+1}$ et $\mathbf{g}^{i+1}$, $\mathbf{g}^{i+1} = \boldsymbol{x}^{i+1} - \frac{1}{\mu^i}\mathbf{m}^i$, $\gamma^{i+1} = \frac{\lambda}{\mu^i}$ et

$$S_{\gamma^{i+1}}(\mathbf{g}_k^{i+1}) = \begin{cases} \mathbf{g}_k^{i+1} - \gamma^{i+1} & \text{si } \mathbf{g}_k^{i+1} > \gamma^{i+1} \\ 0 & \text{si } |\mathbf{g}_k^{i+1}| \leq \gamma^{i+1} \\ \mathbf{g}_k^{i+1} + \gamma^{i+1} & \text{si } \mathbf{g}_k^{i+1} < -\gamma^{i+1} \end{cases} \tag{2.38}$$

- Mise à jour de $\boldsymbol{e}_{\mathrm{C}}$:

La mise à jour du signal d'anomalie continu est identique à celle du signal d'anomalie discret et revient à résoudre le problème de minimisation suivant

$$\widehat{\mathbf{e}}_{\mathrm{C}}^{i+1} = \arg\min_{\mathbf{e}_{\mathrm{C}}} \frac{1}{2} \parallel \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{i+1} - \mathbf{e} \parallel_2^2 + b \sum_{k=1}^{K_{\mathrm{C}}} \parallel \mathbf{e}_{\mathrm{C},k} \parallel_2 . \tag{2.39}$$

De manière analogue au problème d'estimation parcimonieuse discret, la solution du problème (2.39) est donnée par l'opérateur de seuillage par groupe $\widehat{\mathbf{e}}_{\mathrm{C}} = T_{b_{\mathrm{C}}}(\mathbf{h}_{\mathrm{C}}^{i+1})$, avec $\mathbf{h}_{\mathrm{C}}^{i+1} = \boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathrm{C}}\widehat{\boldsymbol{x}}_{\mathrm{C}}^{i+1}$ (la définition de l'opérateur d seuillage par groupe est donnée dans la section précédente).

- Mise à jour de $\mathbf{m}$:

La mise à jour de $\mathbf{m}$ est donnée par

$$\widehat{\mathbf{m}}^{i+1} = \mathbf{m}^i + \mu^i(\boldsymbol{z}^{i+1} - \boldsymbol{x}^{i+1}). \tag{2.40}$$

La résolution du problème d'estimation parcimonieuse continu (2.29) par l'algorithme ADMM est détaillée dans Boyd et al. (2011) et résumée dans l'algorithme 2. La convergence de l'algorithme ADMM est assurée si le Lagrangien augmenté à minimiser est convexe. Cette condition est satisfaite ici étant donné que (2.34) est une somme de fonctions convexes positives Boyd et al. (2011); Eckstein and Bertsekas (1992) .

Algorithm 2 $\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{e}}, \widehat{\boldsymbol{z}}$ =ADMM$(\boldsymbol{y}, \boldsymbol{\Phi})$

Initialiser i=1, $\boldsymbol{z}^0, \mathbf{e}^0, \mathbf{m}^0, \mu^0, \rho, a_{\mathrm{C}}, b_{\mathrm{C}}$
repeat
$\boldsymbol{x}^{i+1} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \mu^i I)^{-1}[\boldsymbol{\Phi}^T(\boldsymbol{y} - \mathbf{e}^i) + \mathbf{m}^i + \mu^i\boldsymbol{z}^i]$
$\boldsymbol{z}^{i+1} = S_{\gamma^{i+1}}(\boldsymbol{x}^{i+1} - \frac{1}{\mu^i}\mathbf{m}^i), \gamma = \frac{a}{\mu^i}$
$\mathbf{e}^{i+1} = T_b(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x})$
$\mathbf{m}^{i+1} = \mathbf{m}^i + \mu^i(\boldsymbol{z}^{i+1} - \boldsymbol{x}^{i+1})$
$\mu^{i+1} = \rho\mu^i$
i = i+1
until critère d'arrêt

2.3.3.3 Règle de décision

Un score d'anomalie, noté $s(\boldsymbol{y})$ est défini à partir des signaux d'anomalies $\widehat{\mathbf{e}}_{\mathrm{D}}$ et $\widehat{\mathbf{e}}_{\mathrm{C}}$ tel que

$$s(\boldsymbol{y}) = \begin{cases} -1 & \text{si } \|\widehat{\mathbf{e}}_{\mathrm{D}}\|_2 > 0 \text{ (i.e., } \mathcal{M} = \emptyset) \\ \parallel \widehat{\boldsymbol{e}}_{\mathrm{C}} \parallel_2 & \text{sinon.} \end{cases} \tag{2.41}$$

La règle de décision appliquée au score est la suivante

$$\text{Anomalie détectée si} \begin{cases} s(\boldsymbol{y}) = -1 \quad \text{(anomalie discrète)} \\ \text{ou} \\ s(\boldsymbol{y}) > S_{\text{PFA}} \quad \text{(anomalie continue/multivariée)} \end{cases} \tag{2.42}$$

où S_{PFA} est un seuil fixé par l'utilisateur. Ce seuil peut être déterminé à l'aide de caractéristiques opérationnelles du récepteur (CORs) si une vérité terrain est disponible, ou par l'expert de manière à respecter certaines contraintes relatives aux non-détections ou aux fausses alarmes. L'algorithme ADDICT est résumé dans l'algorithme 3

Algorithm 3 $\widehat{\boldsymbol{x}}_D, \widehat{\boldsymbol{e}}_D, \widehat{\boldsymbol{x}}_C, \widehat{\boldsymbol{e}}_C, \mathcal{M}$=ADDICT$(\boldsymbol{y}, \boldsymbol{\Phi})$

1- Estimation parcimonieuse discrète et sélection d'atomes
for $l = 1$ to L **do**
$\widehat{\mathbf{e}}_{\text{D},l} = T_{b_{\text{D}}}\left[\boldsymbol{y}_{\text{D}} - \boldsymbol{\phi}_{\text{D},l}\right]$
end for
$\mathcal{M} = \{l \in \{1, ..., L\} | \|\widehat{\mathbf{e}}_{\text{D},l}\|_2 = 0\}$
Une anomalie discrète est détectée si $\mathcal{M} = \emptyset$

2- Estimation parcimonieuse continue si $\mathcal{M} \neq \emptyset$
$\boldsymbol{\Phi}_{\mathcal{M}} = \{\boldsymbol{\Phi}_{\text{C},l} \mid l \in \mathcal{M}\}$
$\widehat{\boldsymbol{e}}_{\text{C}}, \widehat{\boldsymbol{x}}_{\text{C}}, \boldsymbol{z} = \text{ADMM}(\boldsymbol{y}_C, \boldsymbol{\Phi}_{\mathcal{M}})$
Une anomalie continue ou multivariée est détectée si $\|\widehat{\mathbf{e}}_{\text{C}}\|_2 > S_{\text{PFA}}$

Notons que, étant donnée la structure par groupe des signaux d'anomalies $\widehat{\mathbf{e}}_{\text{D}}$ et $\widehat{\mathbf{e}}_{\text{C}}$, la détection d'anomalies peut être découplée en K_{D} et K_{C} sous problèmes et la règle de décision peut être appliquée paramètre par paramètre à partir des signaux $\widehat{\boldsymbol{e}}_{\text{D},k}, k = 1, ..., K_{\text{D}}$ et $\widehat{\boldsymbol{e}}_{\text{C},k}, k = 1, ..., K_{\text{C}}$.

2.3.4 Option d'invariance par translation

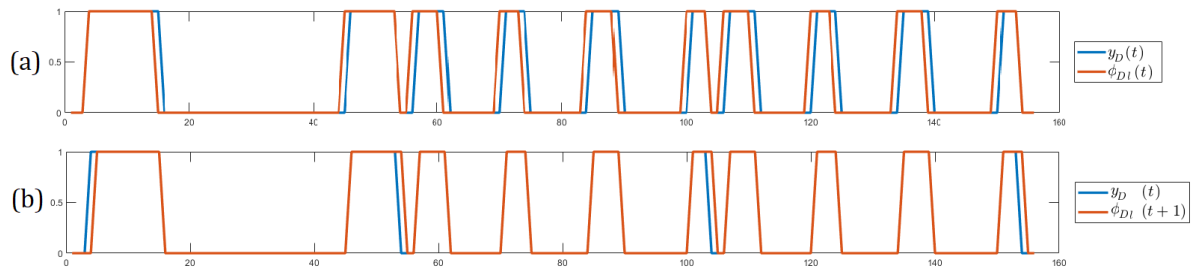

FIGURE 2.5 – (a) signal discret d'intérêt $\boldsymbol{y}_{\text{D}}$ et atome du dictionnaire $\boldsymbol{\phi}_{\text{D},l}$, (b) signal discret d'intérêt $\boldsymbol{y}_{\text{D}}$ et atome du dictionnaire $\boldsymbol{\phi}_{\text{D},l}$ décalé de 1.

L'algorithme ADDICT intègre une option d'invariance par translation permettant de régler un éventuel problème de décalage temporel entre le signal étudié et les atomes du dictionnaire. L'idée est de construire un dictionnaire redondant en appliquant des décalages à tous les atomes du dictionnaire. En d'autres termes, chaque atome $\boldsymbol{\phi}_l$ est décalé de τ instants pour créer un nouvel atome $\boldsymbol{\phi}_{l-\tau}$, avec $\tau \in \{-\tau_{\max}, -(\tau_{\max}-1), ..., -1, 0, 1, ..., \tau_{\max}-1, \tau_{\max}\}$. Le décalage maximal $\tau_{\max}$ est fixé par l'utilisateur. En activant l'option d'invariance par translation, la taille du dictionnaire augmente (le nombre d'atomes passe de L à $2L\tau_{\max}+L$) ainsi que la complexité de calcul. Pour cette dernière raison, l'option n'a été appliquée dans ces travaux qu'aux données discrètes. L'intérêt de cette option est illustré par la figure 2.5 qui représente un signal discret (signal bleu) et un atome du dictionnaire discret (signal orange) lorsque l'option n'est pas activée (a) et lorsque qu'elle est activée et que l'atome est décalé de $\tau_{\max} = 1$ (b). Dans le premier cas (a),il apparaît que la différence entre les deux signaux ($\|\boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D},l}$) est élevée et donc que l'atome $\boldsymbol{\phi}_{\mathrm{D},l}$ ne peut être sélectionné que pour une valeur suffisamment grande de l'hyperparamètre b_{D}. Or, comme nous avons pu l'évoquer plus haut, le choix d'une trop grande valeur pour cet hyperparamètre peut compromettre la détection des anomalies. En recanche, en activant l'option d'invariance par translation et en décalant l'atome $\boldsymbol{\phi}_{\mathrm{D},l}$ de 1 (b), les deux signaux $\boldsymbol{y}_{\mathrm{D}}$ et $\boldsymbol{\phi}_{\mathrm{D},l}$ sont presque égaux et l'atome discret pourra être sélectionné même pour une faible valeur de b_{D}. Cette option est donc intéressante car elle assure une sélection plus exhaustive d'atomes et permet ainsi d'améliorer l'estimation des signaux et la détection d'anomalies.

2.3.5 Résultats expérimentaux

2.3.5.1 Apprentissage de dictionnaire

Il est évident que la qualité de l'estimation des données et par conséquent les performances de détection d'anomalies de l'algorithme ADDICT reposent en grande partie sur le dictionnaire. Malheureusement, l'apprentissage de dictionnaires de données mixtes est un problème qui a été encore peu exploré et les méthodes classiques d'apprentissage de dictionnaires tels que les algorithmes K-SVD ou ODL (voir section 2.2.7 et annexe A.1 et A.2) ne permettent pas d'apprendre des atomes discrets, ou des atomes ayant certaines de leurs composantes discrètes. L'apprentissage de dictionnaire à partir de données mixtes et un sujet très intéressant qui mériterait d'être étudié dans le futur. Dans ces travaux, les premières expérimentations ont été réalisées à partir d'un dictionnaire composé de L atomes sélectionnés parmi les signaux sains de l'ensemble d'apprentissage. Dans un premier temps, le dictionnaire est initialisé en sélectionnant de manière aléatoire L signaux d'apprentissage. Les signaux continus issus des signaux mixtes non sélectionnés sont ensuite décomposés sur le dictionnaire continu initialisé en résolvant le problème LASSO (2.9) à l'aide de l'algorithme ADMM (voir section 2.2.2.2). Les L signaux dont l'erreur d'estimation est la plus élevée sont sélectionnés. Ce processus est répété 100 fois et les L signaux qui ont été les plus souvent sélectionnés constituent les atomes du dictionnaire mixte noté Φ. Cette stratégie de construction du dictionnaire est inspirée de l'implémentation du célèbre algorithme K-SVD pour lequel les auteurs préconisent par exemple d'éliminer à chaque itération les atomes trop corrélés ou les atomes les moins

utilisés et de les remplacer par les signaux d'apprentissage les moins bien représentés. Ces astuces permettraient souvent d'éviter les minimas locaux et le sur-apprentissage.

2.3.5.2 Données de télémesure satellite réelles

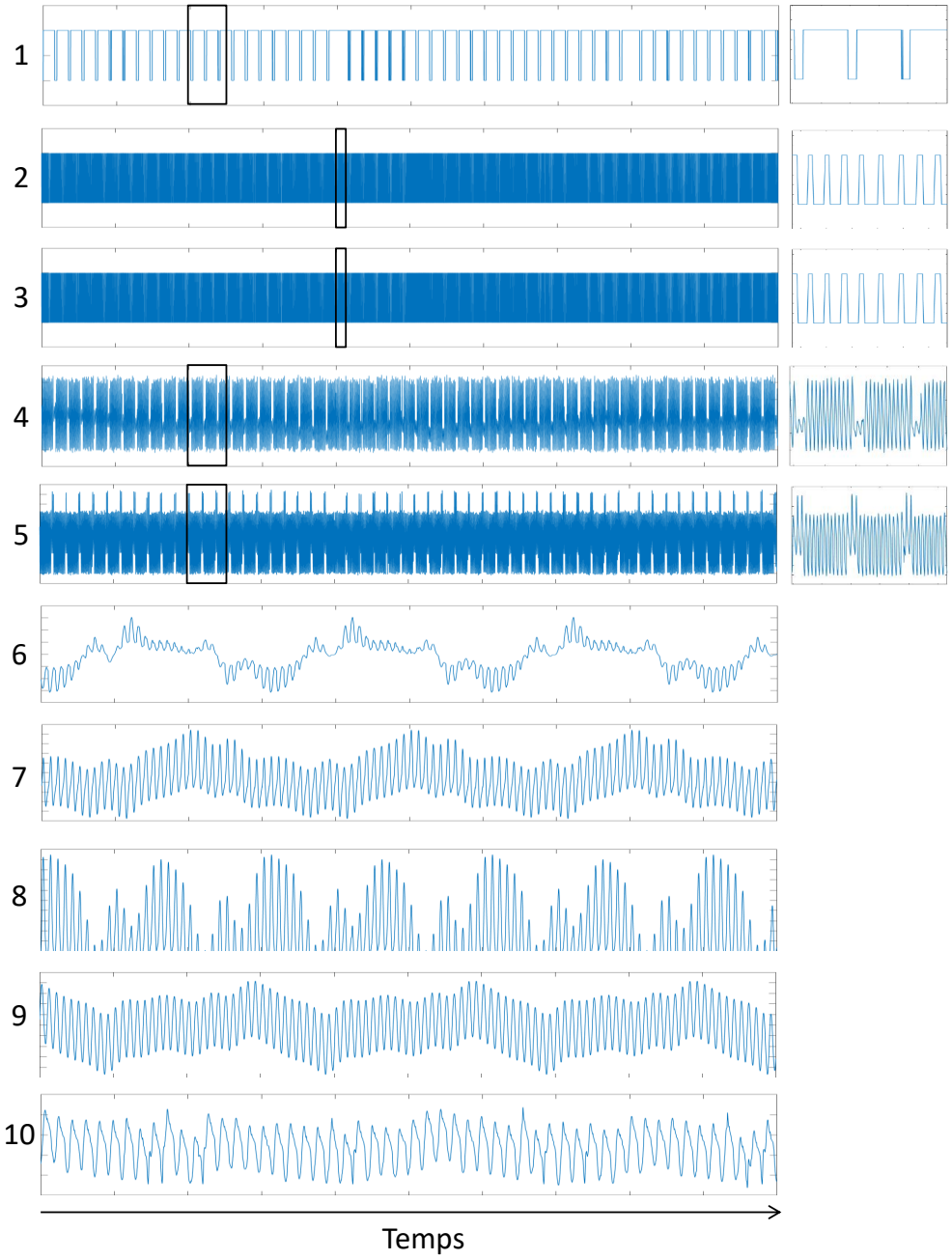

FIGURE 2.6 – Extraits des signaux d'apprentissage issus de 10 paramètres de télémesure dont 3 paramètres discrets (1 à 3) et 7 paramètres continus (4 à 10).

Afin d'évaluer les performances de l'algorithme proposé, des tests ont pu être menés sur de données réelles de télémesure satellite. On considère dans ces expérimentations $K = 10$ paramètres de télémesure dont $K_{\mathrm{D}} = 3$ paramètres discrets décrivant le mode de fonctionnement (ON/OFF) d'équipements et $K_{\mathrm{C}} = 7$ paramètres continus. Le dictionnaire $\mathbf{\Phi}$ a été construit à partir de deux mois de télémesure nominale (sans anomalie), ce qui représente, après pré-traitement, environ 30000 signaux mixtes pour l'apprentissage. Le pré-traitement évoqué correspond au pré-traitement ADDICT (voir section 2.3.1) appliqué avec une taille de fenêtre fixée à $W = 50$ (la dimension des signaux mixtes est alors de $N = 500$) et un décalage de chevauchement fixé à $\delta = 5$. Pour une fréquence d'échantillonnage classique de 1/32Hz, cela signifie que le signal mixte correspond à la concaténation des 10 signaux acquis sur une même période d'environ 26 minutes (50*32 secondes). Un extrait de télémesure des différents paramètres étudiés associé à un fonctionnement normal du satellite est représenté dans la figure 2.6 dans laquelle les trois premiers signaux correspondent aux paramètres discrets et les 7 derniers aux paramètres continus. Pour la détection et l'évaluation des performances, une base d'anomalies a été construite. Il s'agit d'une base de test dont la vérité terrain est connue et qui se compose d'anomalies réelles ou fictives identifiées par des experts. Cette base compte 1000 signaux mixtes de test dont 90 sont affectés par une anomalie. Les signaux affectés peuvent être regroupés en 7 périodes d'anomalies représentées dans la figure 4.1 (cadres rouges). Chaque période d'anomalies peut contenir un ou plusieurs signaux mixtes anormaux contigus. Elles constituent la vérité terrain sur laquelle nous allons pouvoir nous appuyer pour évaluer les performances de l'algorithme. La base d'anomalies étudiée se compose d'anomalies très diverses telles que des anomalies locales (figure 4.1 #3 et #5), des anomalies collectives observées sur des paramètres discrets et continus (figure 4.1 #1, #2 et #4), des anomalies contextuelles univariées (figure 4.1 #6) ou encore des anomalies multivariées (figure 4.1 #4 et #7). La première anomalie multivariée étudiée ici correspond à une désynchronisation du fonctionnement de deux paramètres discrets décrivant le mode de fonctionnement (ON/OFF) de deux équipements. Cette anomalie peut également être vue comme une anomalie collective si l'on considère uniquement la durée anomale en position "ON" de l'équipement dont le fonctionnement est décrit par le signal du haut dans la figure 4.1 #4. La deuxième anomalie multivariée implique un paramètre continu et un paramètre discret. En fonctionnement normal (à l'extérieur du cadre rouge), on constate que la variance des données acquises pour le paramètre continu chute lorsque que l'équipement est actif. Ce phénomène n'est pas observé sur la période encadrée en rouge ce qui représente une anomalie multivariée.

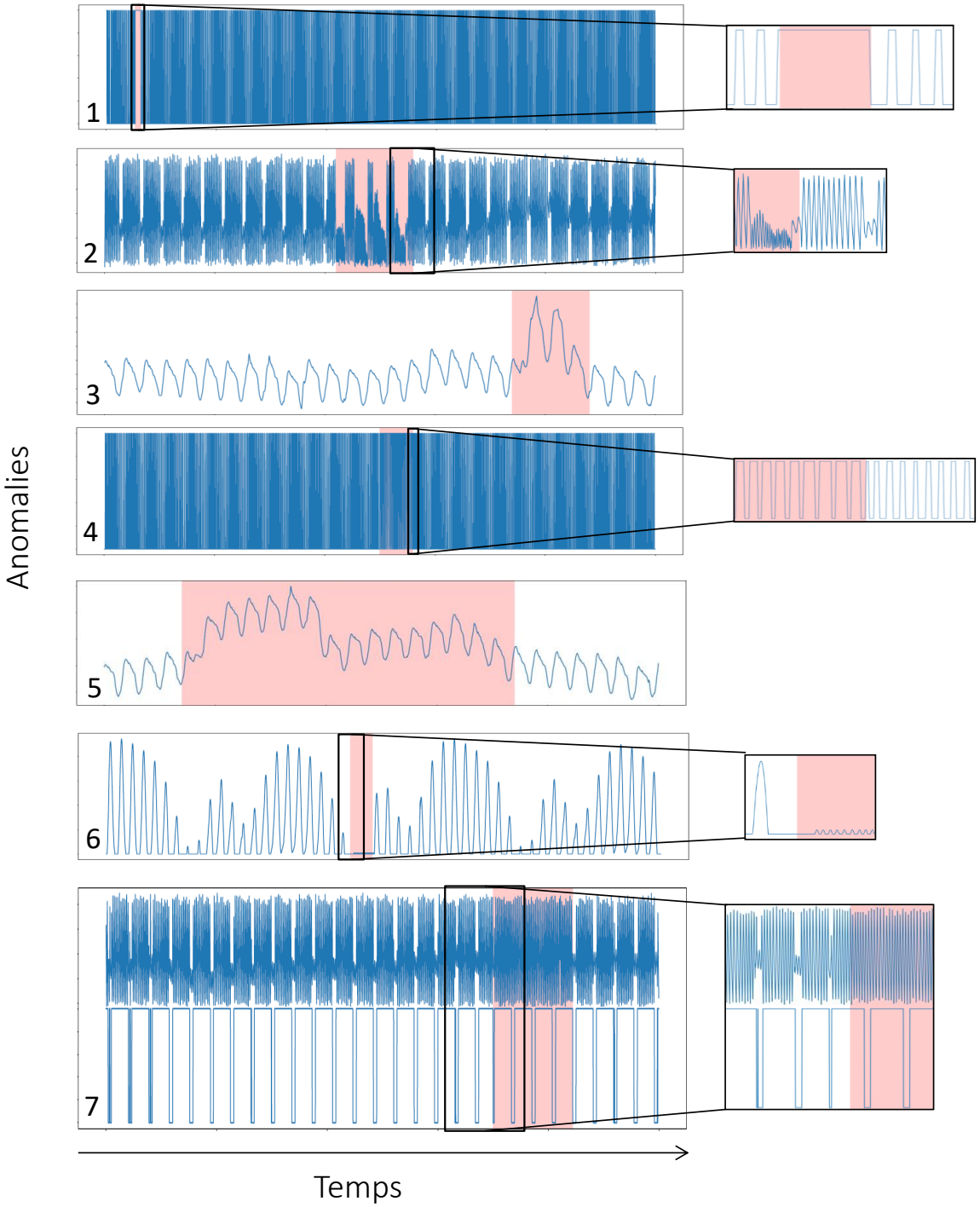

FIGURE 2.7 – Périodes d'anomalies de la vérité terrain composée de 7 périodes d'anomalie diverses et de durée variable. La vérité tarrain compte deux anomalies locales (#3 et #5), trois anomalies collectives (#1 , #2 et #4), une anomalie contextuelle (#6) et une anomalies multivariée entre impliquant un paramètre discret et un paramètre continus (#7)

2.3.5.3 Performances

Les performances de détection de l'algorithme ADDICT enregistrées sur nos données de télémesure satellite ont été comparées à celles de trois autres méthodes de l'état de l'art, à savoir l'algorithme NOSTRADAMUS Fuertes et al. (2016), l'algorithme MPPCAD Yairi et al. (2017) et l'algorithme OC-SVM Schölkopf et al. (2001). Ce dernier a été appliqué à des signaux mixtes obtenus à l'aide du pré-traitement ADDICT (voir section 2.3.1). Ces trois approches ont été présentées en détail dans la section 1.4. Les scores d'anomalies retournés par chacune d'elles pour les différents signaux de la base d'anomalies sont représentés dans la figure 2.8. Notons que les scores retournés par l'algorithme ADDICT (figure 2.8(d)) ont été obtenus avec l'option d'invariance par translation inactive (i.e., $\tau_{\max} = 0$). Les périodes d'anomalies connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1.
L'analyse des scores de la figure 2.8 montre que les anomalies locales #3 et #5 sont bien détectées par toutes les méthodes. En effet, les scores attribués aux signaux affectés par ces anomalies sont significativement plus élevés que la moyenne. La deuxième période d'anomalie correspondant à une anomalie collective est également bien détectée par toutes les méthodes. En revanche, la première anomalie collective qui affecte un paramètre discret n'est détectée que par NOSTRADAMUS. C'est également le cas de l'anomalie #4 qui affecte deux paramètres discrets. Cette non détection par l'algorithme MPPCAD s'explique en partie par le fait que cette approche traite des signaux multivariés composés d'une seule mesure par paramètre (i.e., la taille des fenêtres temporelles étudiées est égale à $W = 1$) alors que ce type d'anomalie nécessite de prendre en compte des collections de mesures (voir section 1.4.4.1 pour plus de d'étails sur le méthode MPPCAD).
D'une manière générale, il apparaît que les anomalies observées sur les paramètres discrets sont moins bien détectées par les méthodes de détection d'anomalies multivariées et l'algorithme ADDICT ne fait pas exception. Cependant, rappelons que dans ces premiers tests l'option d'invariance par translation de l'algorithme ADDICT n'a pas été activée. La période d'anomalie #6 est une anomalie contextuelle univariée qui touche un paramètre continu. Elle n'est détectée que par l'algorithme ADDICT. La non-détection de cette anomalie par la méthode MPPCAD s'explique par les mêmes arguments que ceux avancés pour la non détection de l'anomalie collective #4. Par ailleurs, cette anomalie n'est pas détectée par NOSTRADAMUS car elle n'affecte pas de manière significative les statistiques qui composent le vecteur d'entrée de cet algorithme (voir le pré-traitement NOSTRADAMUS détaillé à la section 1.4.3.3).
L'anomalie #7 correspondant à une anomalie multivariée est bien détectée par l'algorithme ADDICT et par l'algorithme OC-SVM. La non détection de cette anomalie par l'algorithme MPPCAD est moins nette. Enfin, la non-détection de cette anomalie par NOSTRADAMUS était prévisible car cette méthode est univariée et traite les paramètres indépendamment les uns des autres. Cette approche ne permet donc pas de prendre en compte les relations entres les paramètres et d'y détecter tout changement.

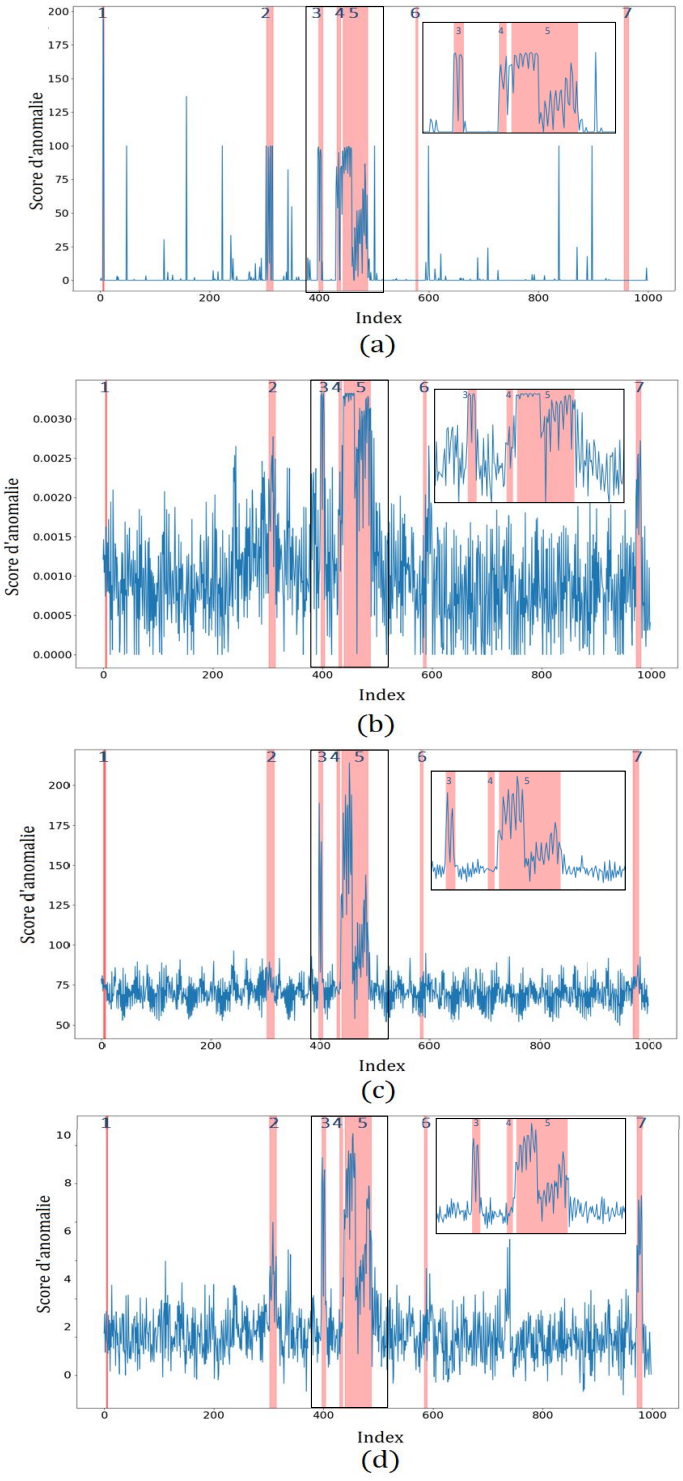

FIGURE 2.8 – Comparaison de scores d'anomalie retournés pour les différents signaux tests de la base d'anomalies par les algorithmes NOSTRADAMUS, OC-SVM, MPCCAD et ADDICT pour lequel l'option d'invariance par translation n'a pas été activée. La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

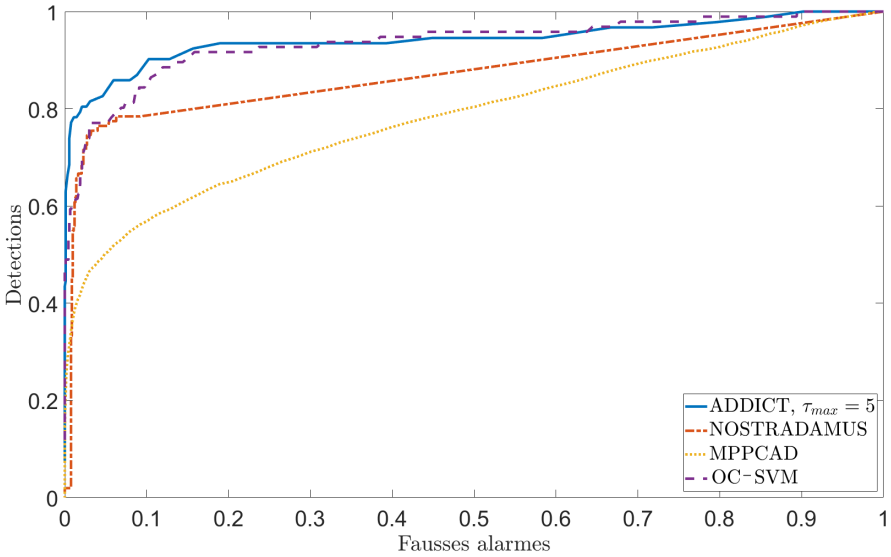

FIGURE 2.9 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par ADDICT (avec l'option d'invariance par translation active et $\tau_{\max} = 5$), NOSTRADAMUS, MPPCAD et OC-SVM.

TABLE 2.1 – 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) de l'algorithme OC-SVM, MPPCAD, NOSTRADAMUS et ADDICT avec l'option d'invariance par translation inactive ($\tau_{\max} = 0$) et active (avec $\tau_{\max} = 5$). Le seuil S_{PFA} correspond au seuil pour lequel le couple (P_{FA},P_{D}) est le plus proche du point optimal (0,1).

Méthode	S_{PFA}	P_{D}	P_{FA}	AUC
OC-SVM	0.0016	89%	12.3%	0.94113
MPPCAD	12	80%	13%	0.8779
NOSTRADAMUS	29	79.6%	6%	0.88
ADDICT ($\tau_{\max} = 0$)	3.8	84.6%	9.8%	0.9246
ADDICT ($\tau_{\max} = 5$)	3.7	89%	10.2%	0.937

Les performances de détection en termes de probabilité de détection P_{D} et de probabilité de fausse alarme P_{FA} calculées pour les différentes méthodes évaluées sur notre base d'anomalies sont représentées sous forme de courbes CORs dans la figure 2.9. La courbe COR d'un détecteur représente les probabilités de détections (en ordonné) et les probabilités de fausses alarmes (en abscisses) estimées pour tous les seuils de détection. Elle ne peut être calculée que si une vérité terrain est disponible. Les résultats présentés dans la figure 2.9 décrivent les meilleures performances enregistrées sur nos données pour chacune des méthodes étudiées et sont associés à un réglage optimal de leurs hyperparamètres obtenus à l'aide de la vérité terrain. Notons qu'une méthode de réglage automatique des hyperparamètres ADDICT sera détaillée à la section 2.5. Quelques résultats quantitatifs, résumés dans la table 2.1 où le seuil S_{PFA} correspond au seuil pour lequel le couple (P_{FA},P_{D}) est le plus proche du point optimal (0,1), mènent aux conclusions suivantes :

- L'algorithme MPPCAD retourne peu de fausses alarmes mais ne détecte que les anomalies locales présentes dans la base d'anomalies étudiée. Il serait intéressant d'appliquer cette méthode à des fenêtres temporelles de plus grande taille (que $W = 1$) pour tenter de détecter les autres anomalies.
- L'algorithme NOSTRADAMUS détecte la majorité des anomalies univariées mais retourne de nombreuses fausses alarmes. Cette étude confirme le fait qu'un traitement indépendant des différents paramètres (comme c'est le cas de la méthode NOSTRADAMUS) ne permet pas de détecter les anomalies multivariées.
- L'algorithme OC-SVM appliqué à des signaux mixtes détecte la plupart des anomalies univariées qui affectent des paramètres continus et l'anomalie multivariée qui touche un paramètre discret et un paramètre continu. Cependant aucune des anomalies de notre base qui affecte uniquement des paramètres discrets n'est détectée par cette méthode.
- Les résultats obtenus avec l'algorithme ADDICT sont très encourageants notamment lorsque l'option d'invariance par translation est activée.

Comme nous avons pu le voir précédemment, l'algorithme de détection d'anomalies ADDICT repose sur l'étude des signaux d'anomalies discret $\widehat{\mathbf{e}}_{\mathrm{D}}$ et continu $\widehat{\mathbf{e}}_{\mathrm{C}}$. Comme évoqué à la section 2.3.1, la structure par groupe de ces signaux permet de découpler la détection en K_{D} et K_{C} détections et la règle de décision peut être appliquée paramètre par paramètre. La figure 2.10 représente les scores univariés, calculés pour chaque paramètre à partir des sous-signaux $\widehat{\boldsymbol{e}}_k, k = 1, ..., K$. Les périodes d'anomalie connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1. L'analyse de ces scores peut constituer un point de départ à l'investigation en cas de détection par l'algorithme ADDICT, notamment pour identifier les paramètres affectés par l'anomalie. En effet, au regard de la figure 2.10, il apparaît que les scores des différents paramètres sont bien supérieurs à la moyenne, ou négatifs, presque exclusivement lorsque qu'ils sont touchés par une anomalie. Par ailleurs, cette représentation met en évidence les relations, parfois méconnues, qui peuvent exister entre les paramètres. En effet, on observe que les scores attribués aux paramètre 2 et 3 sont très corrélés, de même que les scores attribués aux paramètres 4 et 5. La nature des relations qu'entretiennent ces paramètres est visible à la figure 2.6.

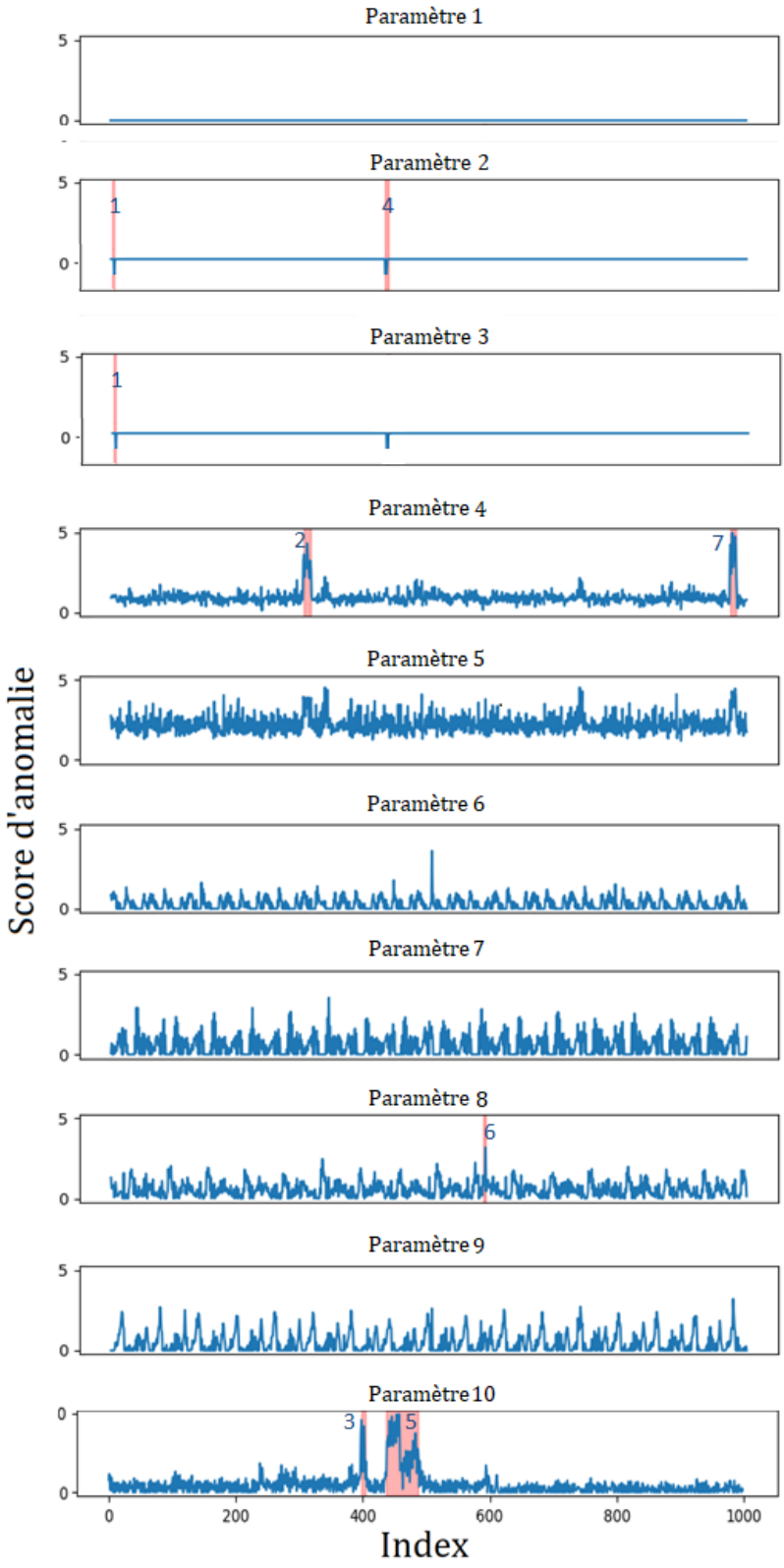

FIGURE 2.10 – Scores d'anomalie univariés retournés par l'algorithme ADDICT lorsque l'option d'invariance par translation est activée avec $\tau_{\max} = 5$.

2.3.5.4 Option d'invariance par translation

Cette section s'intéresse à l'option d'invariance par translation qui a été intégrée à l'algorithme ADDICT afin de corriger les éventuels décalages temporels entres les atomes du dictionnaire et les signaux tests. La figure 2.11 représente les scores retournés par l'algorithme ADDICT pour les signaux de la base d'anomalies lorsque l'option d'invariance par translation est activée avec une invariance maximale fixée à $\tau_{\max} = 5$. Les périodes d'anomalies connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1. On observe que toutes les anomalies de la base sont bien détectées notamment les anomalies #1 et #4 qui affectent des paramètres discrets. Ces anomalies ne sont pas détectées lorsque l'option n'est pas activée (voir figure 2.8 (d)). De plus, l'activation de cette option a permis de sélectionner un plus grand nombre d'atomes et donc d'améliorer l'estimation des signaux continus nominaux et de réduire le nombre de fausses alarmes. Ces résultats confirment l'intérêt de cette option pour notre problème. En revanche, la question se pose de régler l'hyperparamètre $\tau_{\max}$ qui fixe l'invariance maximale autorisée. La figure 2.12 représente les courbes CORs associées à l'algorithme ADDICT pour différentes valeurs de $\tau_{\max}$. Ces résultats montrent qu'une invariance maximale $\tau_{\max} = 5$ représente, pour ces données, un bon compromis en terme de performances de détection et de complexité de calcul (plus $\tau_{\max}$ est élevé, plus le temps d'exécution est important).

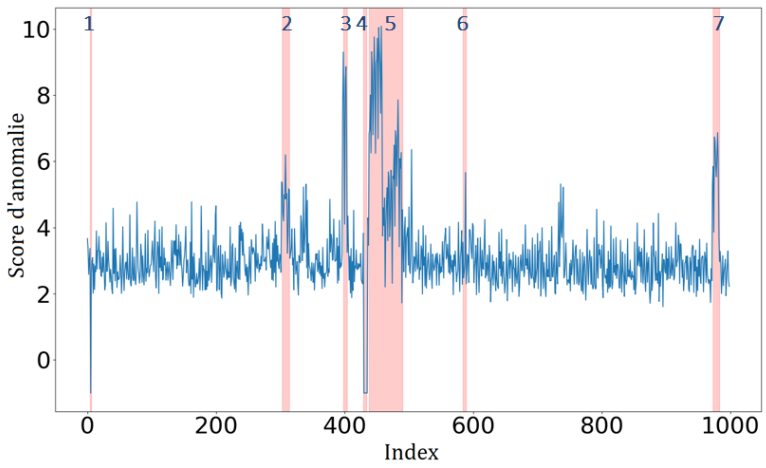

FIGURE 2.11 – Scores d'anomalie retournés par l'algorithme ADDICT lorsque l'option d'invariance par translation est active avec $\tau_{\max} = 5$. La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

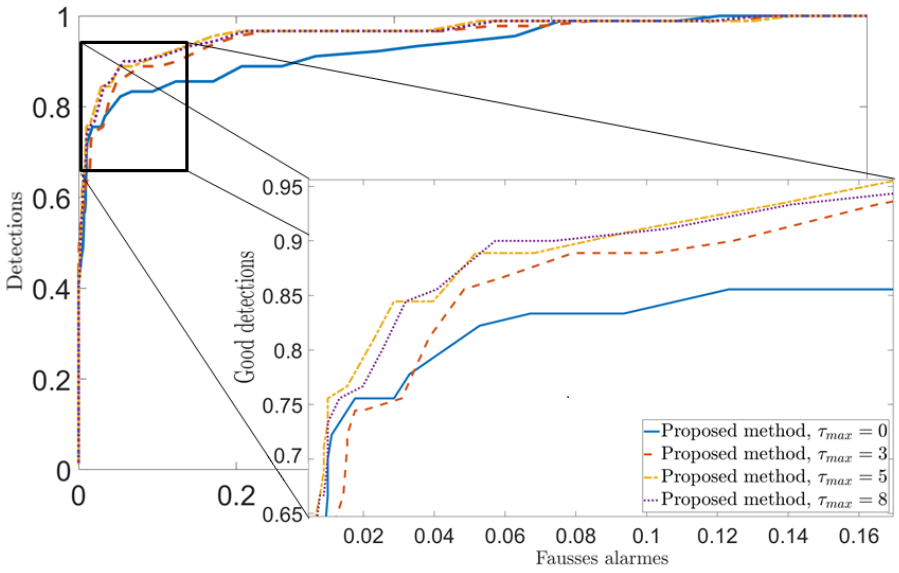

FIGURE 2.12 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par ADDICT pour plusieurs valeurs du paramètre $\tau_{\max}$.

2.3.5.5 Choix du nombre d'atomes

Tout comme la construction du dictionnaire, le choix du nombre d'atomes qui le compose est important car il influence les performances du détecteur. De manière intuitive, plus le nombre d'atomes dans le dictionnaire est élevé, meilleure est la représentation des signaux nominaux et plus faible est la probabilité de fausse alarme. Cependant, plus le nombre d'atomes du dictionnaire est élevé, plus la complexité de calcul est importante. Le choix du nombre d'atomes représente donc un compromis entre qualité d'estimation et complexité de calcul. La figure 2.13 représente les probabilités de détection P_{D} et les probabilités de fausses alarmes P_{FA} retournées par l'algorithme ADDICT pour les signaux de notre base d'anomalies en faisant varier le nombre d'atomes du dictionnaire. Pour chaque valeur du nombre d'atomes, le dictionnaire a été construit à partir de signaux sains (sans anomalie) selon la procédure détaillée à la section 2.3.5.1. Les couples (P_{FA},P_{D}) représentés ici pour chaque valeur du nombre d'atomes ont été déterminés à l'aide de la vérité terrain disponible. Dans un premier temps, les performances de détection s'améliorent à mesure que le nombre d'atomes du dictionnaire augmente. Par exemple, avec dictionnaire de 100 atomes il faut accepter un taux de fausses alarmes de 26% pour pouvoir détecter $77,78\%$ des anomalies alors que l'utilisation d'un dictionnaire de 2000 atomes permet d'atteindre une probabilité de détection P_{D} de 89% et un probabilité de fausse alarme P_{FA} de 10.2%. Au-delà de 2000 atomes les performances de détection ne s'améliorent pas de manière significative. Ces résultats justifient notre choix de construire un dictionnaire composé de 2000 atomes.

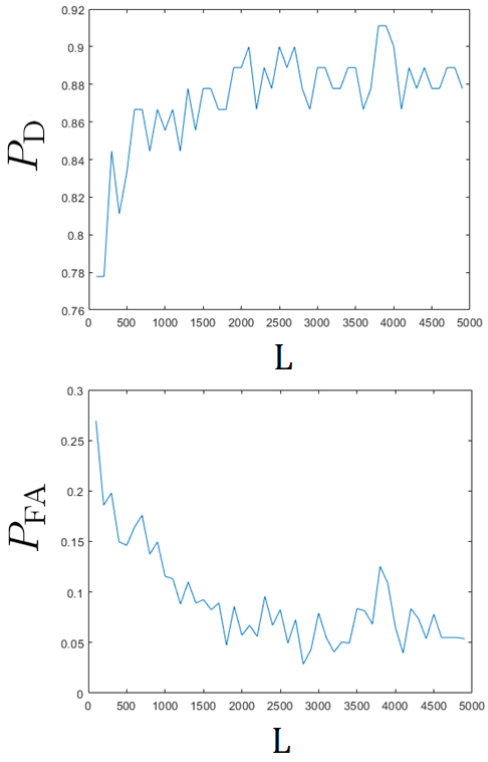

FIGURE 2.13 – Comparaison des performances de détection ADDICT en termes de probabilités de détection P_{D} et de probabilités de fausses alarmes P_{FA} pour différentes valeurs de L, le nombre d'atomes du dictionnaire.

2.3.5.6 Normalisation des données

Lorsque l'on traite des données multivariées, il est naturel de normaliser les données afin de les convertir vers un format standard commun et de limiter l'effet d'échelle lié à leur hétérogénéité. L'objectif de cette section est d'étudier l'influence de la normalisation sur les performances de détection de l'algorithme ADDICT. Deux techniques ont été explorées : la normalisation par centrage-réduction (aussi appelé standardisation) et la normalisation Min-Max. Ces deux normalisations peuvent être définies comme suit

$$\mathbf{Y'}_k = \frac{\mathbf{Y}_k - \mathbf{Y}_{\min,k}}{\mathbf{Y}_{\max,k} - \mathbf{Y}_{min,k}} \tag{2.43}$$

Normalisation Min-Max

$$\mathbf{Y'}_k = \frac{\mathbf{Y}_k - \mathrm{mean}(\mathbf{Y}_k)}{\mathrm{std}(\mathbf{Y}_k)} \tag{2.44}$$

Standardisation

où $\mathbf{Y}_k \in \mathbb{R}^{W \times M}$ est une matrice composée de signaux univariés associés au paramètre $k, k = 1, ..., K$ et extraits d'une matrice $\mathbf{Y} \in \mathbb{R}^{N \times M}$ de signaux multivariés. La matrice

$\mathbf{Y'}_k \in \mathbb{R}^{W \times M}$ est la matrice de signaux univariés associée au paramètre k obtenue par normalisation de la matrice $\mathbf{Y}_k$. Les paramètres $\mathbf{Y}_{\max,k}$ et $\mathbf{Y}_{\min,k}$ correspondent aux valeurs du maximum et du minimum de $\mathbf{Y}_k$ et $\text{mean}(\mathbf{Y}_k)$ et $\text{std}(\mathbf{Y}_k)$ renvoient respectivement la moyenne et l'écart-type calculés sur les données de $\mathbf{Y}_k$. Dans cette étude, les données d'apprentissage et les données de test ont été normalisées avec les mêmes valeurs de $\mathbf{Y}_{\min,k}, \mathbf{Y}_{\max,k}, \text{mean}(\mathbf{Y}_k)$ et $\text{std}(\mathbf{Y}_k)$ calculées pour chaque paramètre k sur les données d'apprentissage associées à un fonctionnement normal du satellite. Les performances de détection de l'algorithme ADDICT appliqué à des données normalisées selon ces deux techniques et non normalisées sont représentées dans la figure 2.14 sous forme de courbes CORs. Dans les trois cas, les courbes CORs sont associées à un réglage optimal des hyperparamètres réalisé à l'aide de la vérité terrain. Pour nos données, la standardisation a détérioré les performances de détection. En effet, pour atteindre une probabilité de bonne détection de 90%, il faut accepter 35.75% de fausses alarmes avec les données standardisées contre 6.5% avec les données normalisées selon la normalisation Min-Max et 10% avec les données non normalisées. La normalisation des données à l'aide d'une normalisation Min-Max ou par standardisation n'a pas permis d'améliorer de manière significative les performances de détection de l'algorithme ADDICT. De plus, ces résultats sont difficiles à interpréter. Dans la suite de cette étude, nous proposons de ne pas utiliser de normalisation.

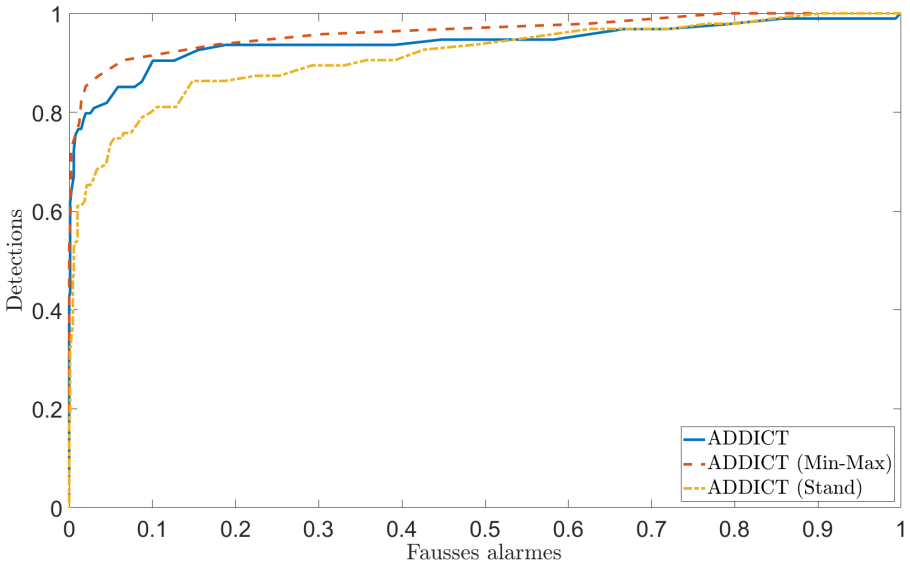

FIGURE 2.14 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par ADDICT appliqué à des données normalisées selon une normalisation par centrage-réduction (Stand) et une normalisation Min-Max (Min-Max), et à des données non normalisées.

2.4 Les extensions ADDICT

2.4.1 L'algorithme G-ADDICT

Cette section propose de généraliser la méthode d'estimation parcimonieuse ADDICT afin l'adapter d'avantage au cadre multivarié et de réaliser une détection d'anomalies efficace malgré l'hétérogénéité des données étudiées. Pour cela nous définissons les problèmes d'estimation parcimonieuse ADDICT généralisés comme suit

$$\widehat{\boldsymbol{x}}_{\mathrm{D}}, \widehat{\boldsymbol{e}}_{\mathrm{D}} = \arg \min_{\boldsymbol{x}_{\mathrm{D}} \in \mathcal{B}, \boldsymbol{e}_{\mathrm{D}} \in \mathbb{R}^{N_{\mathrm{D}}}} \| \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\Phi}_{\mathrm{D}} \boldsymbol{x}_{\mathrm{D}} - \boldsymbol{e}_{\mathrm{D}} \|_2^2 + \sum_{k=1}^{K_{\mathrm{D}}} b_{\mathrm{D},k} \|\mathbf{e}_{\mathrm{D,k}}\|_2 \tag{2.45}$$

Estimation parcimonieuse discrète généralisée

$$\widehat{\boldsymbol{x}}_{\mathrm{C}}, \widehat{\boldsymbol{e}}_{\mathrm{C}} = \arg \min_{\boldsymbol{x}_{\mathrm{C}}, \mathbf{e}_{\mathrm{C}}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathcal{M}} \boldsymbol{x}_{\mathrm{C}} - \mathbf{e}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\boldsymbol{x}_{\mathrm{C}}\|_1 + \sum_{k=1}^{K_{\mathrm{C}}} b_{\mathrm{C},k} \|\mathbf{e}_{\mathrm{C,k}}\|_2 \tag{2.46}$$

Estimation parcimonieuse continue généralisée

La stratégie proposée ici consiste à estimer les signaux d'anomalies $\boldsymbol{e}_{\mathrm{D}}$ et $\boldsymbol{e}_{\mathrm{C}}$ non plus à l'aide d'un unique hyperparamètre par type données (b_{D} et b_{C}) comme c'est le cas dans la version de base de l'algorithme ADDICT, mais à l'aide d'autant d'hyperparamètres qu'il y a de paramètres de télémesure. La résolution des problèmes (2.45) et (2.46) par minimisation du Lagrangien augmenté à l'aide de l'algorithme ADMM est équivalente à celles des problèmes non généralisés (2.28) et (2.29) détaillées à la section 2.3.2. Par exemple, l'estimation à l'itération $i+1$ du $k^{\text{ème}}$ signal d'anomalie continu est donnée par l'opérateur de seuillage par groupe (voir section 2.3.3.1) tel que $\widehat{\mathbf{e}}_{\mathrm{C},k}^{i+1} = T_{b_{\mathrm{C},k}}(\mathbf{h}_{\mathrm{C},k}^{i+1})$, avec $\mathbf{h}_{\mathrm{C},k}^{i+1}$ le $k^{\text{ème}}$ sous-signal de $\mathbf{h}_{\mathrm{C}}^{i+1} = \boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathrm{C}} \widehat{\boldsymbol{x}}_{\mathrm{C}}^{i+1}$. De même, l'estimation du signal d'anomalie discret est donnée par l'opérateur de seuillage par groupe tel que $\widehat{\mathbf{e}}_{\mathrm{D},k} = T_{b_{\mathrm{D},k}}(\mathbf{h}_{\mathrm{D},k})$, avec $\mathbf{h}_{\mathrm{C},k}^{i+1}$ le $k^{\text{ème}}$ sous-signal de $\mathbf{h}_{\mathrm{C}}^{i+1} = \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D}} \widehat{\boldsymbol{x}}_{\mathrm{D}}$. La figure 2.15 représente le score retourné par l'algorithme ADDICT généralisé pour tous les signaux multivariés qui composent notre base d'anomalies. Les périodes d'anomalies connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1. Tous les hyperparamètres ont été déterminés à l'aide de la vérité terrain disponible. On observe que les scores obtenus sont non nuls presque exclusivement en présence d'anomalies ce qui témoigne d'une détection d'anomalies efficace. Par ailleurs, le seuil de détection S_{PFA} est plus facile à déterminer que pour le modèle non généralisé. L'examen des scores univariés (calculés pour chaque paramètre à partir des sous-signaux d'anomalie $\boldsymbol{e}_k$) représentés dans la figure 2.17 mène aux mêmes conclusions. Pour chaque période d'anomalie, l'identification des paramètres de télémesure affectés par l'anomalie est immédiate. Une comparaison des performances de l'algorithme ADDICT et de sa version généralisée en termes de probabilités de détection et de probabilités de fausses alarmes peut être réalisée à travers l'étude des courbes CORs représentées dans la figure 2.16. Ces résultats sont résumés dans la table 2.2 et démontrent de l'intérêt de la généralisation du problème ADDICT pour ces données. En effet, le modèle

ADDICT généralisé détecte 94.7% des anomalies et ne retourne que 4.6% de fausses alarmes contre 89% de détections et 10.2% de fausses alarmes retournées par l'algorithme ADDICT. La généralisation du problème proposée a permis d'améliorer les performances de détection de l'algorithme ADDICT grâce à une parcimonie plus adaptée au cadre multivarié du fait de l'introduction d'hyperparamètres bien choisis. La détection d'anomalies est moins soumise à l'effet d'échelle lié à l'hétérogénéité des données traitées, ce qui engendre des résultats de détection intéressants. Cependant, le modèle généralisé impose de régler un grand nombre d'hyperparamètres. Selon le nombre de paramètres de télémesure étudiés et la disponibilité d'une vérité terrain, le réglage de ces hyperparamètres peut s'avérer très difficile. Plusieurs méthodes de réglage de ces hyperparamètres seront présentées et discutées dans la section 2.4 dédiée à ce problème.

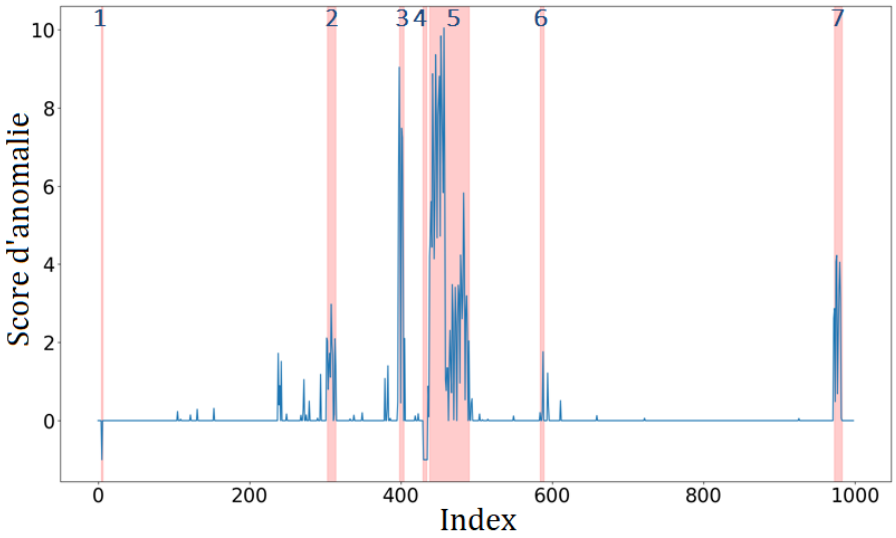

FIGURE 2.15 – Scores d'anomalie retournés par l'algorithme ADDICT généralisé avec l'option d'invariance par translation active et $\tau_{\max} = 5$. La vérité terrain est représentée par les fonds rouges.Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1

TABLE 2.2 – Résultats de détection en termes de probabilités de détection (P_{D}), de probabilités de fausse alarme (P_{FA}) et d'aire sous la courbe COR (AUC) de l'algorithme ADDICT et de sa version généralisée.

Méthode	S_{PFA}	P_{D}	P_{FA}	AUC
ADDICT	3.7	89%	10.2%	0.937
G-ADDICT	$\epsilon > 0$	93.6%	3.3%	0.9775

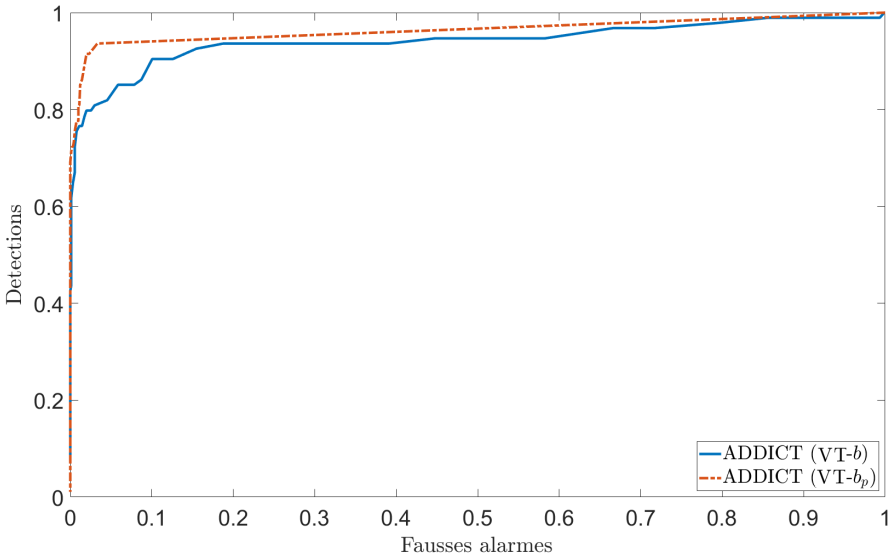

FIGURE 2.16 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par les algorithmes ADDICT et G-ADDICT.

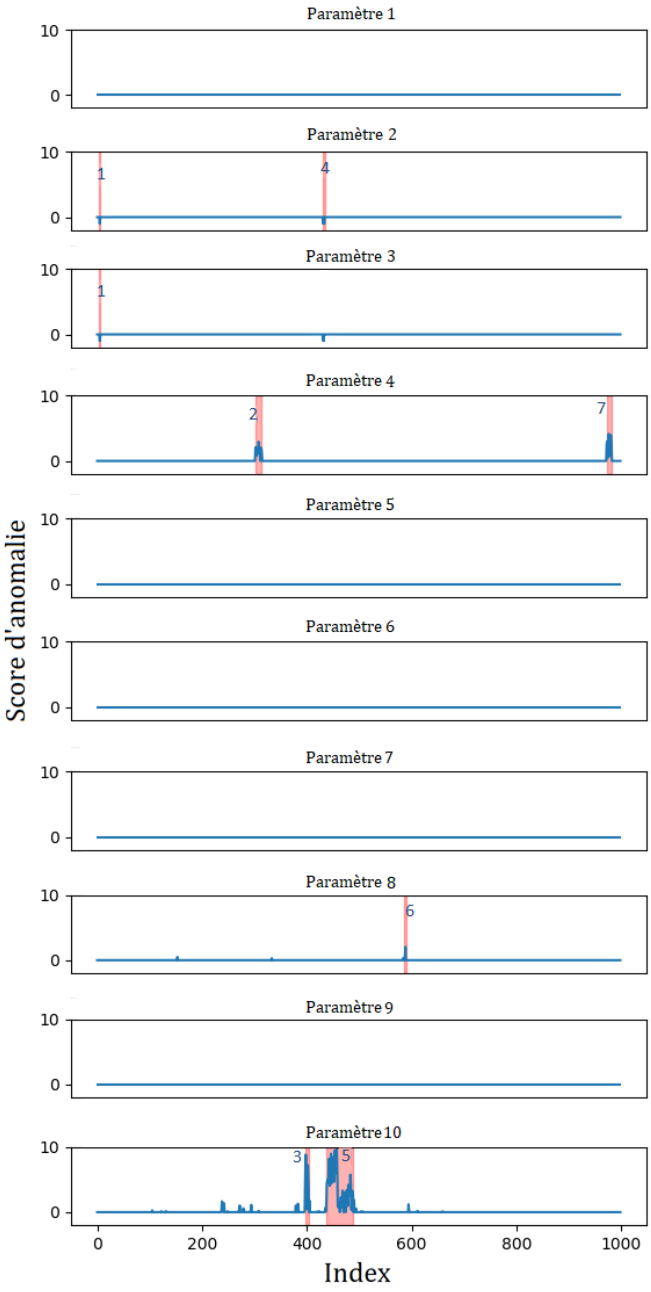

FIGURE 2.17 – Scores d'anomalies paramètre par paramètre retournés pas l'algorithme ADDICT généralisé avec l'option d'invariance par translation active. La vérité terrain est représentée par les fonds rouges.Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

2.4.2 L'algorithme W-ADDICT

Inspiré des algorithmes Reweighted-ℓ_1 et Group-LASSO, l'algorithme Weighted-ADDICT (W-ADDICT) est une extension pondérée de l'algorithme ADDICT. Il a l'avantage de per-

mettre l'intégration d'informations externes par l'intermédiaire d'une pondération appropriée permettant d'améliorer les performances de détection obtenues avec ADDICT. L'intérêt d'une telle extension a été étudié dans ces travaux pour l'estimation parcimonieuse de signaux continus à travers un exemple simple d'information externe issue du coefficient de corrélation entre le signal testé et sa décomposition sur le dictionnaire. Ces travaux ont fait l'objet de deux communications Pilastre et al. (2019b,c).

2.4.2.1 Formulation du problème

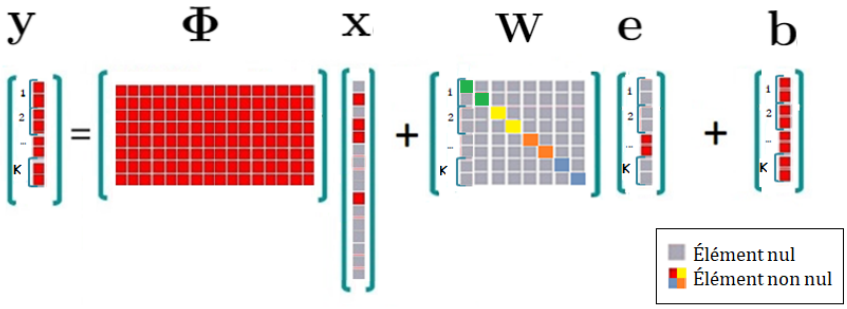

FIGURE 2.18 – Illustration de la stratégie de décomposition W-ADDICT.

La méthode de détection d'anomalies ADDICT pour un signal de données continues multivariées $\boldsymbol{y} \in \mathbb{R}^N$ repose sur l'erreur d'estimation issue de sa décomposition sur un dictionnaire. L'algorithme W-ADDICT propose d'améliorer la détection en intégrant de l'information externe issue de la connaissance métier ou d'autres indicateurs de présence d'anomalies que l'erreur d'estimation. L'objectif principal de cette extension est de limiter les fausses alarmes retournées par l'algorithme ADDICT. Les fausses alarmes observées sur nos données de test sont exclusivement associées à des paramètres continus. Pour cette raison, l'algorithme W-ADDICT présenté ici traite uniquement des données continus. Cependant, une telle extension peut également être appliquée à des données discrètes. Le modèle de représentation parcimonieuse pondérée proposé ici pour la détection d'anomalies multivariées attribue un poids par groupe, en considérant ici qu'un groupe réunit l'ensemble des mesures associées à un paramètre de télémesure donné. Les poids seront actualisés à chaque itération comme dans l'algorithme Reweighted-ℓ_1 (voir section 2.2.1.4). Le problème que nous proposons de résoudre s'écrit alors

$$\arg\min_{\boldsymbol{x},\boldsymbol{e}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} - \mathbf{e}\|_2^2 + a\|\boldsymbol{x}\|_1 + b\sum_{k=1}^{\mathrm{K}} w_k\|\mathbf{e}_k\|_2 \tag{2.47}$$

où $\boldsymbol{y} \in \mathbb{R}^N$ est un signal multivarié de taille N composé des dernières mesures de K paramètres continus de télémesure satellite, $\boldsymbol{\Phi} \in \mathbb{R}^{N\times L}$ est un dictionnaire connu composé de L atomes (formant les colonnes de $\boldsymbol{\Phi}$), $\boldsymbol{x} \in \mathbb{R}^L$ 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 bruit additif. Étant donné le cadre multivarié dans lequel nous nous plaçons et le pré-traitement appliqué (voir section 2.3.1.1), le signal test $\boldsymbol{y}$, ainsi que le signal d'anomalie $\mathbf{e}$ peuvent être divisés en K sous-signaux, où K correspond au

nombre de paramètres étudiés, i.e., $\boldsymbol{y} = [\boldsymbol{y}_1^T, ..., \boldsymbol{y}_K^T]^T$ et $\mathbf{e} = [\mathbf{e}_1^T, ..., \mathbf{e}_K^T]^T, k = 1, ..., K$. La décomposition W-ADDICT associée au problème (2.47) est présentée de façon schématique dans la figure 2.18. Le problème (2.47) peut être résolu par l'algorithme ADMM après avoir introduit une variable auxiliaire notée $\boldsymbol{z}$ et une contrainte d'égalité telles que

$$\arg\min_{\boldsymbol{x},\mathbf{e},\boldsymbol{z}} \frac{1}{2} \parallel \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} - \mathbf{e} \parallel_2^2 + a \parallel \boldsymbol{z} \parallel_1 + b \sum_{k=1}^{K} w_k \parallel \mathbf{e}_k \parallel_2 \text{ s.c. } \boldsymbol{z} = \boldsymbol{x}. \tag{2.48}$$

Le lagrangien augmenté associé au problème (2.48) est défini par

$$\mathcal{L}_A(\boldsymbol{x}, \boldsymbol{z}, \mathbf{e}, \mathbf{m}, \mu) = \frac{1}{2} \parallel \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} - \mathbf{e} \parallel_2^2 + a \parallel \boldsymbol{z} \parallel_1 +$$

$$b \sum_{k=1}^{K} w_k \parallel \mathbf{e}_k \parallel_2 + \mathbf{m}^T(\boldsymbol{z} - \boldsymbol{x}) + \frac{\mu}{2} \parallel \boldsymbol{z} - \boldsymbol{x} \parallel_2^2 . \tag{2.49}$$

La minimisation du Lagrangien augmenté pour la mise à jour des variables $\boldsymbol{x}$ et $\boldsymbol{z}$ mène aux mêmes expressions que celles obtenues pour les mises à jour des variables $\boldsymbol{x}_{\mathrm{C}}$ et $\boldsymbol{z}$ de l'algorithme ADDICT (voir section 2.3.1.3). Les poids $w_k, k = 1, ..., K$ interviennent dans la mise à jour du signal d'anomalie $\boldsymbol{e}$. Son expression à l'itération i+1 est obtenue par la solution du problème suivant

$$\widehat{\mathbf{e}}^{i+1} = \arg\min_{\mathbf{e}} \frac{1}{2} \parallel \mathbf{h}^{i+1} - \mathbf{e} \parallel_2^2 + b \sum_{k=1}^{K} w_k \parallel \mathbf{e}_k \parallel_2 \tag{2.50}$$

où $\mathbf{h}^{i+1} = \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^{i+1}$. Le problème (2.50) peut être découplé en K sous-problèmes qui estiment les sous-vecteurs $\mathbf{e}_k^{i+1}$ séparément

$$\widehat{\mathbf{e}}_k^{i+1} = \arg\min_{\mathbf{e}_k} \frac{1}{2} \parallel \mathbf{h}_k^{i+1} - \mathbf{e}_k \parallel_2^2 + b w_k \parallel \mathbf{e}_k \parallel_2 . \tag{2.51}$$

De manière analogue à l'algorithme non pondéré, la solution du problème (2.51) est donnée par l'opérateur de seuillage par groupe $\widehat{\mathbf{e}}_{\mathrm{k}}^{i+1} = T_{bw_k}(\mathbf{h}_k^{i+1})$, avec $\mathbf{h}_k^{i+1}$ le $k^{ème}$ sous-signal de $\mathbf{h}^{i+1}$.

$$T_b(\mathbf{h}_k^{i+1}) = \begin{cases} \frac{\|\mathbf{h}_k^{i+1}\|_2^2 - bw_k}{\|\mathbf{h}_k^{i+1}\|_2^2} \mathbf{h}_k^{i+1} & \text{si } \parallel \mathbf{h}_k^{i+1} \parallel_2^2 > bw_k \\ 0 & \text{sinon.} \end{cases} \tag{2.52}$$

L'opérateur T_{bw_k} fonde l'estimation des signaux d'anomalies $\mathbf{e}_k$ sur l'erreur d'estimation donnée par $\|\mathbf{h}_k\|_2$ et fait intervenir les poids de manière à favoriser la détection lorsque w_k est inférieur à 1 et inversement à limiter la détection lorsque w_k est supérieur à 1 (quand w_k est supérieur à 1, la première inégalité est plus difficile à respecter d'où un signal d'anomalie $\widehat{\mathbf{e}}_k$ plus souvent égal à 0).

2.4.2.2 Construction des poids

L'étude des résultats retournés par l'algorithme ADDICT sur les données de télémesure satellite présentées à la section 2.3.5 a permis d'examiner les erreurs commises, à savoir les anomalies de la vérité terrain qui n'ont pas été détectées et les anomalies détectées sur des signaux sains (fausses alarmes). La figure 2.19 représente pour deux paramètres k et k' différents, les signaux $\boldsymbol{y}_k$ et $\boldsymbol{y}_{k'}$ associés à ces détections ainsi que leur reconstruction $[\boldsymbol{\Phi}\widehat{\boldsymbol{x}}]_k$ et $[\boldsymbol{\Phi}\widehat{\boldsymbol{x}}]_{k'}$ obtenues par décomposition sur le dictionnaire. Cette figure présente un exemple d'anomalie non détectée (a) et un exemple d'anomalie détectée à tort (b). Dans le premier cas (a), il apparaît nettement que le signal $\boldsymbol{y}_k$ n'a pas pu être estimé correctement par décomposition sur le dictionnaire. Ce constat visuel est confirmé par la faible valeur du coefficient de corrélation (estimée à 0.14) entre ce dernier et sa reconstruction. Cependant, étant données les petites valeurs prises par le paramètre k, l'erreur d'estimation estimée est faible et ne permet pas la détection. Dans le deuxième cas (b), le signal testé $\boldsymbol{y}_{k'}$ et sa reconstruction $[\boldsymbol{\Phi}\widehat{\boldsymbol{x}}]_{k'}$ ont la même allure. Le coefficient de corrélation enregistré entre les deux signaux est d'ailleurs de 0.98. Cependant, l'amplitude du signal $\boldsymbol{y}_{k'}$ est telle que la valeur de l'erreur d'estimation (estimée à 2.88) favorise, à tort, la détection d'une anomalie. Au regard de ces éléments, c'est tout naturellement que nous nous sommes intéressés à l'intérêt du coefficient de corrélation comme indicateur de la présence d'anomalies, en supposant que plus la valeur du coefficient de corrélation entre le signal testé et sa reconstruction est faible, plus la probabilité que les données soient affectées par une anomalies est élevée.

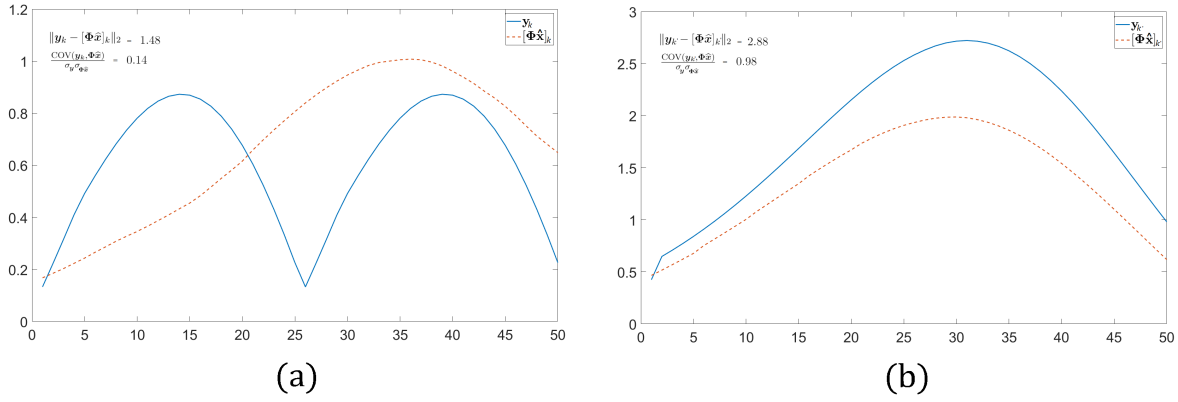

FIGURE 2.19 – Illustration d'une anomalie non détectée et d'une anomalie détectée à tort par l'algorithme ADDICT à travers la représentation d'un signal $\boldsymbol{y}_k$ affecté par une anomalie (a) et d'un signal $\boldsymbol{y}_{k'}$ sain (b) et de leur reconstruction $[\boldsymbol{\Phi}\widehat{\boldsymbol{x}}]_k$ et $[\boldsymbol{\Phi}\widehat{\boldsymbol{x}}]_{k'}$.

Comme nous avons pu le voir dans la section précédente, les poids $w_k, k = 1, ..., K$ interviennent dans l'estimation des signaux $\mathbf{e}_k$. La détection d'anomalie pour le paramètre k, est pénalisée si le poids associé w_k est élevé et est favorisée si ce poids est faible. De manière à respecter cette propriété, nous proposons de considérer la fonction suivante

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

qui est une fonction croissante de c_k, où c_k est le coefficient de corrélation entre le signal testé du paramètre k 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 $c_k = \alpha$, on a $w_k = 1$ (ce qui correspond à l'algorithme non pondéré), que pour $c_k < \alpha$, on a $w_k < 1$ (l'algorithme pondéré favorise la présence d'une anomalie par rapport à l'algorithme non pondéré), et enfin que pour $c_k > \alpha$, on a $w_k > 1$ (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. Un exemple de représentation graphique de la fonction f_α pour $c_k \in]-1, +1[$ et $\alpha = 0.8$ est représenté dans la figure 2.20.

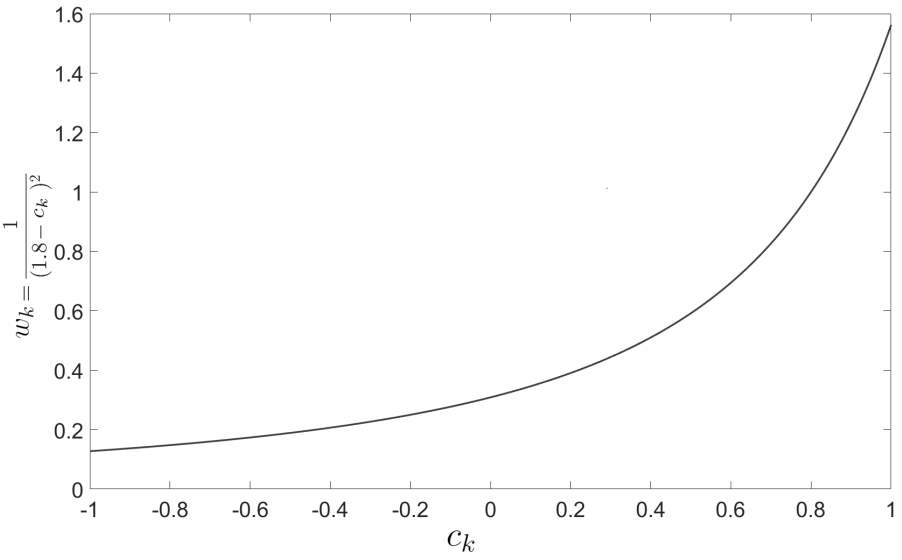

FIGURE 2.20 – Fonction de construction des poids utilisée dans l'algorithme W-ADDICT.

2.4.2.3 Résultats expérimentaux

L'algorithme ADDICT et sa version pondérée W-ADDICT ont été évaluées à l'aide d'une base de données composée de 7 paramètres de télémesure satellite continus affectés par des anomalies connues. Le dictionnaire a été appris à l'aide de l'algorithme K-SVD Aharon et al. (2006) à partir de deux mois de télémesures décrivant un fonctionnement normal du système. Les données de test se composent de 1000 signaux dont 82 sont affectés par une anomalie et répertoriés dans la vérité terrain. Les anomalies ont été observées sur 5 périodes de différentes durées. Les 5 périodes d'anomalies sont visibles dans la figure 2.21.
Les périodes d'anomalie connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 2.21. Les seuils de détection donnés dans la table 2.3 sont représentés par les droites horizontales rouges et ont été fixés à l'aide de courbes CORs calculées grâce à la vérité terrain. On observe que des scores supérieurs à la moyenne ont été retournés pour toutes les périodes anomalies, ce qui témoigne d'une détection efficace des anomalies et notamment de l'anomalie

multivariée correspondant à l'anomalie #5 visible dans la figure 2.21. D'une manière générale, on observe que l'intégration de l'information portée par le coefficient de corrélation entre entre chaque signal et sa reconstruction (par l'intermédiaire d'une pondération adaptée) a permis de fait un grande majorité des scores retournés. De plus, il est désomrais possible de déterminer une seuil de détection pour lequel certains scores en dehors des périodes d'anomalies sont inférieurs et donc de limiter les fausses alarmes. C'est le cas par exemple des signaux tests $237, 535$ et 921 repérables par des flèches bleues. Certaines fausses alarmes subsistent encore pour les signaux tests $337, 341, 737$ et 741 par exemple.

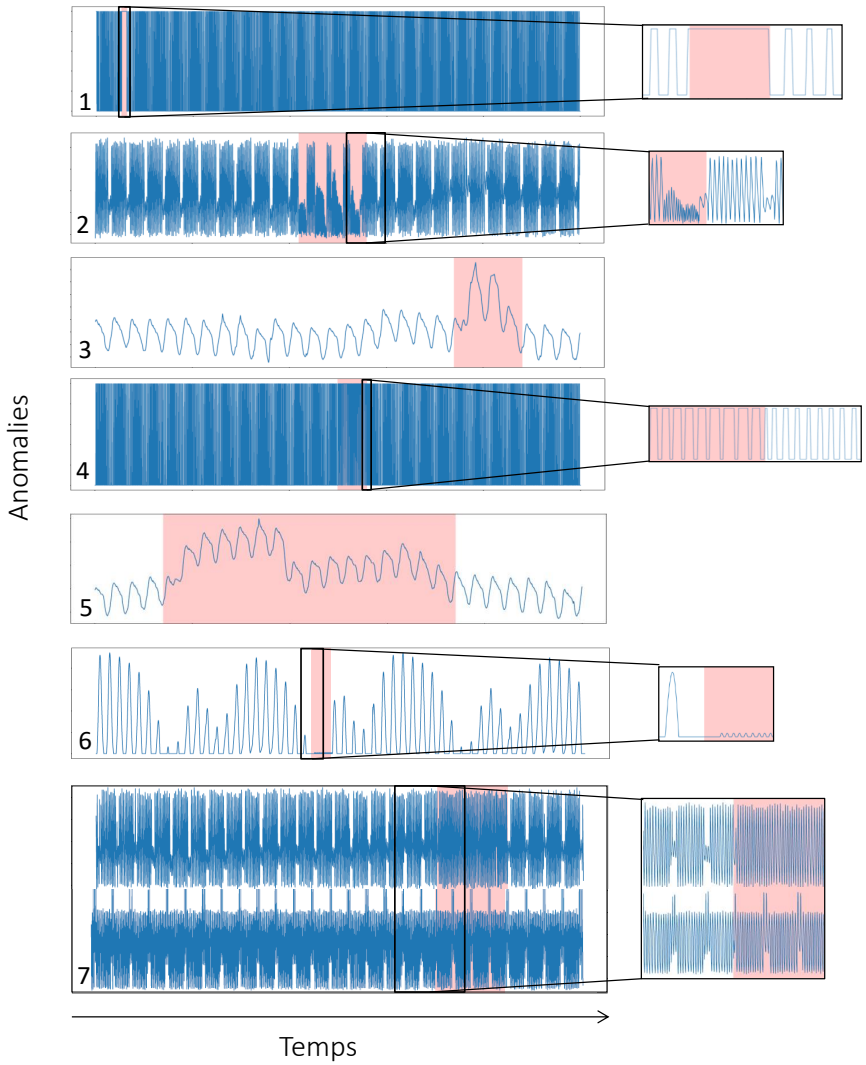

FIGURE 2.21 – Anomalies univariées (1,2,3,4) et multivariée (5) affectant des paramètres continues de télémesure satellites.

Les performances de détection en termes de probabilités de détection (P_{D}) et de probabilités de fausses alarmes (P_{FA}) sont représentées sous forme de courbes CORs dans la figure 2.24. Pour chacune des méthodes, la courbe COR est associée à un réglage optimal des hyperparamètres obtenus à l'aide de la vérité terrain. Au regard des deux courbes CORs, le gain apporté par la pondération semble marginal. Cependant, si l'on s'intéresse aux résultats reportés dans la table 2.3, on observe que la probabilité de fausse alarme est passée de 9.84% pour le modèle non pondéré à 3.06% pour le modèle pondéré pour une probabilité de détection de plus de 90%. La réduction de la probabilité de fausse alarme représente un enjeux opérationnel important de mobilisation de ressources car chaque anomalie détectée doit être examinée par les experts afin de determiner s'il s'agit réellement d'une anomalies ou d'une fausse alarme. Cette section montre que l'introduction d'information externe est intéressante pour réduire le nombre de fausses alarmes.

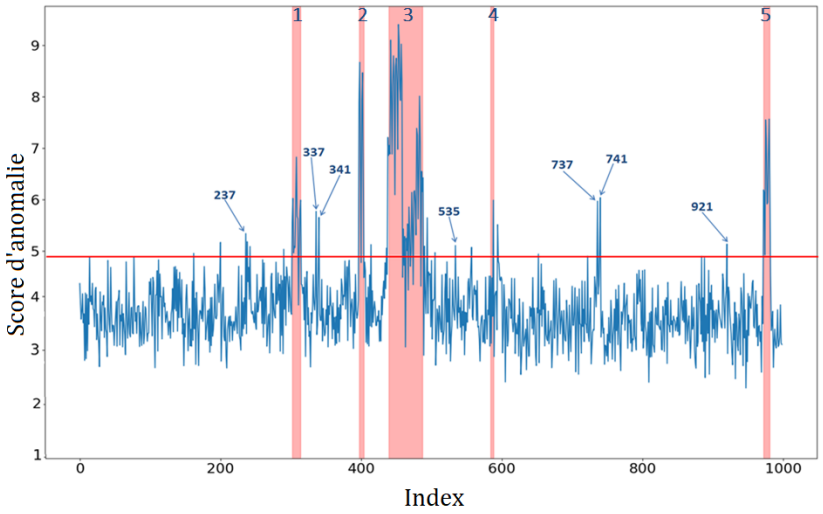

FIGURE 2.22 – représentation des scores d'anomalie retournés par l'algorithme ADDICT pour les signaux de la base d'anomalies. La vérité terrain est représentée par les fonds rouges et est numérotée de manière à pouvoir être identifiée dans la figure 2.21. Le seuil de détection a été déterminé à l'aide de courbes CORs et est représenté par une droite horizontale rouge. Certaines fausses alarmes sont identifiables par leur index et repérables par une flèche bleue.

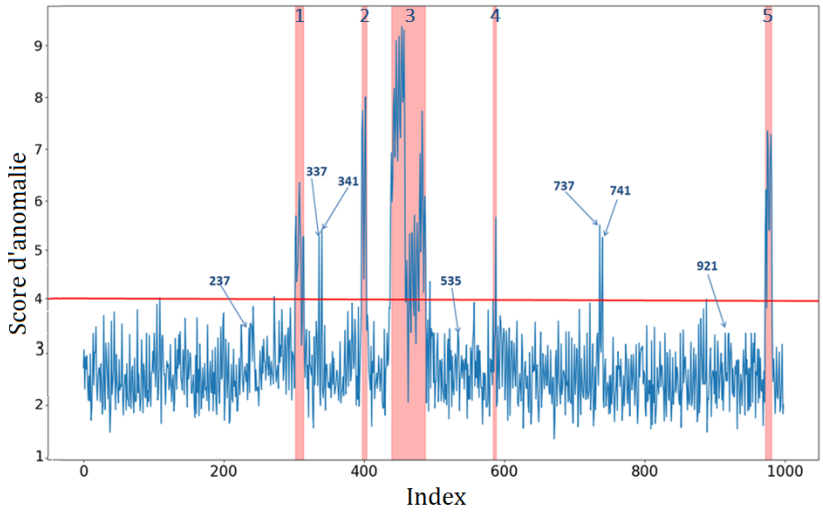

FIGURE 2.23 – Représentation des scores d'anomalie retournés par l'algorithme W-ADDICT pour les signaux de la base d'anomalies. La vérité terrain est représentée par les fonds rouges et est numérotée de manière à pouvoir être identifiée dans la figure 2.21. Le seuil de détection a été déterminé à l'aide de courbes CORs et est représenté par une droite horizontale rouge. Certaines fausses alarmes sont identifiables par leur index et repérables par une flèche bleue.

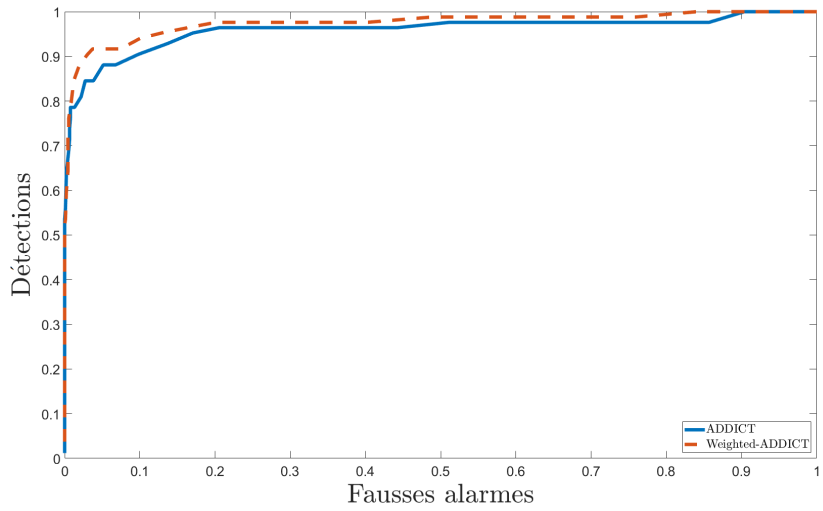

FIGURE 2.24 – Comparaison des courbes CORs calculées à partir des résultats de détection retournés par l'algorithme ADDICT et sa version pondérée W-ADDICT.

TABLE 2.3 – Résultats de détection en termes de probabilité de détection (P_{D}), de probabilité de fausses alarmes (P_{FA}) et d'aire sous la courbe COR (AUC) pour l'algorithme ADDICT et sa version pondérée W-ADDICT.

Méthode	Seuil	P_{D}	P_{FA}	AUC
ADDICT	4.7	90.48%	9.84%	0.937
W-ADDICT	3.9	90.48%	3.06%	0.96

2.5 Réglage des hyperparamètres

2.5.1 Les hyperparamètres de l'algorithme ADDICT

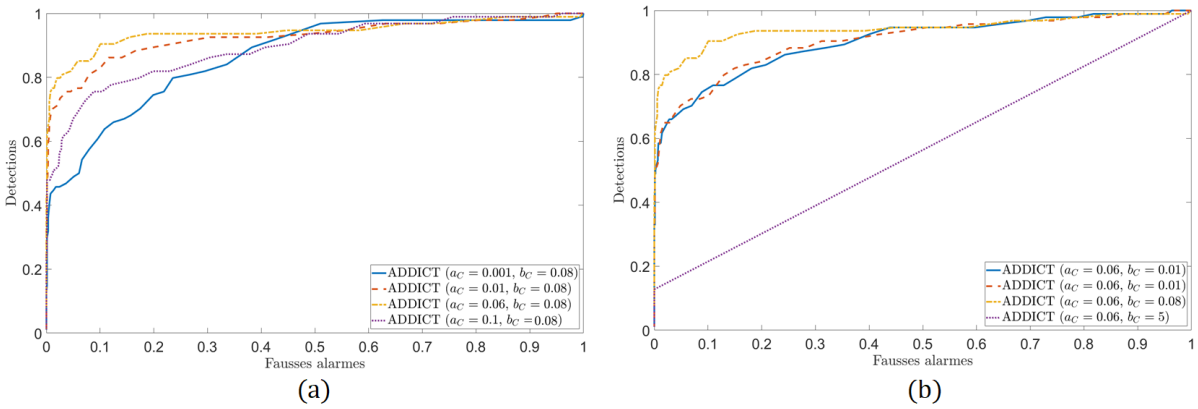

FIGURE 2.25 – Comparaison des courbes CORs calculées à partir des résultats de détection retournés par l'algorithme ADDICT en faisant varier la valeur de a_C (a) et en faisant varier la valeur de b_C (b).

L'algorithme ADDICT dans sa version de base compte trois hyperparamètres, à savoir b_D, a_C, b_C, et un seuil de détection S_{PFA}. Ce dernier dépend de la probabilité de fausse alarme fixée par l'utilisateur. Le choix du seuil de détection est souvent lié au contexte d'application car il représente un compromis entre détection et fausses alarmes. Au regard des résultats présentés par la figure 2.25 il est évident que les hyperparamètres a_C et b_C intervenant dans l'estimation parcimonieuse des données continues influencent les performances de détection. Il en est de même pour l'hyperparamètre b_D qui intervient dans le problème d'estimation parcimonieuse des données discrètes. La figure 2.26 représente les scores associés aux courbes CORs de la figure 2.25 (a) qui décrivent les performances de détection de l'algorithme ADDICT pour plusieurs valeurs de a_C et une valeur de b_C fixée à 0.8. Les périodes d'anomalies connues de la vérité terrain sont indiquées par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1. On observe que plus la valeur de a_C est faible, plus les scores d'anomalies retournés sont faibles. Ceci s'explique par le fait qu'une faible valeur de a_C contraint moins la parcimonie de la combinaison linéaire qui offre alors une meilleure estimation du signal testé. Cependant, une trop faible valeur de a_C rend plus difficile la détection des anomalies car les signaux affectés peuvent eux aussi être très bien estimés par une combinaison linéaire peu parcimonieuse. Pour ce qui est des anomalies de notre vérité terrain, il est clair que les performances de détection présentées dans la figure 2.25(a) se dégradent à mesure que la valeur de a_C diminue. Mais d'après l'analyse des scores de la figure 2.26 il apparaît que même avec une très faible valeur de a_C, il est possible de trouver un seuil de détection pour lequel peu de fausses alarmes sont retournées et au moins un signal affecté par période d'anomalie est détecté. À l'inverse, une valeur trop grande de a_C impose une trop forte parcimonie de la combinaison linéaire et les signaux sains, au même titre que les signaux affectés, sont très mal estimés. Des scores d'anomalie élevés sont retournés

et le taux de fausse alarmes augmente. La même expérience a été réalisée en faisant varier b_C et en fixant la valeur de a_C à 0.05. Les résultats sont présentés dans la figure 2.27 qui présente les scores d'anomalies retournés et la figure 2.25 (b) qui décrit les performances de détection de l'algorithme ADDICT sous forme de courbes CORs. On observe que pour une trop faible valeur de b_C (2.27 (a) et (b)), une proportion importante des anomalies de notre vérité terrain n'est pas détectée. C'est le cas notamment des anomalies #1 et #6. Par ailleurs, une trop grande valeur de ce paramètre mène à l'annulation du signal d'anomalie continu et empêche la détection. Ces résultats témoignent de la difficulté de régler l'hyperparamètre b_C car l'intervalle de valeur pour ce paramètre permettant d'assurer la détection est a priori inconnu. Ces résultats peuvent être extrapolés à la détection d'anomalies dans les données discrètes même si les résultats ne sont pas présentés dans ce manuscrit.

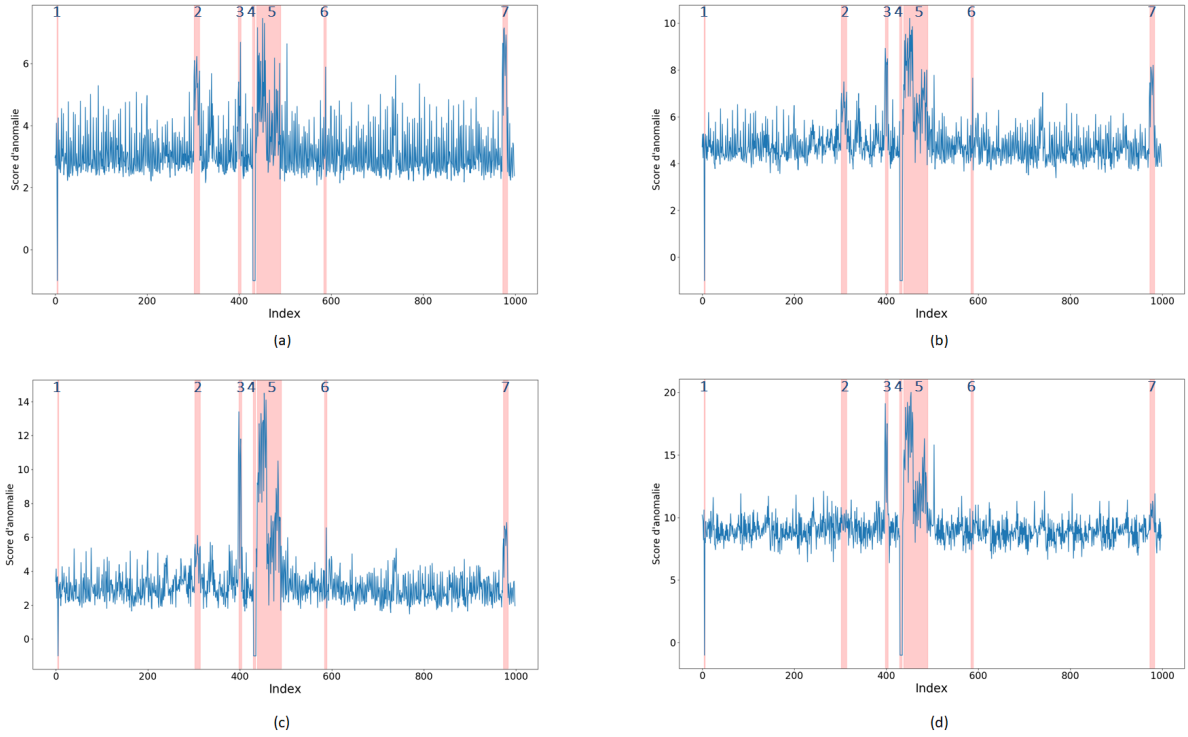

FIGURE 2.26 – Comparaison de scores d'anomalie retournés par l'algorithme ADDICT pour différentes valeurs de a_C et pour $b_C = 0.8$, $b_D = 4$ et $\tau_{max} = 5$. (a) $a_C = 0.001$, (b) $a_C = 0.01$, (c) $a_C = 0.05$, (d) $a_C = 1.5$. La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

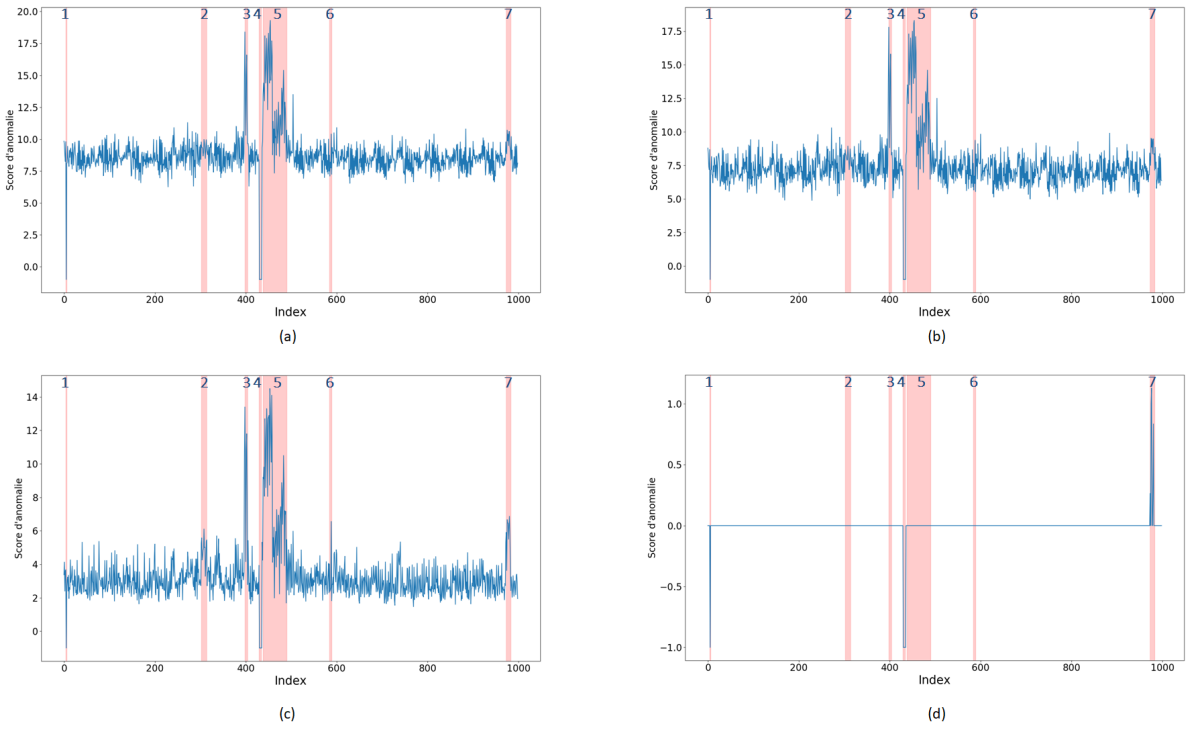

FIGURE 2.27 – Comparaison de scores d'anomalie retournés par l'algorithme ADDICT pour différentes valeurs de b_C et pour $a_C = 0.05$, $b_D = 4$ et $\tau_{max} = 5$. (a) $b_C = 0.01$, (b) $b_C = 0.1$, (c) $b_C = 0.05$, (d) $b_C = 4$. La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

2.5.2 Utilisation de la vérité terrain

La base d'anomalies sur laquelle nous avons mené nos expérimentations nous a permis de réaliser une première évaluation de la méthode proposée et de comparer ses performances à celles d'approches concurrentes. Cette évaluation a été possible grâce à la vérité terrain disponible à partir de laquelle des probabilités de détection et de fausse alarme ont pu être calculées. Au delà de cela, la vérité terrain a également permis de déterminer par une recherche exhaustive (aussi appelée méthode "grid search" d'optimisation d'hyperparamètre) la valeur des hyperparamètres optimaux pour ces données. Le réglage des hyperparamètres par une recherche exhaustive consiste à tester l'algorithme pour un ensemble de combinaisons de valeurs possibles des hyperparamètres et de calculer pour chacune d'elle les performances de l'algorithme obtenues sous forme de courbes CORs. Les courbes CORs ainsi obtenues sont comparées entre elles afin de déterminer la combinaison d'hyperparamètres qui enregistre les meilleures performances. Il existe de nombreuses façons de comparer des courbes CORs. Dans cette étude, la courbe COR retenue correspond à celle admettant un seuil pour lequel le point de coordonné (P_{FA},P_D) est le plus proche du point optimal (0,1). Sous réserve d'une vérité terrain suffisamment exhaustive et représentative, le réglage des hyperparamètres à l'aide de la vérité terrain par une recherche exhaustive assure un réglage au plus juste des hyperparamètres. Cependant la représentativité de la vérité terrain est rarement assurée. Par

ailleurs, plus le nombre d'hyperparamètres à déterminer est important, plus la complexité de calcul est élevée. En effet, si l'on considère par exemple 10 paramètres de télémesure et que l'on souhaite appliquer la méthode G-ADDICT cela suppose de régler 11 hyperparamètres. Le réglage de ces hyperparamètres à l'aide de la vérité terrain par la méthode de la recherche exhaustive pour 20 valeurs différentes de chaque hyperparamètre à tester représente 2.048×10^{14} vecteurs d'hyperparamètres à étudier. Cette méthode n'est donc pas toujours applicable. En conclusion, la méthode de recherche exhaustive est intéressante mais nécessite d'utiliser une base de données expertisée et elle peut conduire à une complexité calculatoire importante. La suite de cette partie présente plusieurs méthodes permettant d'estimer les hyperparamètres directement à partir des données.

2.5.3 Réglage automatique : OC-SDT

Cette partie présente une méthode d'estimation des hyperparamètres à partir de données saines (sans anomalie). Pour simplifier le problème, nous proposons de nous concentrer sur les hyperparamètres b_{D} et b_{C} pour lesquels le réglage peut être réalisé de manière indépendante. Pour ce qui est du réglage de a_{C}, nous nous appuyons sur les résultats obtenus à la section 2.5.1 et préconisons de choisir une faible valeur afin d'assurer la détection au moins partielle des périodes d'anomalie. À défaut d'avoir une vérité terrain disponible, nous supposons que nous disposons d'une quantité suffisante de données saines (sans anomalie) pour apprendre le dictionnaire et estimer les hyperparamètres b_{D} et b_{C}. Initialement, ces données étaient exclusivement réservées à l'apprentissage du dictionnaire. Nous proposons ici de découper l'ensemble de données saines en deux sous-ensembles : un sous-ensemble d'apprentissage utilisé pour l'apprentissage du dictionnaire et un sous-ensemble d'estimation pour le réglage des hyperpramètres b_{D} et b_{C}. Notons que le sous-ensemble d'apprentissage représente une proportion beaucoup plus importante de données que le sous-ensemble d'estimation. À partir d'un dictionnaire appris précédemment, les signaux multivariés discrets et continus du sous-ensemble d'estimation sont décomposés sur leur dictionnaire respectif selon la stratégie ADDICT. Cependant, étant donné que l'on traite des données saines, le signal d'anomalie est supposé nul. La décomposition des données d'estimation s'écrit alors

$$\widehat{\boldsymbol{x}}_{\mathrm{D}} = \arg\min_{\boldsymbol{x}_{\mathrm{D}} \in \mathcal{B}} \parallel \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\Phi}_{\mathrm{D}} \boldsymbol{x}_{\mathrm{D}} \parallel_2^2 \tag{2.53}$$

Estimation parcimonieuse discrète (cas sain)

$$\widehat{\boldsymbol{x}}_{\mathrm{C}} = \arg\min_{\boldsymbol{x}_{\mathrm{C}}} \frac{1}{2} \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathrm{C}} \boldsymbol{x}_{\mathrm{C}}\|_2^2 + a_{\mathrm{C}} \|\boldsymbol{x}_{\mathrm{C}}\|_1 \tag{2.54}$$

Estimation parcimonieuse continue (cas sain)

L'idée de cette section est d'estimer les hyperparamètres b_{D} et b_{C} à partir des erreurs de reconstruction $\mathbf{h}_{\mathrm{D},i} = \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathrm{C}} \widehat{\boldsymbol{x}}_{\mathrm{C}}\|_2^2$ et $\mathbf{h}_{\mathrm{D},i} = \|\boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\Phi}_{\mathrm{C}} \widehat{\boldsymbol{x}}_{\mathrm{C}}\|_2^2, i = 1, ..., I$ issus de la décomposition des signaux d'estimation discrets et continus sur les dictionnaires $\boldsymbol{\Phi}_{\mathrm{D}}$ et $\boldsymbol{\Phi}_{\mathrm{C}}$. Ce choix est motivé par le fait que l'esstimation des signaux d'anomalies $\mathbf{e}_{\mathrm{D}}$ et $\mathbf{e}_{\mathrm{C}}$ est donnée par l'opérateur de seuillage par groupe tel que $\widehat{\mathbf{e}}_{\mathrm{D}} = T_{b_{\mathrm{D}}}(\mathbf{h}_{\mathrm{D}})$, avec $\mathbf{h}_{\mathrm{D}} = \boldsymbol{y}_{\mathrm{D}} - \boldsymbol{\phi}_{\mathrm{D}} \widehat{\boldsymbol{x}}_{\mathrm{D}}$ et

$\widehat{\mathbf{e}}_{\mathrm{C}} = T_{b_{\mathrm{C}}}(\mathbf{h}_{\mathrm{C}})$, avec $\mathbf{h}_{\mathrm{C}} = \boldsymbol{y}_{\mathrm{C}} - \boldsymbol{\phi}_{\mathcal{M}}\widehat{\boldsymbol{x}}_{\mathrm{C}}$. D'une manière générale, sans tenir compte de la nature des données (afin de simplifier les notations) l'estimation du signal d'anomalies $\boldsymbol{e}$ associé à un signal multivarié d'intérêt $\boldsymbol{y}$ (discret ou continu) est obtenue par l'opérateur de seuillage par groupe tel que $\widehat{\mathbf{e}} = T_b(\mathbf{h})$ avec $\mathbf{h} = \boldsymbol{y} - \boldsymbol{\phi}\widehat{\boldsymbol{x}}$ et

$$[T_b(\mathbf{h})]_k = \begin{cases} \frac{\|\mathbf{h}_k\|_2 - b}{\|\mathbf{h}_k\|_2}\mathbf{h}_k & \text{si } \|\mathbf{h}_k\|_2 > b \\ 0 & \text{sinon} \end{cases} \tag{2.55}$$

où $\mathbf{h}_k$ est le $k^{\text{ème}}$ sous-signal de $\mathbf{h}$ associé au $k^{\text{ème}}$ paramètre $k = 1, ..., K$. L'hyperparamètre b joue alors un rôle de seuil de manière à annuler le signal d'anomalie associé au paramètre k si l'erreur d'estimation $\|\mathbf{h}_k\|_2$ est inférieur à b. Un signal d'anomalie nul suppose que le signal associé au paramètre k a pu être correctement estimé par une combinaison linéaire parcimonieuse des atomes du dictionnaire et donc que le paramètre n'est pas affecté par une anomalie. L'idée des méthodes étudiées dans cette partie pour le réglage de l'hyperapramètre b est d'étudier les erreurs d'estimation issues de la décomposition de données saines sur un dictionnaire. De manière naïve, étant donnée la définition de l'opérateur de seuillage par groupe, la première idée est d'estimer l'hyperpramètre b par le maximum des erreurs d'estimation univariées $\|\mathbf{h}_{i,k}\|$, calculées pour chaque paramètre $k, k = 1, ..., K$ de chaque signal d'estimation i pour $i = 1, ..., I$. Cependant, cette stratégie a montré quelques limitations car elle n'est pas du tout robuste à la présence de signaux atypiques dans l'ensemble d'estimation.

2.5.3.1 Formulation du problème OC-SDT

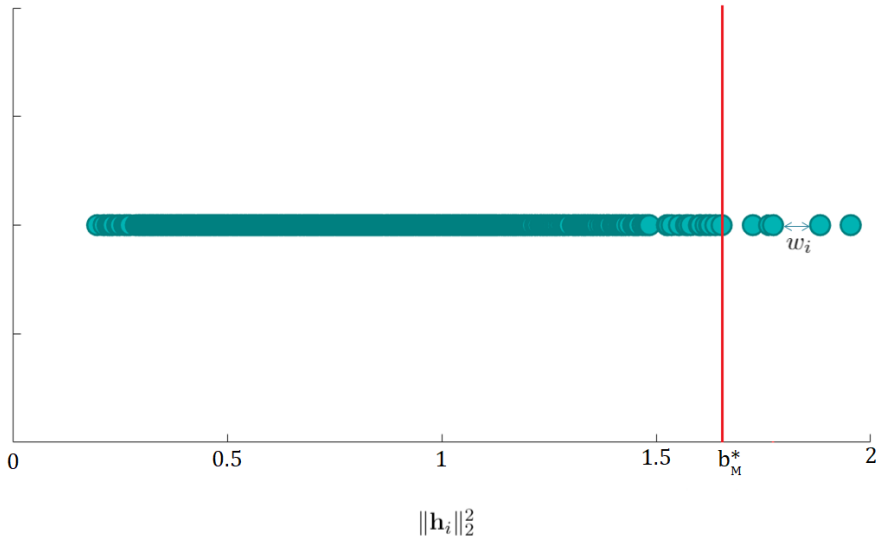

FIGURE 2.28 – Illustration de la stratégie OC-SDT d'estimation des hyperparmètres.

Inspirée des méthodes OC-SVM Schölkopf et al. (2001) et SVDD Tax and Duin (2004), l'approche proposée, appelée One-Class Support Data Thresholding (OC-SDT), est une mé-

thode robuste d'estimation de l'hyperparamètre b. À partir de I signaux multivariés supposés majoritairement sains, l'idée est d'estimer l'hyperparamètre b par le majorant des erreurs d'estimation univariées $\|\mathbf{h}_{i,k}\|_2, i = 1, ..., I, k = 1, ..., K$ tout en tenant compte de la présence éventuelle de quelques signaux plus atypiques dans l'ensemble d'estimation, c'est à dire en autorisant quelques signaux à enregistrer une erreur d'estimation supérieure à b. Afin de limiter la complexité de calcul, nous proposons de rechercher un seuil multivarié, noté b_{M}, qui majore les erreurs d'estimation multivariées $\|\mathbf{h}_i\|_2^2 = \sum_k \|\mathbf{h}_{i,k}\|_2^2$ et d'en déduire le seuil univarié b. Le caractère robuste de la méthode proposée repose sur l'hypothèse qu'une certaine proportion (inconnue) des signaux d'estimation ont des erreurs d'estimation supérieures à b_M. D'un point de vue mathématique, nous proposons de résoudre le problème suivant

$$\min_{b_{\mathrm{M}}} b_{\mathrm{M}} + C \sum_i w_i \xi_i \tag{2.56}$$

$$\text{s.c. } \|\mathbf{h_i}\|_2^2 < b_{\mathrm{M}} + w_i \xi_i \tag{2.57}$$

$$\xi_i \geq 0 \tag{2.58}$$

où $\|\mathbf{h_i}\|_2^2$ est l'erreur d'estimation du $i^{\text{ème}}$ signal multivarié décomposé sur le dictionnaire tel que $\|\mathbf{h_i}\|_2^2 < \|\mathbf{h_{i+1}}\|_2^2, i = 1, ..., I$, $\xi_i, i = 1, ..., I$ sont des variables ressort (slack variables en anglais) introduites afin d'autoriser certain signaux à être mal estimés et donc à enregistrer une erreur d'estimation supérieur au seuil b_{M}, w_i est le poids associé au $i^{\text{ème}}$ signal d'estimation tel que $w_i = \|\mathbf{h_i}\|_2^2 - \|\mathbf{h_{i-1}}\|_2^2$ et C est un hyperparamètre donné par $C = 1 + \frac{w_{\max}}{\sum_i w_i}$ où $w_{\max}$ correspond à la plus grande valeur de w_i enregistrée. La stratégie proposée pour estimer le seuil b_{M} est illustrée par la figure 2.28 dans laquelle b_{M}^* représentent la valeur de b_{M} recherchée. L'introduction des contraintes (2.57) et (2.58) dans (2.56) par la méthode des multiplicateurs de Lagrange mène à l'expression suivante

$$\mathcal{L}(b, \xi, \alpha_i, \lambda_i) = b + \frac{1}{\nu I} \sum_{i=1}^{I} \xi_i - \sum_{i=1}^{I} \alpha_i (b + \xi_i - \|\mathbf{h_i}\|_2^2) - \sum_{i=1}^{I} \lambda_i \xi_i \tag{2.59}$$

où $\alpha_i \geq 0$ et $\lambda_i \geq 0$ correspondent aux multiplicateurs de Lagrange. L'annulation des dérivées partielles mène au problème dual suivant

$$\max_{\alpha_i} \sum_i \alpha_i \|\mathbf{h_i}\|_2^2 \tag{2.60}$$

$$\text{s.c. } \sum_i \alpha_i = 1 \tag{2.61}$$

$$\alpha_i \in [0, C w_i]. \tag{2.62}$$

Le problème dual obtenu est un problème simple de programmation linéaire pouvant être résolu par l'algorithme du simplex Dantzig (1990) par exemple. Ce nouveau problème met en lumière le rôle des poids w_i et de l'hyperparamètre C. En effet, un poids élevé du $i^{\text{ème}}$ signal d'estimation signifie que ce dernier a été beaucoup moins bien estimé que le signal qui

enregistre la première erreur d'estimation inférieure. Ceci peut laisser penser que malgré sa présence dans le sous-ensemble d'estimation supposé exempt d'anomalie, ce signal ne décrit pas réellement un fonctionnement normal du satellite. Dans un réel contexte de détection, il doit donc être détecté par l'algorithme et la valeur du seuil b_{M} doit alors être inférieure à son erreur d'estimation. Par ailleurs plus le poids du $i^{\text{ème}}$ signal est élevée, plus la valeur retournée pour le seuil b_{M} a de chance de correspondre à l'erreur d'estimation du signal (i-1). L'hyperparamètre C joue également un rôle important dans l'estimation du seuil b_{M}. En effet, une valeur trop grande de C revient à toujours estimer le seuil par la deuxième plus forte erreur d'estimation $\|\mathbf{h_{I-1}}\|_2^2$. D'autre part, plus la valeur de C sera élevée, plus le seuil b_{M} retourné sera élevé. Pour nos premières expérimentations, il nous a paru judicieux de déterminer C à l'aide du poids maximal car il représente "la frontière" entre les sains et les signaux les signaux les plus atypiques de l'ensemble de signaux d'estimation. Par cette méthode il est possible de distinguer les signaux atypiques présents dans l'ensemble d'apprentissage sans avoir à connaitre quelle proportion ils représentent. Enfin, notons que afin de satisfaire les contraintes (2.61) et (2.62), l'hyperparamètre C ne peut être inférieur à 1. L'estimation du paramètre b à partir du seuil b_{M} obtenu par la méthode OC-SDT est donné par

$$\widehat{b} = \sqrt{\frac{b_{\mathrm{M}}}{K}} \tag{2.63}$$

avec K le nombre de paramètres de télémesure étudiés. La méthode OC-SDT peut également être appliquée pour déterminer les hyperparamètres du modèle G-ADDICT. Il s'agit alors de résoudre pour chaque paramètre k le problème (2.56) sous les contraintes (2.57) et (2.58) à partir des erreurs d'estimation univariées (calculées pour chaque paramètre) notées $\|\mathbf{h}_{i,k}\|_2$, et non plus à partir des erreurs $\|\mathbf{h}_i\|_2^2$. L'estimation du paramètre b_k associé au $k^{\text{ème}}$ paramètre de télémesure est alors directement donné par résolution du problème OC-SDT pour le $k^{\text{ème}}$ paramètre.

2.5.3.2 Méthode concurrente : SURE

Cette section s'intéresse à une méthode concurrente de l'état de l'art permettant d'estimer l'hyperparamètre b à l'aide de l'estimateur non biaisé de Stein ou Stein's Unbiased Risk Estimator (SURE) Stein (1981). Cette méthode a été appliquée avec succès pour de nombreuses applications Donoho and Johnstone (1994); Luisier et al. (2011); Giryes et al.; Pesquet et al. (2009); Eldar (2009). Par ailleurs, le choix d'étudier cette approche est motivé par le fait que son application ne nécessite de régler aucun hyperparamètre. Cependant elle impose de travailler avec des données issues d'une distribution normale, ce que nous supposerons pour les erreurs d'estimation $\|\mathbf{h}_i\|_2^2$.
On considère le vecteur d'anomalie $\mathbf{Q} \in \mathbb{R}^I$ inconnu, associé aux données saines du sous-ensemble d'estimation et défini tel que $\mathbf{Q}(i) = \|\boldsymbol{e}_i\|_2^2, i = 1, ..., I$ où $\mathbf{Q}(i)$ représente la $i^{\text{ème}}$ composante de $\mathbf{Q}$ et $\boldsymbol{e}_i$ correspond au signal d'anomalie inconnu associé au $i^{\text{ème}}$ signal du sous-ensemble d'estimation. L'idée est alors d'estimer le vecteur $\mathbf{Q}$ à partir du vecteur $\mathbf{h}$ composé des erreurs d'estimation des différents signaux multivariés du sous-ensemble d'estimation tel que $\mathbf{h}(i) = \|\mathbf{h}_i\|_2^2$. La méthode SURE suppose que les erreurs d'estimation qui composent

le vecteur $\mathbf{h}$ sont supposées indépendantes et approximativement distribuées selon une loi normale de variance $\sigma^2 = 1$ et de moyennes μ_i inconnues. Étant donné que l'on traite des signaux sains, les signaux d'anomalie $\boldsymbol{e}_i, i = 1, ..., I$ associés à ces derniers sont souvent nuls et le vecteur $\mathbf{Q}$ compte alors de nombreuses valeurs nulles. Un estimateur pertinent de ce vecteur est alors donné par $\widehat{\mathbf{Q}} = S_{b_\mathrm{M}}(\mathbf{h})$ où $S_{b_\mathrm{M}}(\mathbf{h})$ correspond à l'opérateur de seuillage par élément dont nous rappelons la définition

$$[S_{b_\mathrm{M}}(\mathbf{h})]_i = \begin{cases} \mathbf{h}(i) - b_\mathrm{M} & \text{si } \mathbf{h}(i) > b_\mathrm{M} \\ 0 & \text{si } |\mathbf{h}(i)| \leq b_\mathrm{M} \\ \mathbf{h}(i) + b_\mathrm{M} & \text{si } \mathbf{h}(i) < -b_\mathrm{M} \end{cases} \tag{2.64}$$

où $[S_{b_\mathrm{M}}(\mathbf{h})]_i$ est la $i^{ème}$ composante de $S_{b_\mathrm{M}}(\mathbf{h})$ et $\mathbf{h}(i)$ est la $i^{ème}$ composante de $\mathbf{h}$. La stratégie étudiée ici consiste à déterminer la valeur du seuil b_M permettant d'estimer au mieux le vecteur $\mathbf{Q}$ à l'aide de la méthode SURE. Cette dernière consiste à trouver la valeur de b_M qui minimise l'erreur quadratique moyenne, ou Mean Square Error (MSE) an anglais, de l'estimateur $\widehat{\mathbf{Q}}$ définie comme suit

$$\mathrm{MSE}(\widehat{\mathbf{Q}}) = \mathbb{E}_\mathbf{Q}\|\widehat{\mathbf{Q}} - \mathbf{Q}\|_2^2 \tag{2.65}$$

$$= \mathbb{E}_\mathbf{Q}\|S_{b_\mathrm{M}}(\mathbf{h}) - \mathbf{Q}\|_2^2. \tag{2.66}$$

Comme démontré par Stein dans Stein (1981), le MSE peut être estimé par

$$\mathrm{MSE}(\widehat{\mathbf{Q}}) = \mathbb{E}_\mathbf{Q}[\mathrm{SURE}(\mathbf{Q})] \tag{2.67}$$

où

$$\mathrm{SURE}(\mathbf{Q}) = I + \|g(\mathbf{h})\|_2^2 + 2\sum_{i=1}^{I} \frac{\partial}{\partial \mathbf{h}(i)} g_i(\mathbf{h}) \tag{2.68}$$

avec $g(\mathbf{h}) = S_{b_\mathrm{M}}(\mathbf{h}) - \mathbf{Q}$ et $g_i(\mathbf{h})$ la $i^{ème}$ composante de $g(\mathbf{h})$. D'après Donoho and Johnstone (1994) l'expression de l'estimateur SURE pour l'opérateur de seuillage par élément est donnée par

$$\mathrm{SURE}(\mathbf{Q}) = I - 2\#\{i : |\mathbf{h}(i)| \leq b_\mathrm{M}\} + \sum_{i=1}^{I} (\min(\mathbf{h}^2(i), b_\mathrm{M}^2). \tag{2.69}$$

La minimisation de l'estimateur SURE défini par (2.69) mène à la valeur du seuil b_M qui minimise l'erreur quadratique moyenne (2.66). L'estimation du paramètre b à partir du seuil b_M est donné par

$$\widehat{b} = \sqrt{\frac{b_\mathrm{M}}{K}} \tag{2.70}$$

avec K le nombre de paramètres de télémesure étudiés. La méthode SURE peut également être appliquée pour de déterminer les hyperparamètres du modèle G-ADDICT. Il s'agit alors de minimiser l'estimateur SURE Tibshirani and Wasserman (2015) à partir des erreurs d'estimation notées $\|\mathbf{h}_{i,k}\|_2$ et non plus des erreurs $\|\mathbf{h}_i\|_2^2$. L'estimation du paramètre b_k associé au $k^{\text{ème}}$ paramètre de télémesure est alors directement donné par le seuil $b_{\mathrm{M,k}}$ issue de la résolution du problème SURE pour le $k^{\text{ème}}$ paramètre.

2.5.3.3 Expérimentations

La méthode proposée OC-SDT et la méthode SURE ont été appliquées aux erreurs d'estimation de 1000 signaux multivariés sains décomposés sur un dictionnaire. L'algorithme ADDICT a ensuite pu être évalué avec les valeurs des hyperparamètres obtenues par ces méthodes. La figure 2.29 représente les courbes CORs décrivant les performances sur notre base d'anomalies de l'algorithme ADDICT lorsque les hyperparamètres b_{D} et b_{C} ont été réglés à l'aide de la méthode OC-SDT, de la méthode SURE et de la vérité terrain disponible (VT). On observe que les méthodes étudiées permettent un réglage automatique des paramètres pour lesquels les performances enregistrées sont proches des performances obtenues avec les hyperparamètres obtenus à l'aide la vérité terrain. Notons par ailleurs que l'application des méthodes OC-SDT et SURE est beaucoup moins coûteuse en temps de calcul que la mise en place de la recherche exhaustive.
Comme évoqué dans les sections précédentes, les méthodes SURE et OC-SDT peuvent également permettre de régler les nombreux hyperparamètres du modèle G-ADDICT et ainsi faciliter son utilisation. La figure 2.30 représente les courbes CORs décrivant les performances sur notre base d'anomalies de l'algorithme G-ADDICT avec des hyperparamètres $b_{\mathrm{k}}, k = 1, ..., K$ estimés à l'aide la méthode OC-SDT, de la méthode SURE et de la vérité terrain (VT). On observe que les performances de l'algorithme avec les hyperparamètres OC-SDT correspondent à celles obtenues avec les hyperparamètres optimaux déterminés à l'aide de la vérité terrain. Des mesures de performances quantitatives sont données dans la table 2.4 qui montre des probabilités de détection de 93.6% et 94.7% et des probabilités de fausses alarmes de 3.3% et 4.6%. Avec les hyperparamètres déterminés par la méthode SURE, la probabilité de fausse alarme est inférieur à 1% mais 25% des signaux affectés par une anomalies ne sont pas détectés. Les scores d'anomalie retournés par la méthode G-ADDICT à partir desquels ont été calculées les performances de détection de la figure 2.30 sont représentés dans la figure 2.31. On observe que les scores sont non nuls presque exclusivement en période d'anomalie ce qui témoigne d'une détection d'anomalies efficace. Il apparaît également que la méthode retourne plus de fausses alarmes avec le réglage OC-SDT qu'avec le réglage SURE. Cependant, une proportion importante des signaux affectés de la période d'anomalies #5 ne sont pas détectés avec un réglage réalisé à l'aide de la méthode SURE ce qui explique les résultats reportés dans la figure 2.30 et la table 2.4.

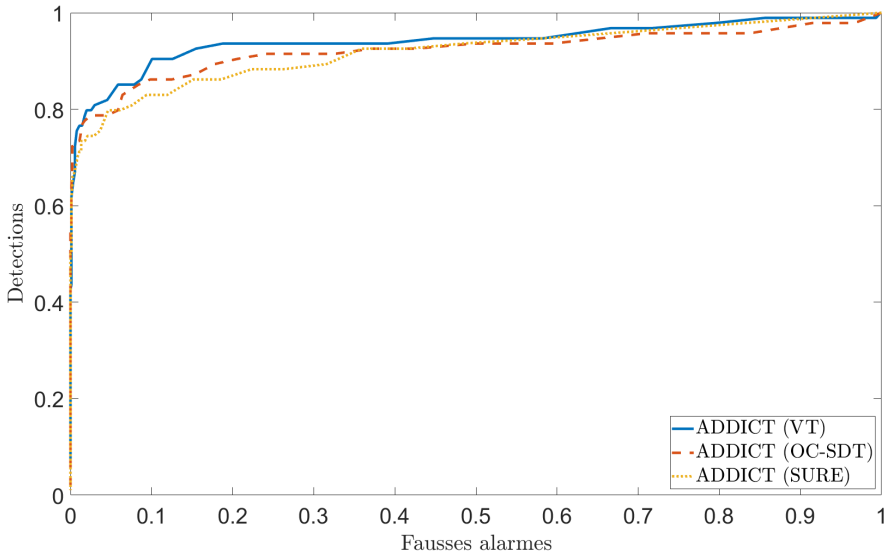

FIGURE 2.29 – Comparaison des courbes CORs obtenus à partir des résultats de détection retournés par ADDICT dont les hyperparamètres ont été réglés à l'aide de la vérité terrain, de la méthode OC-SDT et de la méthode SURE.

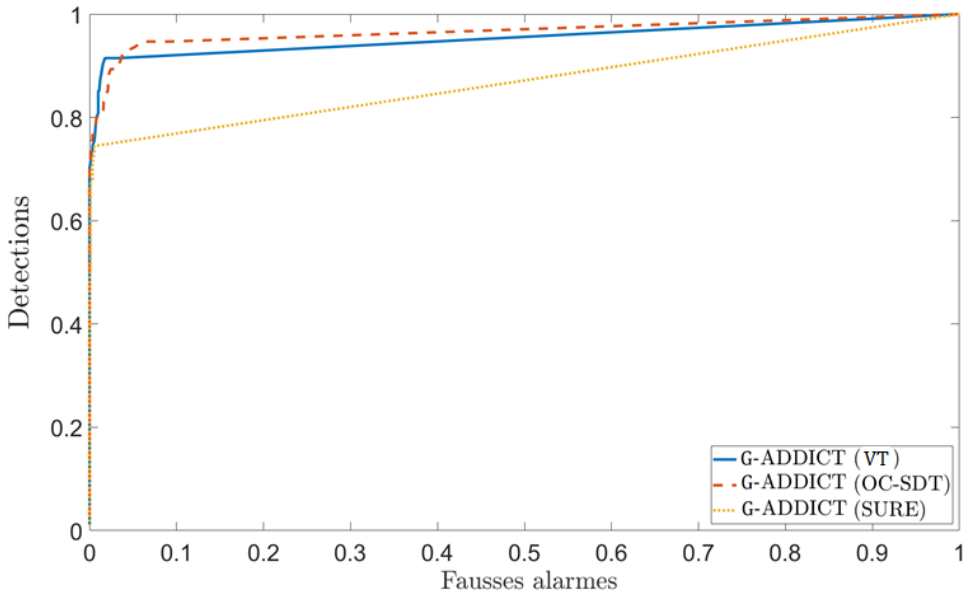

FIGURE 2.30 – Comparaison des courbes CORs obtenus à partir des résultats de détection retournés par G-ADDICT dont les hyperparamètres ont été réglés à l'aide de la vérité terrain, de la méthode OC-SDT et de la méthode SURE.

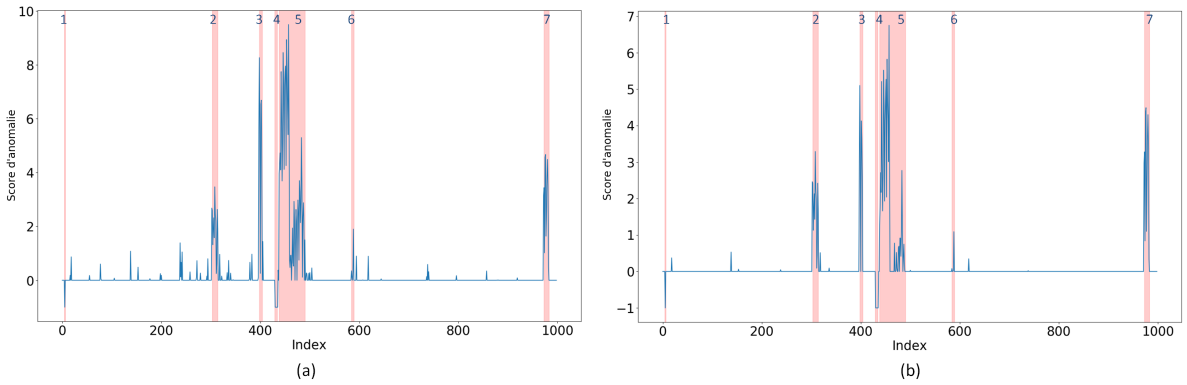

FIGURE 2.31 – Comparaison des scores d'anomalie retournés par l'algorithme G-ADDICT avec des hyperparamètres estimés à l'aide de la méthode OC-SDT (a) et SURE (b). La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

TABLE 2.4 – 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) de la méthode G-ADDICT dont les hyperparamètres ont été déterminés à l'aide de la méthode OC-SDT, SURE et de la vérité terrain.

Méthode	S_{PFA}	P_{D}	P_{FA}	AUC
G-ADDICT (VT)	$\epsilon > 0$	93.6%	3.3%	0.9775
G-ADDICT (OC-SDT)	$\epsilon > 0$	94.7%	4.6%	0.9452
G-ADDICT (SURE)	$\epsilon > 0$	74.5%	0.6%	0.8851

2.6 Conclusion

Ce chapitre présente l'algorithme ADDICT, une nouvelle méthode de détection d'anomalies multivariées qui repose sur un modèle d'estimation parcimonieuse. L'originalité de cette méthode réside dans le fait qu'elle est capable de traiter conjointement des séries temporelles discrètes et continues et de prendre en compte les relations et corrélations qui existent entre elles. Les premiers résultats obtenus sur une base de test composée de diverses anomalies connues ont confirmé l'intérêt des méthodes d'estimation parcimonieuse et d'apprentissage de dictionnaires pour la détection d'anomalies et ont démontré la compétitivité de l'algorithme proposé vis à vis de plusieurs méthodes de l'état de l'art évaluées sur les mêmes données. Il est également apparu que l'activation de l'option d'invariance par translation semble indispensable pour représenter efficacement les données, notamment discrètes, à l'aide d'une décomposition parcimonieuse.

Par la suite, plusieurs extensions ont été étudiées. Tout d'abord, une version généralisée appelée G-ADDICT permettant d'adapter la stratégie ADDICT au cadre multivarié a été étudiée. Les premiers résultats obtenus témoignent de l'intérêt de la version généralisée qui

assure une détection d'anomalies plus juste et facilite le choix du seuil de détection, ce que n'a pas permis la normalisation des données. Celle-ci autorise un réglage plus fin de l'algorithme, au niveau de chaque paramètre de télémesure. En effet, rappelons que la généralisation du problème ADDICT a permis de détecter presque 95% des anomalies de notre base pour moins de 5% de fausses retournées, tandis qu'il faut accepter presque 10% de fausses alarmes pour assurer 90% de détection avec l'algorithme ADDICT. Cependant l'application de la méthode généralisée impose de régler un nombre important d'hyperparamètres. Étant donné le défi que cela peut représenter dans certains cas, cette méthode n'est pas toujours applicable.
Une version pondérée appelé W-ADDICT a été proposée afin d'intégrer de l'information externe par l'intermédiaire d'une pondération adaptée à chaque paramètre. Plus précisément, nous avons étudié l'intérêt du coefficient de corrélation entre le signal mesuré et sa reconstruction pour la détection d'anomalies. La pondération proposée utilise ce coefficient de corrélation pour tirer profit des dynamiques différentes de chaque signal ce qui permet d'améliorer les performances.

Enfin, nous nous sommes intéressés au réglage des hyperparamètres de l'algorithme ADDICT qui peut s'avérer plus ou moins difficile selon que l'on dispose ou non d'une vérité terrain et de capacité de calcul suffisante. Nous avons alors proposé une nouvelle méthode d'estimation des hyperparamètres, appelée OC-SDT, qui permet de déterminer simplement certains hyperparamètres ADDICT à l'aide de données saines décrivant un fonctionnement normal du satellite. Ces mêmes paramètres ont également été estimés à l'aide d'une méthode de l'état de l'art appliquée aux mêmes données. Les premiers résultats obtenus témoignent de l'intérêt de ces méthodes qui facilitent considérablement l'utilisation de l'algorithme et notamment de l'algorithme G-ADDICT et assure une détection efficace des anomalies.
Au regard des résultats expérimentaux obtenus dans cette étude, nos recommandations d'utilisation de l'algorithme ADDICT sont les suivantes :

- Si les données sont très hétérogènes, nous conseillons d'utiliser l'algorithme G-ADDICT car il semble plus adapté au cadre multivarié et facilite le réglage du seuil de détection.
- Pour la mise en oeuvre de l'algorithme choisi, nous préconisons d'initialiser les hyperparamètres à l'aide de la méthode OC-SDT. Cette méthode a l'avantage de ne pas avoir d'hyperparamètre à régler, de ne pas avoir besoin d'une vérité terrain. Les hyperparamètres pourront par la suite être ajustés compte tenu de l'expérience des premiers résultats obtenus et des contraintes opérationnelles.
- Nous pensons qu'il peut être judicieux de combiner les extensions G-ADDICT et W-ADDICT avec des poids construits à partir du coefficient de corrélation car ces derniers permettent d'améliorer les performances de détection de l'algorithme sans complexifier les calculs. Par ailleurs, il peut être intéressant d'étudier l'intérêt d'une pondération différentes permettant d'intégrer d'autres formes d'information utiles à la détection.

Chapitre 3

Estimation parcimonieuse convolutive

Sommaire

3.1 Introduction

Nous nous intéressons dans ce chapitre aux méthodes d'estimation parcimonieuse convolutive et proposons l'algorithme C-ADDICT, une extension de l'algorithme ADDICT. Contrairement aux approches d'estimation parcimonieuse classiques telle que celle sur laquelle repose l'algorithme ADDICT, ces méthodes sont par définition invariantes par translation et peuvent donc s'avérer très intéressantes pour notre problème de détection d'anomalies multivariées dans des données mixtes. Les travaux présentés dans cette partie ont été menés en collaboration avec Gustavo Silva et Paul Rodriguez de la Pontificia Universidad Catolica del Peru (PUCP) et ont été publiés dans Pilastre et al. (2020b).
Le premier modèle d'estimation parcimonieuse invariant par translation à été formulé par Lewiski and Sejnowski (1999) en utilisant un dictionnaire convolutif d'éléments prédéfinis. Plus tard, une formulation du problème d'apprentissage de dictionnaire pour l'estimation parcimonieuse invariante par translation, appelée maintenant estimation parcimonieuse convolutive (Convolutional Sparse Coding, CSC en anglais), a été proposée dans Grosse et al. (2007). Ces approches ont été étudiées pour de nombreuses applications de traitement du signal et des images Barthélemy et al. (2013); Zeiler et al. (2010); Heide et al. (2010); Wohlberg (2016a); Moreau (2017); Dupre la Tour (2018) telle que la détection d'anomalies Carrera et al. (2015). Les sections qui suivent présentent succinctement les aspects théoriques relatifs au problème non contraint d'estimation parcimonieuse convolutive et présente l'algorithme C-ADDICT. De par sa propriété d'invariance par translation, l'approche C-ADDICT permet de traiter conjointement des signaux de télémesure mixtes selon une unique stratégie qui prend en compte les liens possibles entres les paramètres et assure la détection des anomalies multivariées. De plus, contrairement au modèle ADDICT, le modèle C-ADDICT proposé pour la détection d'anomalies a pu être étendu au problème d'apprentissage de dictionnaire convolutif de données mixtes. Les performances de détection de l'algorithme C-ADDICT ont été évaluées sur la même base d'anomalies utilisée pour l'étude de l'algorithme ADDICT et ses extensions, et ses performances de détection ont été comparées à celles de plusieurs méthodes de l'état de l'art.

3.2 Estimation parcimonieuse convolutive et apprentissage de dictionnaire

3.2.1 Estimation parcimonieuse convolutive

L'estimation parcimonieuse convolutive consiste à exprimer un signal $\boldsymbol{y} \in \mathbb{R}^N$ comme une somme de convolutions entre les éléments, appelés filtres, d'un dictionnaire $\boldsymbol{\Phi} \in \mathbb{R}^{M \times L}$, et les cartes de coefficients parcimonieuses, notées $\boldsymbol{x}_l \in \mathbb{R}^N$ (coefficient maps en anglais), avec $M < N$, i.e.,

$$\boldsymbol{y} = \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_l + \boldsymbol{b} \tag{3.1}$$

où * indique un produit de convolution, $\boldsymbol{\phi}_l \in \mathbb{R}^M, l = 1, ..., L$ est le $l^{\text{ème}}$ filtre du dictionnaire et $\boldsymbol{b}$ est un bruit additif. La manière la plus courante d'estimer les cartes de coefficients est d'utiliser une extension du problème LASSO telle que

$$\widehat{\boldsymbol{x}}_l = \arg\min_{\boldsymbol{x}_l} \frac{1}{2} \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_l - \boldsymbol{y} \|_2^2 + \lambda \sum_l \| \boldsymbol{x}_l \|_1 \tag{3.2}$$

où λ est un paramètre de régularisation (ou hyperparamètre) contrôlant la parcimonie des cartes de coefficients $\boldsymbol{x}_l$ à l'aide d'une pénalisation ℓ_1. Sous réserve d'un dictionnaire composé de filtres représentatifs des données étudiées, le modèle convolutif permet d'estimer tout nouveau signal $\boldsymbol{y}$ de manière invariante par translation par activation des filtres $\boldsymbol{\phi}_l$ (de taille inférieure à celle du signal d'entrée $\boldsymbol{y}$) à l'instant exact où ils sont observés dans le signal à estimer. La figure 3.1 permet d'illustrer l'intérêt de l'invariance par translation pour l'estimation parcimonieuse à travers un exemple simple dans lequel le signal à estimer est une image composée de chiffres. Les trois chiffres présents sur l'image ont été appris et composent les filtres d'un dictionnaire convolutif et les atomes d'un dictionnaire fixe. Á partir de ces filtres, le modèle convolutif est capable de reconstruire parfaitement l'image. Cependant, faute d'avoir appris la bonne position du chiffre 7 dans l'image, l'estimation parcimonieuse à l'aide du dictionnaire fixe est plus difficile. Comme présenté dans la figure 3.2, l'estimation parcimonieuse convolutive classique peut être adaptée afin d'estimer conjointement et de manière invariante par translation des signaux de télémesure multivariés, ce qui a motivé cette étude.

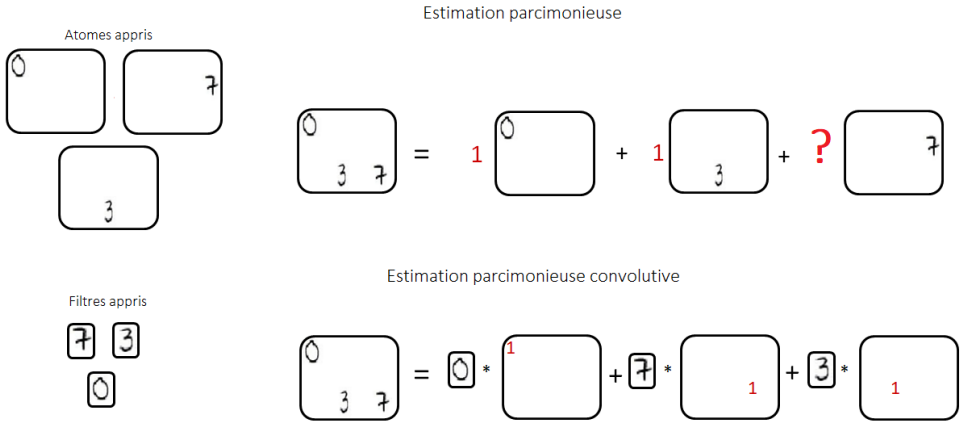

FIGURE 3.1 – Illustration de l'intérêt de l'estimation parcimonieuse convolutive, invariante par translation, par rapport à l'estimation parcimonieuse classique.

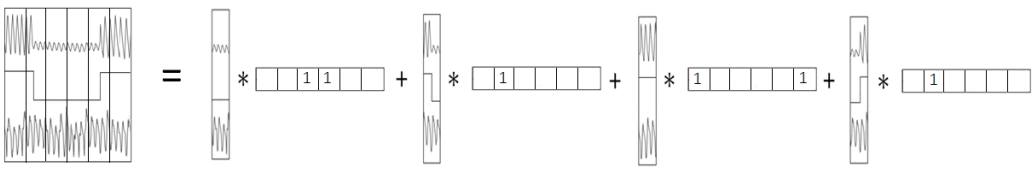

FIGURE 3.2 – Illustration de l'estimation parcimonieuse convolutive appliquée à des signaux de télémesure satellite.

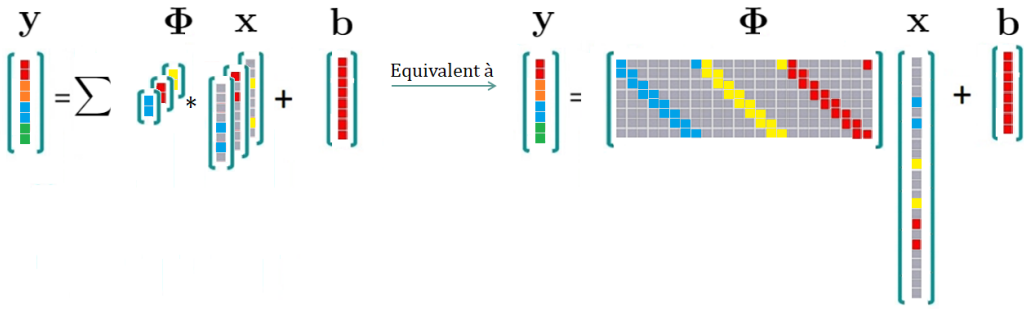

FIGURE 3.3 – Illustration de la reformulation du problème d'estimation parcimonieuse convolutive (gauche) comme un problème d'estimation parcimonieuse non contraint (droite) en introduisant des matrices de Toeplitz.

Le problème (3.2) est convexe et peut donc être résolu efficacement par l'algorithme FISTA Beck and Teboulle (2009b); Chalasani et al. (2013), par un algorithme de descente de coordonnées adapté Kavukcuoglu et al. (2010) ou encore par l'algortihme ADMM Boyd et al. (2011). Rappelons que l'algorithme ADMM est itératif et met à jour les différentes composantes à estimer de manière alternée par minimisation du Lagrangien augmenté. Il peut être appliqué pour résoudre (3.2) après avoir introduit les variables auxiliaires $\boldsymbol{z}_l$ et les contraintes d'égalité telles que

$$\widehat{\boldsymbol{x}}_l, \widehat{\boldsymbol{z}}_l = \arg\min_{\boldsymbol{x}_l, \boldsymbol{z}_l} \frac{1}{2} \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_l - \boldsymbol{y} \|_2^2 + \lambda \sum_l \|\boldsymbol{z}_l\|_1 \text{ s.c. } \boldsymbol{x}_l = \boldsymbol{z}_l \ \forall l. \tag{3.3}$$

Comme illustré dans la figure 3.3, le problème (3.3) peut être reformulé comme un problème d'estimation parcimonieuse classique faisant intervenir un dictionnaire fixe correspondant à une concaténation des matrices de Toeplitz $\boldsymbol{\Phi}_l \in \mathbb{R}^{N \times N}$ telles que $\boldsymbol{\Phi}_l \boldsymbol{x}_l = \boldsymbol{\phi}_l * \boldsymbol{x}_l$ et les matrices $\boldsymbol{\Phi} \in \mathbb{R}^{N \times NL}$ et $\boldsymbol{x} \in \mathbb{R}^{NL}$ définies ci-dessous

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_L \end{bmatrix} \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_L \end{pmatrix}. \tag{3.4}$$

Le problème (3.3) devient alors

$$\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{z}\|_1 \text{ s.c. } \boldsymbol{x} = \boldsymbol{z} \tag{3.5}$$

qui correspond au problème LASSO (3.5) dont la résolution à l'aide de l'algorithme ADMM est détaillée dans la section 2.2.2.2. Rappelons cependant que la mise à jour de $\boldsymbol{x}$ à l'itération (i+1) nécessite de résoudre le problème de minimisation suivant

$$\widehat{\boldsymbol{x}}^{i+1} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \boldsymbol{m}^i(\boldsymbol{z}^i - \boldsymbol{x}) + \frac{\mu^i}{2} \|\boldsymbol{z}^i - \boldsymbol{x}\|_2^2, \tag{3.6}$$

menant à l'équation

$$(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \rho\boldsymbol{I})\boldsymbol{x} = \boldsymbol{\Phi}^T\boldsymbol{y} + \mu^i(\boldsymbol{z}^i - \boldsymbol{m}^i). \tag{3.7}$$

Étant donné la taille importante de la matrice $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ et la complexité de calcul qu'implique son inversion, l'estimation des cartes de coefficients est coûteuse. En revanche, à l'aide du théorème de convolution, il est facile de montrer que le problème (3.7) est équivalent au problème

$$(\tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}} + \rho\boldsymbol{I})\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{y}} + \rho(\tilde{\boldsymbol{z}} - \tilde{\boldsymbol{m}}). \tag{3.8}$$

où $\tilde{\boldsymbol{\Phi}}$ correspond à la transformée de Fourier discrète de $\boldsymbol{\Phi}$. La mise à jour des cartes de coefficients peut alors être réalisée dans le domaine fréquentiel comme cela a pu être étudié dans Garcia-Cardona and Wohlberg (2018); Silva and Rodríguez (2018); Silva and Rodriguez (2018); Wohlberg (2016b, 2014).

3.2.2 Apprentissage de dictionnaire convolutif

L'apprentissage d'un dictionnaire convolutif à partir des données consiste à capturer les comportements récurrents observés dans les signaux d'apprentissage à partir desquels tout nouveau signal peut être estimé par une somme de convolutions entre quelques filtres du dictionnaire et les cartes de coefficients associées. À partir d'un ensemble de signaux d'apprentissage, notés $\boldsymbol{y}_t, = 1, ..., T$, le problème d'apprentissage d'un dictionnaire convolutif consiste à résoudre le problème suivant

$$\widehat{\boldsymbol{x}}_l, \widehat{\boldsymbol{\phi}}_l = \arg\min_{\boldsymbol{x}_l, \boldsymbol{\phi}_l \in \mathcal{C}} \frac{1}{2}\sum_t \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_{t,l} - \boldsymbol{y}_t \|_2^2 + \lambda \sum_{i,l} \|\boldsymbol{x}_{t,l}\|_1 \tag{3.9}$$

où

$$\mathcal{C} = \{\boldsymbol{\phi}_l \in \mathbb{R}^M, \|\boldsymbol{\phi_l}\|_2 \leq 1, l = 1, ..., L\}$$

La contrainte sur les filtres $\boldsymbol{\phi}_l$ est importante car elle évite que les filtres aient des valeurs arbitrairement élevées et que les cartes de coefficients tendent vers 0. Notons que cette normalisation des filtres du dictionnaire est parfois, et de manière sous-optimale, réalisée à posteriori, après la mise à jour du dictionnaire. Le problème (3.9) est non convexe si l'on considère les variables $\boldsymbol{\phi}_l$ et $\boldsymbol{x}_l$ conjointement. Cependant, il peut être découplé en deux sous-problèmes convexes. L'idée est alors d'estimer les filtres du dictionnaire et les cartes de coefficients de manière itérative et alternée : à chaque itération, une première étape consiste à estimer les cartes de coefficients $\boldsymbol{x}_l$ en considérant les filtres $\boldsymbol{\phi}_l$ fixes, et une seconde étape à mettre à jour les filtres avec les cartes de coefficients connues, estimées à l'étape précédente. La première étape revient à résoudre le problème d'estimation parcimonieuse convolutive suivant

$$\widehat{\boldsymbol{x}}_l = \arg\min_{\boldsymbol{x}_l} \frac{1}{2}\sum_t \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_{t,l} - \boldsymbol{y}_t \|_2^2 + \lambda \sum_{t,l} \|\boldsymbol{x}_{t,l}\|_1 \tag{3.10}$$

qui peut être reformulé comme un problème classique d'estimation parcimonieuse en introduisant les matrices de Toeplitz $\boldsymbol{\Phi}_l \in \mathbb{R}^{N\times N}$ telles que $\boldsymbol{\Phi}_l\boldsymbol{x}_l = \boldsymbol{\phi}_l * \boldsymbol{x}_l$ et les matrices $\boldsymbol{\Phi} \in \mathbb{R}^{N\times NL}$, $\boldsymbol{X} \in \mathbb{R}^{NL\times T}$ et $\boldsymbol{Y} \in \mathbb{R}^{N\times T}$

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \dots & \boldsymbol{\Phi}_L \end{bmatrix} \quad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \dots & \boldsymbol{y}_T \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_{1,1} & \boldsymbol{x}_{2,1} & \dots & \boldsymbol{x}_{T,1} \\ \boldsymbol{x}_{1,2} & \boldsymbol{x}_{2,2} & \dots & \boldsymbol{x}_{T,2} \\ \vdots & \vdots & & \\ \boldsymbol{x}_{T,L} & \boldsymbol{x}_{2,L} & \dots & \boldsymbol{x}_{T,L} \end{bmatrix}. \tag{3.11}$$

On se ramène ainsi au problème LASSO

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \frac{1}{2}\|\boldsymbol{\Phi}\boldsymbol{X} - \boldsymbol{Y}\|_2^2 + \lambda\|\boldsymbol{X}\|_1 \tag{3.12}$$

qui peut être résolu à l'aide de l'algorithme ADMM comme détaillé dans la section 2.2.2.2.

La seconde étape cherche à résoudre le problème suivant

$$\widehat{\boldsymbol{\phi}}_l = \arg\min_{\boldsymbol{\phi}_l \in \mathcal{C}} \frac{1}{2} \sum_t \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_{t,l} - \boldsymbol{y}_t \|_2^2 \tag{3.13}$$

qui peut être reformulé de manière non pénalisée après avoir défini la fonction indicatrice $\mathcal{I}_{\mathcal{C}}$

$$\mathcal{I}_{\mathcal{C}}(\boldsymbol{\phi}_l) = \begin{cases} 0 & \text{si } \boldsymbol{\phi}_l \in \mathcal{C} \\ \infty & \text{sinon} \end{cases} \tag{3.14}$$

tel que

$$\widehat{\boldsymbol{\phi}}_l = \arg\min_{\boldsymbol{\phi}_l} \sum_t \| \sum_l \boldsymbol{\phi}_l * \boldsymbol{x}_{t,l} - \boldsymbol{y}_t \|_2^2 + \sum_l \mathcal{I}_{\mathcal{C}}(\boldsymbol{\phi}_l). \tag{3.15}$$

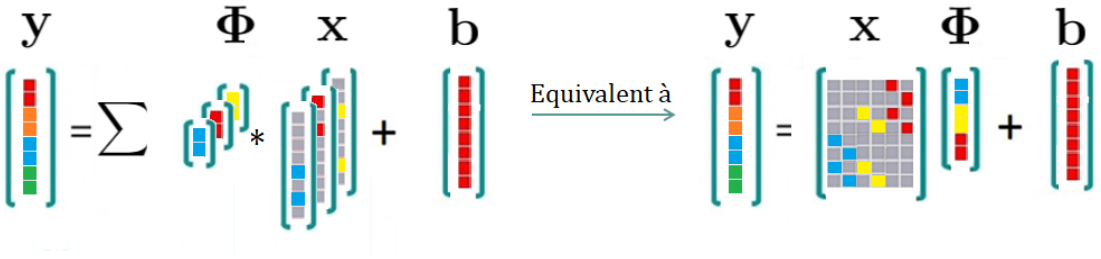

FIGURE 3.4 – Illustration de la reformulation du problème d'estimation parcimonieuse convolutive pour la mise à jour des filtres du dictionnaire.

De même que pour l'estimation des cartes de coefficients, le problème de mise à jour des filtres du dictionnaire peut être reformulé comme un problème de mise à jour des atomes d'un dictionnaire fixe en introduisant les matrices $\boldsymbol{X}_{t,l} \in \mathbb{R}^{N \times M}$ telles que $\boldsymbol{x}_{t,l}\boldsymbol{\phi}_l = \boldsymbol{x}_{t,l} * \boldsymbol{\phi}_l$ et les

matrices $\boldsymbol{X}$, ϕ et $\boldsymbol{y}$ telles que

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{1,1} & \boldsymbol{X}_{1,2} & \dots & \boldsymbol{X}_{1,L} \\ \boldsymbol{X}_{2,1} & \dots & & \\ \dots & \dots & & \\ \boldsymbol{X}_{T,1} & \dots & & \boldsymbol{X}_{T,L} \end{pmatrix} \quad \boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \vdots \\ \boldsymbol{\phi}_L \end{pmatrix} \quad \boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix} \tag{3.16}$$

La reformulation proposée est représentée de manière schématique par la figure 3.4. Le problème (3.15) admet alors une forme simplifiée

$$\widehat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2}\|\boldsymbol{X}\boldsymbol{\phi} - \boldsymbol{y}\|_2^2 + \mathcal{I}_{\mathcal{C}}(\boldsymbol{\phi}) \tag{3.17}$$

La mise à jour des filtres du dictionnaire peut être réalisée à l'aide d'une décomposition en valeurs singulières à l'image du célèbre algorithme K-SVD adapté au problème d'apprentissage de dictionnaires convolutifs dans Yellin et al. (2017), ou par un algorithme de descente de gradient tel que l'algorithme de descente de gradient proximal détaillé dans l'algorithme 4 ou par sa version accélérée détaillée dans l'algorithme 5 Nesterov (1983) où $L > 0$ est un hyperparamètre qui contrôle le pas de la descente de gradient, i désigne l'indice de l'itération, $\text{prox}_{\mathcal{C}}(\boldsymbol{\phi}^{i+1})$ est le projecteur de $\boldsymbol{\phi}^{i+1}$ sur l'ensemble $\mathcal{C}$ et $\nabla_{\boldsymbol{\phi}} f(\boldsymbol{\phi}^i)$ désigne le gradient de la fonction f au point $\boldsymbol{\phi}^i$ avec $f(\boldsymbol{\phi}) = \|\boldsymbol{x}\boldsymbol{\phi} - \boldsymbol{y}\|_2^2$. Une extension de l'algorithme 5 a été proposée afin de résoudre le problème (3.17) dans le domaine fréquentiel et de limiter la complexité de calcul Silva and Rodriguez (2018).

Algorithm 4 Algorithme de descente de gradient proximal

Input $\boldsymbol{X}$, $\boldsymbol{y}$, hyperparamètre L
Init $\boldsymbol{\phi}^0$
i=0
repeat
  $\boldsymbol{\phi}' = \boldsymbol{\phi}^i - \frac{1}{L}\nabla_{\boldsymbol{\phi}} f(\boldsymbol{\phi}^i)$
  $\boldsymbol{\phi}^{i+1} = \text{prox}_{\mathcal{C}}(\boldsymbol{\phi}^{i+1}) = \frac{\boldsymbol{\phi}'}{\max(\|\boldsymbol{\phi}'\|_2, 1)}$
  $i = i + 1$
until $\|\boldsymbol{\phi}^i - \boldsymbol{\phi}^{i+1}\|_2 < \epsilon$

Par ailleurs, l'algorithme ADMM a également été appliqué pour résoudre le problème de la mise à jour des filtres du dictionnaire Garcia-Cardona and Wohlberg (2018) après l'introduction d'une variable auxiliaire $\boldsymbol{d}$ et d'une contrainte d'égalité telles que

$$\widehat{\boldsymbol{\phi}}_l = \arg\min_{\boldsymbol{\phi}_l} \frac{1}{2}\|\boldsymbol{x}\boldsymbol{\phi} - \boldsymbol{y}\|_2^2 + \mathcal{I}_{\mathcal{C}}(\boldsymbol{d}) \text{ s.c } \boldsymbol{d} = \boldsymbol{\phi}. \tag{3.18}$$

La mise à jour des différentes variables est obtenue par minimisation du Lagrangien augmenté dont l'expression est donnée par

$$\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{d}, \boldsymbol{m}, \mu) = \frac{1}{2}\|\boldsymbol{x}\boldsymbol{\phi} - \boldsymbol{y}\|_2^2 + \mathcal{I}_{\mathcal{C}}(\boldsymbol{d}) + \boldsymbol{m}(\boldsymbol{\phi} - \boldsymbol{d}) + \frac{\mu}{2}\|\boldsymbol{\phi} - \boldsymbol{d}\|_2^2. \tag{3.19}$$

Algorithm 5 Algorithme de descente de gradient proximal accéléré

Input $\boldsymbol{X}$, $\boldsymbol{y}$, hyperparamètre L
Init $\boldsymbol{\phi}^0$,γ^0,$\boldsymbol{a}^0 = \boldsymbol{\phi}$
i=0
repeat
  Étape de calcul du Gradient
  $\boldsymbol{\phi}' = \boldsymbol{a}^i - \frac{1}{L}\nabla_{\boldsymbol{a}} f(\boldsymbol{\phi}^i)$
  $\boldsymbol{\phi}^{i+1} = \text{prox}_{\mathcal{C}}(\boldsymbol{\phi}') = \frac{\boldsymbol{\phi}'}{\max(\|\boldsymbol{\phi}'\|_2, 1)}$
  Mise à jour du coefficient momentum
  $\gamma^{i+1} = \frac{1+\sqrt{1+4\gamma^{i\,2}}}{2}$
  Étape du momentum de Nesterov
  $\boldsymbol{a}^{i+1} = \boldsymbol{\phi}^{i+1} + \frac{\gamma^i - 1}{\gamma^{i+1}}(\boldsymbol{\phi}^{i+1} - \boldsymbol{\phi}^i)$
  $i = i + 1$
until $\|\boldsymbol{\phi}^i - \boldsymbol{\phi}^{i+1}\|_2 < \epsilon$

Par exemple, la mise à jour des filtres découle du problème de minimisation suivant

$$\widehat{\boldsymbol{\phi}}^{i+1} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2}\|\boldsymbol{x}^{i+1}\boldsymbol{\phi} - \boldsymbol{y}\|_2^2 + \boldsymbol{m}^i(\boldsymbol{\phi} - \boldsymbol{d}^i) + \frac{\mu^i}{2}\|\boldsymbol{\phi} - \boldsymbol{d}^i\|_2^2, \tag{3.20}$$

qui revient à résoudre le problème linéaire suivant

$$\widehat{\boldsymbol{\phi}}^{i+1} = (\boldsymbol{x}^{i+1}\boldsymbol{x}^{i+1^T} + \mu^i I)^{-1}(\boldsymbol{x}^{i+1^T}\boldsymbol{y} + \boldsymbol{m}^i + \mu^i \boldsymbol{d}^i). \tag{3.21}$$

La résolution de (3.21) a été étudiée dans le domaine fréquentiel dans Silva and Rodriguez (2018) afin de limiter sa complexité de calcul.

3.3 L'algorithme C-ADDICT

Cette partie présente l'algorithme C-ADDICT, une extension de l'algorithme ADDICT basée sur des méthodes d'estimation parcimonieuse convolutive et d'apprentissage de dictionnaires convolutif, qui est par nature un algorithme invariant par translation.

3.3.1 Prétraitement

Les signaux de télémesure reçus sous forme de séries temporelles sont découpés en fenêtres de données multivariées de taille fixe W. Chaque fenêtre représente un intervalle de temps durant lequel les mesures des différents paramètres considérés ont été acquises simultanément. La taille des fenêtres doit être supérieure à la taille des filtres (i.e, $W > M$) qui composent le dictionnaire. La taille des filtres et des fenêtres sont deux paramètres fixés par l'utilisateur et dépendent du contexte d'application. Par exemple, si l'on souhaite surveiller le bon fonctionnement d'un satellite en orbite basse, il peut être opérationellement intéressant de réaliser la

détection d'anomalies tous les jours mais de construire des filtres dont la taille correspond à la durée d'une orbite (environ 100 minutes). Les fenêtres obtenues à l'issue de la segmentation sont des matrices composées d'autant de lignes que de paramètres et d'autant de colonnes que de mesures.

3.3.2 Formulation du problème

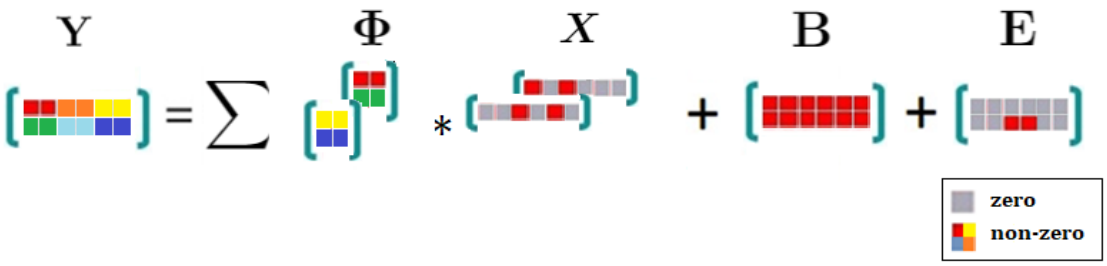

FIGURE 3.5 – Illustration de la stratégie de décomposition C-ADDICT à l'aide d'un dictionnaire convolutif.

À partir d'un dictionnaire $\boldsymbol{\Phi} \in \mathbb{R}^{K \times M \times L}$ composé de L filtres multivariés, notés $\boldsymbol{\Phi}_l \in \mathbb{R}^{K \times M}$, la stratégie C-ADDICT consiste à décomposer chaque nouvelle fenêtre de données multivariées $\boldsymbol{Y} \in \mathbb{R}^{K \times W}$ (avec $W > M$), comme une somme de trois matrices : une matrice nominale issue de la décomposition sur le dictionnaire, une matrice d'anomalie $\boldsymbol{E} \in \mathbb{R}^{K \times W}$ décrivant le caractère atypique des signaux de télémesure étudiés et une matrice de bruit additif $\boldsymbol{B} \in \mathbb{R}^{K \times W}$, i.e.,

$$\boldsymbol{Y} = \sum_l \boldsymbol{\Phi}_l * \boldsymbol{x}_l + \boldsymbol{E} + \boldsymbol{B} \tag{3.22}$$

où $\sum_l \boldsymbol{\Phi}_l * \boldsymbol{x}_l$ représente la matrice nominale associée à $\boldsymbol{Y}$ et les $\boldsymbol{x}_l \in \mathbb{R}^{K \times W}, l = 1, ..., L$ sont les cartes de coefficients parcimonieuses. Notons que ce modèle correspond au modèle ADDICT dans lequel la combinaison linéaire d'éléments du dictionnaire a été remplacée par une somme de produits de convolution. L'intérêt de l'introduction des produits de convolution est qu'ils permettent d'exprimer les signaux de télémesure sains (sans anomalie) comme une superposition d'éléments d'un dictionnaire (appelés ici filtres) de manière invariante par translation. Comme illustré dans la figure 3.5, les données de télémesure sont alors décomposées selon une unique stratégie, et ce, quelque soit leur nature (discrète ou continue). L'estimation conjointe des différents signaux de télémesure qui composent $\boldsymbol{Y}$ revient à résoudre le problème de minimisation suivant

$$\widehat{\boldsymbol{x}}_l, \widehat{\boldsymbol{E}} = \arg\min_{\boldsymbol{x}_l, \boldsymbol{E}} \frac{1}{2} \| \sum_l \boldsymbol{\Phi}_l * \boldsymbol{x}_l + \boldsymbol{E} - \boldsymbol{Y} \|_2^2 + \lambda \sum_l \|\boldsymbol{x}_l\|_1 + \beta \sum_{k,n} \|\boldsymbol{E}_{k,n}\|_2 \tag{3.23}$$

où $\boldsymbol{E}_{k,n} \in \mathbb{R}^M$ est le signal d'anomalie associé à la $n^{ème}$ sous-fenêtre du $k^{ème}$ paramètre de télémesure. De même que dans le modèle ADDICT, le modèle C-ADDICT impose une double contrainte de parcimonie : une sur les cartes de coefficients $\boldsymbol{x}_l$ et une sur la matrice d'anomalies $\boldsymbol{E}$. La première correspond à notre hypothèse de départ selon laquelle des signaux qui ne sont

pas affectés par une anomalie peuvent être correctement estimés par une somme de produits de convolution entre les cartes de coefficients et quelques filtres du dictionnaire associés au fonctionnement normal du satellite. La deuxième contrainte de parcimonie concerne la matrice d'anomalie $\boldsymbol{E}$ et traduit le fait que les anomalies sont rares, qu'elles n'affectent pas tous les paramètres de télémesure en même temps, et qu'elles ne durent pas nécessairement pendant toute la fenêtre temporelle étudiée. Cette parcimonie structurée a donc un double intérêt car elle permet d'identifier automatiquement les paramètres touchés par l'anomalie et de localiser précisément les périodes d'anomalies en estimant un signal d'anomalie par paramètre et par sous-fenêtre dont la taille correspond à celle des filtres. Le problème (3.23) permet de traiter conjointement, par une unique stratégie invariante par translation, les signaux de télémesure continus et discrets. Cependant, selon le nombre de paramètres de télémesure étudiés, il peut être coûteux à résoudre.
Étant donné leur structure, les matrices $\boldsymbol{Y}$, $\boldsymbol{\Phi}$, $\boldsymbol{E}$ et $\boldsymbol{B}$ peuvent être divisées en K sous-matrices où K est le nombre de paramètres de télémesure étudiés. Par ailleurs, rappelons que le dictionnaire multivarié $\boldsymbol{\Phi}$ a été appris de manière à prendre en compte les relations qu'entretiennent les paramètres entre eux. En d'autres termes, le $l^{ème}$ filtre du dictionnaire décrit les comportements des différents paramètres de télémesure observés aux mêmes instants. À partir de ces considérations et afin de réduire la complexité de calcul de (3.23), nous proposons de découpler le problème en K sous-problèmes et de résoudre le problème suivant

$$\widehat{\boldsymbol{x}}_{k,l}, \widehat{\boldsymbol{e}}_k = \arg\min_{\boldsymbol{x}_{k,l}, \boldsymbol{e}_k} \frac{1}{2} \sum_k \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{k,l} + \boldsymbol{e}_k - \boldsymbol{y}_k \|_2^2 + \lambda \sum_{k,l} \|\boldsymbol{x}_{k,l}\|_1 + \beta \sum_{k,n} \|\boldsymbol{e}_{k,n}\|_2$$

$$\text{s.c. } \boldsymbol{x}_{1,l} = \boldsymbol{x}_{2,l} = \cdots = \boldsymbol{x}_{K,l} \; \forall l = 1, ..., L \tag{3.24}$$

où $\boldsymbol{y}_k \in \mathbb{R}^W$ est le signal test associé au $k^{ème}$ paramètre de télémesnure, $\boldsymbol{e}_k \in \mathbb{R}^W$ le signal d'anomalie associé, $\boldsymbol{\phi}_k \in \mathbb{R}^{M \times L}$ le dictionnaire convolutif pour ce $k^{ème}$ paramètre et $\boldsymbol{x}_l \in \mathbb{R}^{W \times L}$ les cartes de coefficients associées au $k^{ème}$ paramètre de télémesure. Comme présenté dans (3.25), les problèmes (3.23) et (3.24) sont équivalents. Notons que la contrainte dans (3.24) est très importante car elle permet de conserver les liens entres les paramètres et donc d'assurer la détection des anomalies multivariées. Le problème d'estimation parcimonieuse convolutive découplé (3.24) est inspiré du problème consensus présenté dans Boyd et al. (2011), Chap 7. Ce dernier a également été étudié dans Garcia-Cardona and Wohlberg (2018) pour un problème similaire d'apprentissage de dictionnaire convolutif. Les deux approches seront appliquées à des données satellite réelles et comparées dans une prochaine section.

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} \boldsymbol{x} \xrightarrow{\text{équivalent à}} \begin{bmatrix} \boldsymbol{y}_1 = \Phi_1 \boldsymbol{x}_1 \\ \boldsymbol{y}_2 = \Phi_2 \boldsymbol{x}_2 \end{bmatrix} \text{ s.c. } \boldsymbol{x}_1 = \boldsymbol{x}_2. \tag{3.25}$$

3.3.3 Résolution du problème

Le problème (3.24) peut être reformulé comme un problème d'estimation parcimonieuse avec un dictionnaire fixe en introduisant les matrices de Toeplitz $\boldsymbol{\Phi}_{k,l} \in \mathbb{R}^{W \times W}$ telles que

$\boldsymbol{\Phi}_{k,l}\boldsymbol{x}_{k,l} = \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{k,l}$ et les deux matrices

$$\boldsymbol{\Phi}_k = \begin{pmatrix} \boldsymbol{\Phi}_{k,1} & \boldsymbol{\Phi}_{k,2} & \dots & \boldsymbol{\Phi}_{k,L} \end{pmatrix}, \; \boldsymbol{x}_k = \begin{pmatrix} \boldsymbol{x}_{k,1} \\ \boldsymbol{x}_{k,2} \\ \dots \\ \boldsymbol{x}_{k,L} \end{pmatrix} \tag{3.26}$$

Le problème (3.24) devient :

$$\widehat{\boldsymbol{x}}_k, \widehat{\boldsymbol{e}}_k = \arg\min_{\boldsymbol{x}_k, \boldsymbol{e}_k} \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{e}_k - \boldsymbol{y}_k\|_2^2 + \lambda \sum_k \|\boldsymbol{x}_k\|_1 + \beta \sum_{k,n} \|\boldsymbol{e}_{k,n}\|_2$$

$$\text{s.c. } \boldsymbol{x}_1 = \boldsymbol{x}_2 = \dots = \boldsymbol{x}_K \tag{3.27}$$

Le nouveau problème (3.27) peut être résolu à l'aide de l'algorithme ADMM après avoir introduit une variable auxiliaire $\boldsymbol{z}$ et une contrainte d'égalité telles que

$$\widehat{\boldsymbol{x}}_k, \widehat{\boldsymbol{e}}_k, \widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{x}_k, \boldsymbol{e}_k} \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{e}_k - \boldsymbol{y}_k\|_2^2 + \lambda \|\boldsymbol{z}\|_1 + \beta \sum_{k,m} \|\boldsymbol{e}_{k,m}\|_2$$

$$\text{s.c. } \boldsymbol{x}_k = \boldsymbol{z}, \; \forall k. \tag{3.28}$$

Le Lagrangien augmenté associé au problème précédent est donné par

$$\mathcal{L}_A(\boldsymbol{x}_k, \boldsymbol{z}, \boldsymbol{e}_k, \boldsymbol{m}_k, \boldsymbol{\mu}) = \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{e}_k - \boldsymbol{y}_k\|_2^2 + \lambda \|\boldsymbol{z}\|_1$$

$$+\beta \sum_{k,m} \|\boldsymbol{e}_{k,m}\|_2 + \sum_k \boldsymbol{m}_k^T (\boldsymbol{z} - \boldsymbol{x}_k) + \frac{\mu}{2} \sum_k \|\boldsymbol{z} - \boldsymbol{x}_k\|_2^2 \tag{3.29}$$

où $\boldsymbol{m}_k$ est un vecteur contenant les multiplicateurs de Lagrange et μ est un paramètre de régularisation fixant le niveau de pénalité de l'écart entre $\boldsymbol{z}$ et $\boldsymbol{x}_k$ par rapport à la contrainte d'égalité. Les mises à jour des composantes $\boldsymbol{x}_k, \boldsymbol{z}$, $\boldsymbol{e}_k$ et $\boldsymbol{m}_k$ à l'itération $(i+1)$ sont détaillées ci-dessous

- Mise à jour de $\boldsymbol{x}_k$:

La mise à jour de $\boldsymbol{x}_k$ découle du problème de minimisation suivant

$$\boldsymbol{x}_k^{i+1} = \arg\min_{\boldsymbol{x}_k} \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{e}_k^i - \boldsymbol{y}_k\|_2^2 + \sum_k \boldsymbol{m}_k^{i\,T} (\boldsymbol{z}^i - \boldsymbol{x}_k) + \frac{\mu}{2} \sum_k \|\boldsymbol{z}^i - \boldsymbol{x}_k\|_2^2. \tag{3.30}$$

Le problème (3.30) peut être découplé en K sous-problèmes permettant d'estimer les cartes de coefficients $\boldsymbol{x}_k, k = 1, ..., K$ de manière indépendante tel que

$$\boldsymbol{x}_k^{i+1} = \arg\min_{\boldsymbol{x}_k} \frac{1}{2} \|\boldsymbol{\Phi}_k \boldsymbol{x}_k + \boldsymbol{e}_k^i - \boldsymbol{y}_k\|_2^2 + \boldsymbol{m}_k^{i\,T} (\boldsymbol{z}^i - \boldsymbol{x}_k) + \frac{\mu}{2} \|\boldsymbol{z}^i - \boldsymbol{x}_k\|_2^2. \tag{3.31}$$

L'estimation des cartes de coefficients est alors donnée par la solution du problème suivant

$$\boldsymbol{x}_k^{i+1} = (\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k + \rho \boldsymbol{I})[\boldsymbol{\Phi}_k^T \boldsymbol{y}_k + \mu^i (\boldsymbol{z}^i - \boldsymbol{m}_k^i)]. \tag{3.32}$$

qui peut être résolu dans le domaine fréquentiel Garcia-Cardona and Wohlberg (2018) afin de limiter la complexité de calcul liée à la taille de la matrice $\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k$.

- Mise à jour de $\boldsymbol{e}_k$:

La mise à jour des signaux d'anomalie $\boldsymbol{e}_k$ consiste à résoudre le problème de minimisation suivant

$$\boldsymbol{e}_k^{i+1} = \arg\min_{\boldsymbol{e}_{k,p}} \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{x}_k^{i+1} + \boldsymbol{e}_k - \boldsymbol{y}_k\|_2^2 + \beta \sum_{k,n} \|\boldsymbol{e}_{k,n}\|_2. \tag{3.33}$$

La solution du problème (3.33) est donnée par l'opérateur de seuillage par groupe T_β tel que $\widehat{\boldsymbol{e}}_k^{i+1} = T_\beta(\boldsymbol{h}_k^{i+1})$ avec $\boldsymbol{h}_k^{i+1} = \boldsymbol{y}_k - \boldsymbol{\Phi}_k \boldsymbol{x}_k^{i+1}$. La définition de l'opérateur de seuillage par groupe est rappelée ci-dessous

$$[T_\beta(\boldsymbol{h})]_n = \begin{cases} \frac{\|\boldsymbol{h}_n\|_2 - \beta}{\|\boldsymbol{h}_n\|_2} \boldsymbol{h}_n & \text{si} \quad \|\boldsymbol{h}_n\|_2 > \beta \\ 0 & \text{sinon} \end{cases} \tag{3.34}$$

où $\boldsymbol{h}_n \in \mathbb{R}^M$ correspond au $n^{\text{ème}}$ sous-signal de $\boldsymbol{h}$.

- Mise à jour de $\boldsymbol{z}$:

La mise à jour de $\boldsymbol{z}$ consiste à résoudre le problème de minimisation suivant

$$\widehat{\boldsymbol{z}}^{i+1} = \arg\min_{\boldsymbol{z}} \lambda \|\boldsymbol{z}\|_1 + \sum_k \boldsymbol{m}_k^{i\,T} (\boldsymbol{z} - \boldsymbol{x}_k^{i+1}) + \frac{\mu}{2} \sum_k \|\boldsymbol{z} - \boldsymbol{x}_k^{i+1}\|_2^2 \tag{3.35}$$

dont la solution est donnée par le seuillage doux par élément tel que

$$\widehat{\boldsymbol{z}}^{i+1} = S_\gamma \left(\frac{1}{K} \sum_k \boldsymbol{g}_k^{i+1} \right) \tag{3.36}$$

avec $\boldsymbol{g}_k^{i+1} = \boldsymbol{x}_k^{i+1} - \frac{1}{\mu}{}^i \boldsymbol{m}_k^{i+1}$ et $\gamma^i = \frac{\lambda}{\mu^i}$. La définition de l'opérateur de seuillage doux par élément est rappelée ci-dessous

$$[S_\gamma^{i+1}(\boldsymbol{g}^{i+1})]_j = \begin{cases} \boldsymbol{g}_j^{i+1} - \gamma^{i+1} & \text{si } \boldsymbol{g}_j^{i+1} > \gamma^{i+1} \\ 0 & \text{si } |\boldsymbol{g}_j^{i+1}| \leq \gamma^{i+1} \\ \boldsymbol{g}_j^{i+1} + \gamma^{i+1} & \text{si } \boldsymbol{g}_j^{i+1} < -\gamma^{i+1} \end{cases} \tag{3.37}$$

où $\boldsymbol{g}_j$ est la $j^{\text{ème}}$ composante de $\boldsymbol{g}$.

- Mise à jour de $\boldsymbol{m}_k$:

La mise à jour de $\boldsymbol{m}_k$ est donnée par

$$\boldsymbol{m}_k^{i+1} = \boldsymbol{m}_k^i + \boldsymbol{x}_k^{i+1} - \boldsymbol{z}^{i+1} \tag{3.38}$$

3.3.4 Détection d'anomalies

De manière opérationnelle, la détection d'anomalies est réalisée à partir d'un score d'anomalie, noté $s(\boldsymbol{Y}_n)$ est défini à partir des signaux d'anomalies $\widehat{\boldsymbol{E}}_n$ tel que

$$s(\boldsymbol{Y}_n) = \|\widehat{\boldsymbol{E}}_n\|_F^2. \tag{3.39}$$

où $\boldsymbol{Y}_n$ est la $n^{\text{ème}}$ sous-fenêtre de $\boldsymbol{Y}$. La règle de détection d'anomalie proposée est définie par

$$\text{Anomalie détectée sur la période } n \text{ si } s(\boldsymbol{Y}_n) > S_{\text{PFA}} \tag{3.40}$$

où S_{PFA} est un seuil fixé par l'utilisateur. Ce seuil peut être déterminé à l'aide de caractéristiques opérationnelles du récepteur (CORs) si une vérité terrain est disponible, ou par l'expert de manière à respecter certaines contraintes relatives aux non-détections ou aux fausses alarmes. L'algorithme C-ADDICT est résumé dans l'algorithme 6. Notons que, la détection d'anomalies peut être réalisée paramètre par paramètre à partir des signaux d'anomalies $\widehat{\boldsymbol{e}}_{n,k}, k = 1, ..., K$ de manière à pouvoir identifier les paramètres affectés par l'anomalie.

Algorithm 6 $\widehat{\boldsymbol{x}}_k, \widehat{\boldsymbol{e}}_k$=C-ADDICT$(\boldsymbol{y}_k, \boldsymbol{\Phi}_k)$, $k = 1, ..., K$

Initialiser i=1, $\boldsymbol{z}^0, \boldsymbol{e}_k^0, \boldsymbol{m}_k^0, \mu^0, \rho, \lambda, \beta$
Estimation avec ADMM
repeat
$\boldsymbol{x}_k^{i+1} = (\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k + \rho \boldsymbol{I})[\boldsymbol{\Phi}_k^T \boldsymbol{y}_k + \mu^i(\boldsymbol{z}^i - \boldsymbol{m}_k^i)]$
$\boldsymbol{z}^{i+1} = S_{\gamma^{i+1}}(\frac{1}{K}\sum_k \boldsymbol{g}_k^{i+1})$, $\boldsymbol{g}_k^{i+1} = \boldsymbol{x}_k^{i+1} - \frac{1}{\mu^i}\boldsymbol{m}_k^i$, $\gamma^{i+1} = \frac{\lambda}{\mu^i}$
$\boldsymbol{e}_k^{i+1} = T_\beta(\boldsymbol{y}_k - \boldsymbol{\Phi}_k \boldsymbol{x}_k^{i+1})$
$\boldsymbol{m}_k^{i+1} = \boldsymbol{m}_k^i + \boldsymbol{x}_k^{i+1} - \boldsymbol{z}^{i+1}$
$\mu^{i+1} = \rho\mu^i$
i = i+1
until critère d'arrêt
Détection d'anomalie
Une anomalie est détectée dans la sous-fenêtre n du paramètre k si $\|\widehat{\boldsymbol{E}}_n\|_2^2 > S_{\text{PFA}}$

3.3.5 Apprentissage du dictionnaire C-ADDICT

Le problème d'estimation parcimonieuse convolutive C-ADDICT pour la détection d'anomalies multivariées a pu être étendu au problème d'apprentissage de dictionnaires convolutifs de données mixtes prenant en compte les relations et corrélations entre les différents paramètres. À partir d'un ensemble de T signaux sains acquis pour chaque paramètre aux mêmes instants et noté $\boldsymbol{y}_{t,k}, t = 1, ..., T,\ k = 1, ..., K$, l'apprentissage du dictionnaire consiste à résoudre le problème de minimisation suivant

$$\widehat{\boldsymbol{x}}_{k,l}, \widehat{\boldsymbol{\phi}}_{k,l} = \arg \min_{\boldsymbol{x}_{k,l}, \boldsymbol{\phi}_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_{t,k} \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{t,k,l} - \boldsymbol{y}_{t,k} \|_2^2 + \lambda \sum_{t,k,l} \|\boldsymbol{x}_{t,k,l}\|_1$$

$$\text{s.c. } \boldsymbol{x}_{t,1,l} = \boldsymbol{x}_{t,2,l} = \cdots = \boldsymbol{x}_{t,K,l}, \forall t, l \tag{3.41}$$

dans lequel les signaux d'anomalies sont absents (i.e, $\boldsymbol{e}_k = 0$). Le problème (3.41) n'est pas convexe si l'on considère conjointement les cartes de coefficients $\boldsymbol{x}_{t,k,l}$ et les filtres $\boldsymbol{\phi}_{t,k,l}$. Il est alors classique de le découpler en deux problèmes convexes qui peuvent être résolus de manière itérative et alternée. À chaque itération, une première étape permet d'estimer les cartes de coefficients en considérant les filtres fixes, et une deuxième étape met à jour les filtres avec les cartes de coefficients fixes estimées à l'étape précédente. L'estimation des cartes de coefficients revient à résoudre le problème suivant

$$\widehat{\boldsymbol{x}}_{t,k,l} = \arg \min_{\boldsymbol{x}_{t,k,l}} \frac{1}{2} \sum_{t,k} \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{t,k,l} - \boldsymbol{y}_{t,k} \|_2^2 + \lambda \sum_{t,k} \|\boldsymbol{x}_{t,k}\|_1$$

$$\text{s.c. } \boldsymbol{x}_{t,1,l} = \boldsymbol{x}_{t,2,l} = \cdots = \boldsymbol{x}_{t,K,l}, \forall t, l \tag{3.42}$$

qui peut être reformulé comme un problème d'estimation parcimonieuse avec un dictionnaire fixe en introduisant les matrices de Toeplitz $\boldsymbol{\Phi}_{k,l} \in \mathbb{R}^{M \times M}$ telles que $\boldsymbol{\Phi}_{k,l}\boldsymbol{x}_{t,k,l} = \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{t,k,l}$ et les matrices $\boldsymbol{\Phi}_k \in \mathbb{R}^{W \times WL}$, $\boldsymbol{x}_k \in \mathbb{R}^{WL \times T}$ et $\boldsymbol{y}_k \in \mathbb{R}^{WT}$

$$\boldsymbol{\Phi}_k = \begin{pmatrix} \boldsymbol{\Phi}_{k,1} & \boldsymbol{\Phi}_{k,2} & \dots & \boldsymbol{\Phi}_{k,L} \end{pmatrix} \quad \boldsymbol{y}_k = \begin{pmatrix} \boldsymbol{y}_{1,k} \\ \boldsymbol{y}_{2,k} \\ \dots \\ \boldsymbol{y}_{T,k} \end{pmatrix} \boldsymbol{X}_k = \begin{pmatrix} \boldsymbol{x}_{1,k,1} & \boldsymbol{x}_{2,k,1} & \dots & \boldsymbol{x}_{T,k,1} \\ \boldsymbol{x}_{1,k,2} & \boldsymbol{x}_{2,k,2} & \dots & \boldsymbol{x}_{T,k,2} \\ \vdots & \vdots & & \\ \boldsymbol{x}_{1,k,L} & \boldsymbol{x}_{2,k,L} & \dots & \boldsymbol{x}_{T,k,L} \end{pmatrix} \tag{3.43}$$

Le problème (3.42) devient

$$\widehat{\boldsymbol{X}}_k = \arg\min_{\boldsymbol{x}_k} \frac{1}{2} \sum_k \|\boldsymbol{\Phi}_k \boldsymbol{X}_k - \boldsymbol{y}_k\|_2^2 + \lambda \sum_k \|\boldsymbol{X}_k\|_1$$

$$\text{s.c. } \boldsymbol{X}_1 = \boldsymbol{X}_2 = \dots = \boldsymbol{X}_K, \tag{3.44}$$

qui correspond au problème d'estimation parcimonieuse convolutive de l'algorithme C-ADDICT pour lequel on ne considère pas les signaux d'anomalies (i.e, $\boldsymbol{e}_k = 0, k = 1, ..., K$). Il peut être résolu à l'aide de l'algorithme ADMM comme détaillé dans la section 3.3.3.

La mise à jour des filtres du dictionnaire avec les cartes de coefficients connues consiste à résoudre le problème de minimisation suivant

$$\widehat{\boldsymbol{\phi}}_{k,l} = \arg\min_{\boldsymbol{\phi}_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_{t,k} \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{t,k,l} - \boldsymbol{y}_{t,k}\|_2^2 \tag{3.45}$$

qui peut être découplé en K sous problèmes afin de mettre à jour les filtres associés aux différents paramètres de télémesure de manière indépendante à l'aide des cartes de coefficients estimées à l'étape précédente. La mise à jour à l'itération $(i+1)$ des filtres associés au paramètre k revient à résoudre le problème suivant

$$\widehat{\boldsymbol{\phi}}_{k,l}^{i+1} = \arg\min_{\boldsymbol{\phi}_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_t \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{t,k,l}^{i+1} - \boldsymbol{y}_t\|_2^2. \tag{3.46}$$

Nous nous sommes ramenés à un problème classique de mise à jour des filtres d'un dictionnaire convolutif. Comme présenté précédemment (voir section 3.2.2), il existe plusieurs méthodes pour le résoudre. Dans ces travaux, nous avons choisi d'utiliser l'algorithme de descente de gradient proximal accéléré détaillé dans l'algorithme 5.

3.3.6 Convergence de l'algorithme proposé

La convergence de l'algorithme ADMM utilisé pour la détection C-ADDICT est assurée si le Lagrangien augmenté à minimiser est convexe. Cette condition est satisfaite étant donné que (3.29) est une somme de fonctions convexes positives Boyd et al. (2011); Eckstein

and Bertsekas (1992). Pour ce qui est de l'algorithme d'apprentissage d'un dictionnaire C-ADDICT, la convergence à chaque itération de l'algorithme de descente de gradient proximal pour la mise à jour des filtres du dictionnaire et de l'algorithme ADMM pour la mise à jour des cartes de coefficients est assurée pour les même raisons que celles évoquées précédemment. La convergence de l'algorithme d'apprentissage de dictionnaire convolutif est donc garantie.

3.4 Résultats Expérimentaux

3.4.1 Données de télémesure satellite réelles

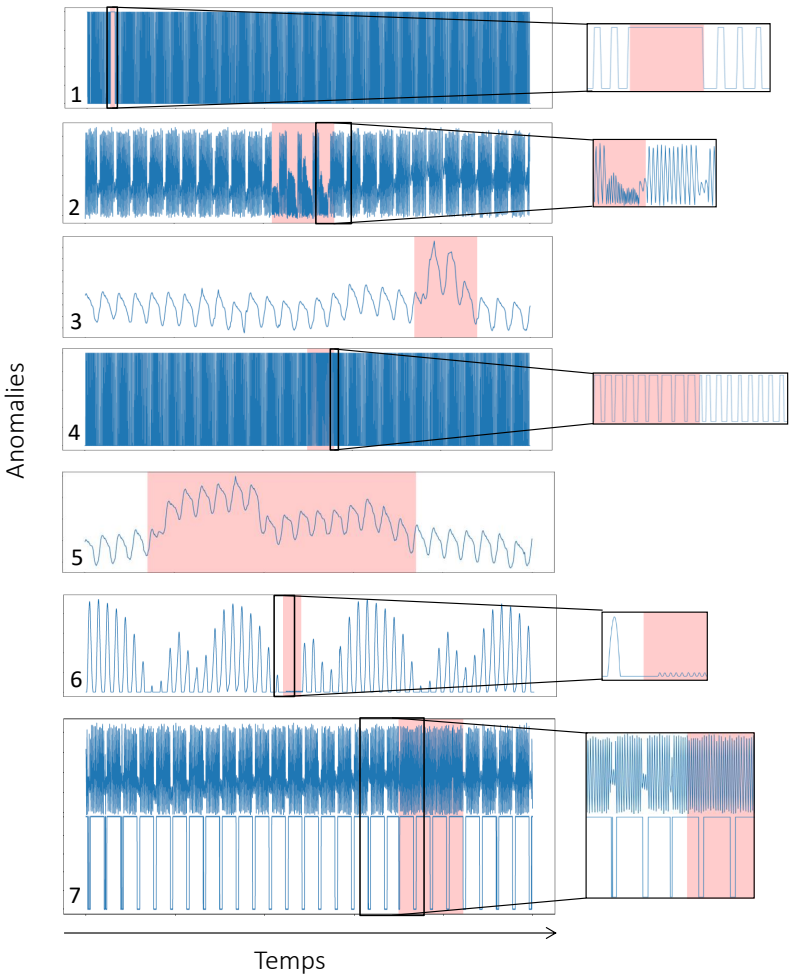

FIGURE 3.6 – Périodes d'anomalies de la vérité terrain

Dans ces travaux, un dictionnaire convolutif a été appris à l'aide de l'algorithme d'apprentissage de dictionnaire C-ADDICT à partir de 3 mois de données de télémesure réelles décrivant le fonctionnement normal d'un satellite à travers 10 paramètres de télémesure dont 3 discrets et 7 continus. Afin de pouvoir comparer les performances de l'algorithme C-ADDICT à celles de ADDICT, la taille des filtres correspond à celle des atomes construits dans nos précédentes expérimentations, i.e, $M = 50$. Bien que le modèle C-ADDICT soit invariant par translation, les signaux de test sont estimés par superposition de filtres plus ou moins décalés les uns par rapport aux autres. De même que pour l'algorithme ADDICT, cela implique qu'une quantité suffisante de filtres est nécessaire pour assurer la qualité de la représentation des signaux sains et ainsi limiter le nombre de fausses alarmes. Pour ces premières expérimentations, nous avons fait le choix de construire un dictionnaire composé de $L = 100$ filtres. Ce choix représente un compromis entre complexité de calcul et qualité d'estimation. Un extrait des signaux d'apprentissage et des exemples de filtres appris par notre algorithme sont représentés dans la figure 3.7. L'algorithme a été capable de tenir compte de la nature des données et de capturer les corrélations et relations entre les différents paramètres de télémesure pour construire des filtres mixtes fidèles aux données.
L'évaluation de l'algorithme C-ADDICT a été réalisée à partir de données issues des 10 mêmes paramètres de télémesure que ceux utilisés pour apprendre le dictionnaire. Les signaux de cette base de test représente environ 18 jours de télémesure durant lesquels 7 périodes d'anomalies connues et identifiées par des experts ont été observées. Les périodes d'anomalies que nous étudions dans ces travaux sont représentées dans les encadrés rouges de la figure 2.6 qui sont rappelés ci-dessus pour faciliter la lecture.

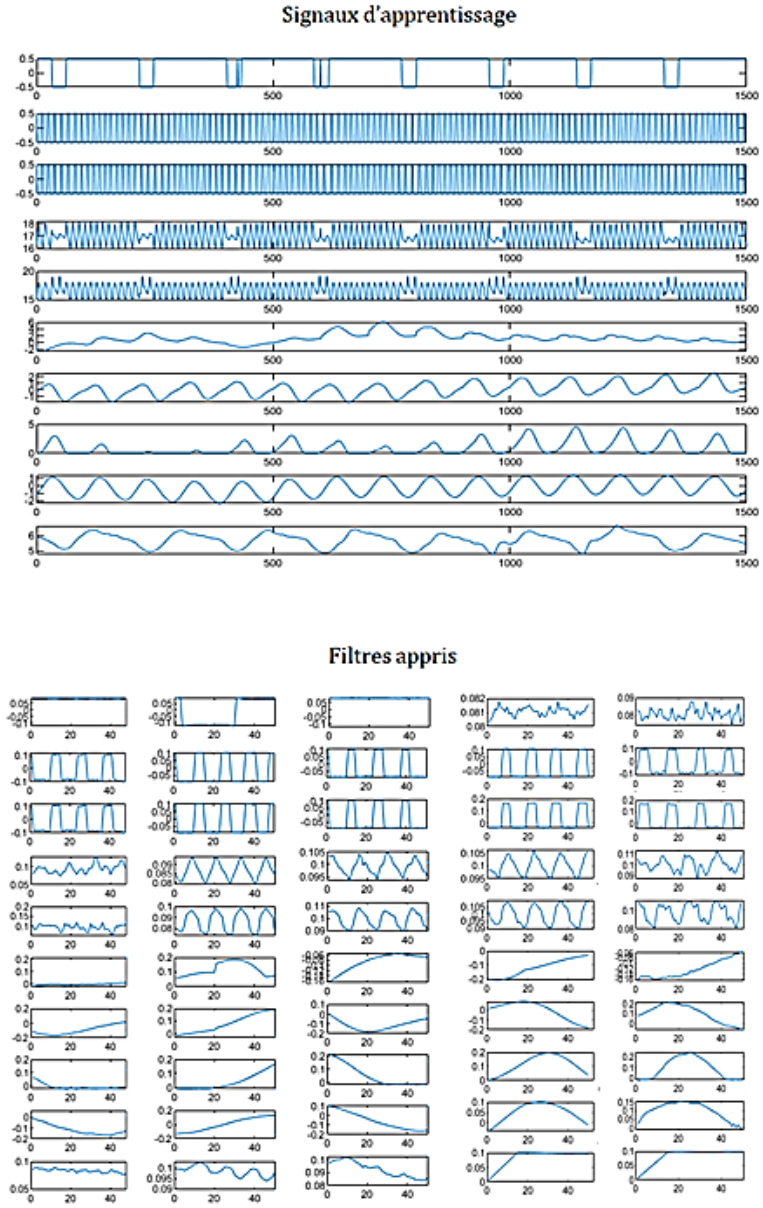

FIGURE 3.7 – Extraits de signaux d'apprentissage associés à un fonctionnement normal du satellite et représentation de cinq filtres du dictionnaire convolutif appris à l'aide des signaux d'apprentissage.

3.4.2 Comparaison de modèle

Cette section revient sur la définition du modèle C-ADDICT pour la détection d'anomalies multivariées et compare les performances de détection et de calcul du problème non contraint

$$\widehat{\boldsymbol{x}}_l, \widehat{\boldsymbol{E}} = \arg\min_{\boldsymbol{x}_l, \boldsymbol{E}} \frac{1}{2} \| \sum_l \boldsymbol{\Phi}_l * \boldsymbol{x}_l + \boldsymbol{E} - \boldsymbol{Y} \|_2^2 + \lambda \sum_l \|\boldsymbol{x}_l\|_1 + \beta \sum_{k,m} \|\boldsymbol{E}_{k,n}\|_2 \tag{3.47}$$

et de sa version découplée

$$\widehat{\boldsymbol{x}}_{k,l}, \widehat{\boldsymbol{e}}_k = \arg\min_{\boldsymbol{x}_{k,l}, \boldsymbol{e}_k} \frac{1}{2} \sum_k \| \sum_l \boldsymbol{\phi}_{k,l} * \boldsymbol{x}_{k,l} + \boldsymbol{e}_k - \boldsymbol{y}_k \|_2^2 + \lambda \sum_{k,l} \|\boldsymbol{x}_{k,l}\|_1 + \beta \sum_{k,n} \|\boldsymbol{e}_{k,n}\|_2 \quad (3.48)$$

$$\text{s.c. } \boldsymbol{x}_{1,l} = \boldsymbol{x}_{2,l} = \cdots = \boldsymbol{x}_{K,l} \; \forall l = 1, ..., L \quad (3.49)$$

où $\boldsymbol{Y} \in \mathbb{R}^{K \times W}$ est une matrice de signaux test telle que $\boldsymbol{y}_k \in \mathbb{R}^W$ est le signal test du $k^{\text{ème}}$ paramètre, $\boldsymbol{E} \in \mathbb{R}^{K \times W}$ est une matrice d'anomalies telle que $\boldsymbol{e}_k \in \mathbb{R}^W$ est le signal d'anomalie associé au $k^{\text{ème}}$ paramètre, $\boldsymbol{\Phi} \in \mathbb{R}^{K \times M \times L}$ est un dictionnaire convolutif composé de L filtres mixtes, notés $\Phi_l \in \mathbb{R}^{K \times M}$, tels que $\boldsymbol{\phi}_{k,l} \in \mathbb{R}^M$ est le $l^{\text{ème}}$ filtre associé au $k^{\text{ème}}$ paramètre, $\boldsymbol{x}_l \in \mathbb{R}^W$ est la $l^{\text{ème}}$ carte de coefficients associée au $l^{\text{ème}}$ filtre, $\boldsymbol{x}_{k,l}$ est la $l^{\text{ème}}$ carte de coefficients univariée associée au $k^{\text{ème}}$ paramètre et λ et β sont deux paramètres de régularisation contrôlant respectivement la parcimonie de $\boldsymbol{x}_l$ ou $\boldsymbol{x}_{k,l}$ et de $\boldsymbol{E}$ ou $\boldsymbol{e}_k$. D'un point de vue théorique, le problème (3.48) est équivalent au problème (3.47). Rappelons que (3.47) a simplement été découplé en K sous problèmes plus faciles à résoudre et que la contrainte (3.49) a été ajoutée afin d'assurer la préservation des liens entres les paramètres et la détection des anomalies multivariées.

Nous avons évalué ces deux approches sur nos données de télémesure satellite. En termes de temps de calcul, le modèle découplé s'est montré plus performant pour surveiller 10 paramètres de télémesure. En effet, la détection sur une journée de télémesure a été réalisée en 58 secondes pour le modèle contraint contre 369 secondes pour le modèle non contraint (voir tableau 2.1). En terme de performances de détection, les scores retournés par les deux approches et représentés dans les figures 3.8 et 3.9 témoignent de l'intérêt du modèle découplé qui retourne des scores d'anomalies positifs presque exclusivement en période d'anomalie (fonds rouges) et les detecte presque en totalité. Le modèle non contraint ne détecte que partiellement les périodes d'anomalie, comme c'est le cas de l'anomalie #5 par exemple, et retourne plus de fausses d'alarmes. Découpler le problème en introduisant une contrainte adaptée a alors permis : 1) de limiter la complexité de calcul, et 2) d'améliorer les performances de détection de la méthode. Ce dernier point s'explique en partie par le fait que l'égalité imposée aux cartes de coefficients des différents paramètres de télémesure est en réalité rarement parfaite. Cette souplesse assure une bonne estimation des signaux sains et la détection des anomalies, notamment multivariées.

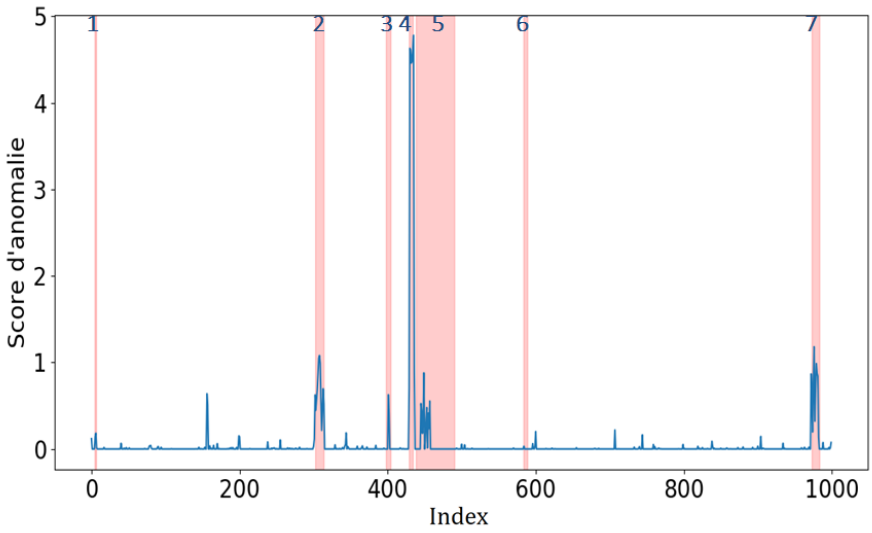

FIGURE 3.8 – Scores d'anomalie obtenus par minimisation du problème non contraint d'estimation parcimonieuse convolutive pour la détection d'anomalies dans des données mixtes (3.47). La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

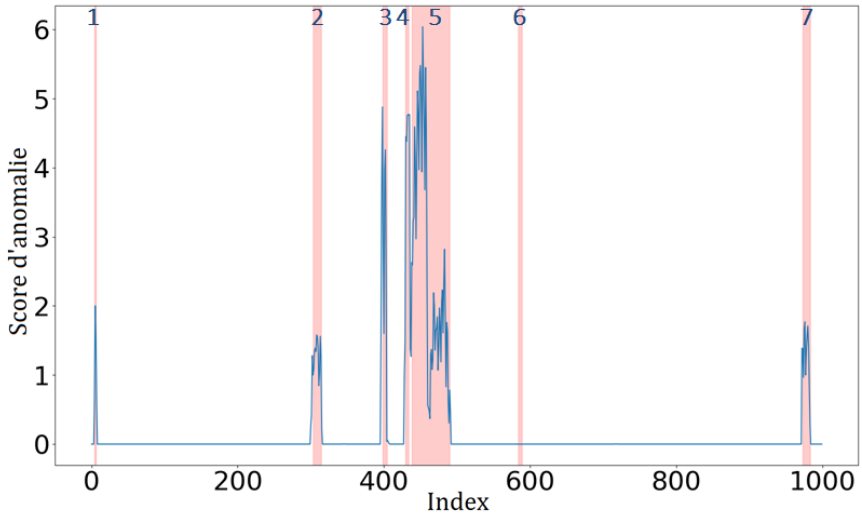

FIGURE 3.9 – Scores d'anomalie obtenus par minimisation du problème découplé d'estimation parcimonieuse convolutive pour la détection d'anomalies dans des données mixtes (3.48). La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1

Méthode	Seuil	P_{D}	P_{FA}	Temps (en secondes)
Classique(3.47)	$\epsilon > 0$	47.9%	9.2%	369’
Découplé (3.48) (3.49)	$\epsilon > 0$	94.7%	1.7%	58’

TABLE 3.1 – Résultats de détection en termes de probabilités de détection (P_{D}) et de probabilités de fausses alarmes (P_{FA}) obtenues par minimisation du problème non contraint d’estimation parcimonieuse convolutive et du problème contraint.

3.4.3 Comparaison à l’état de l’art

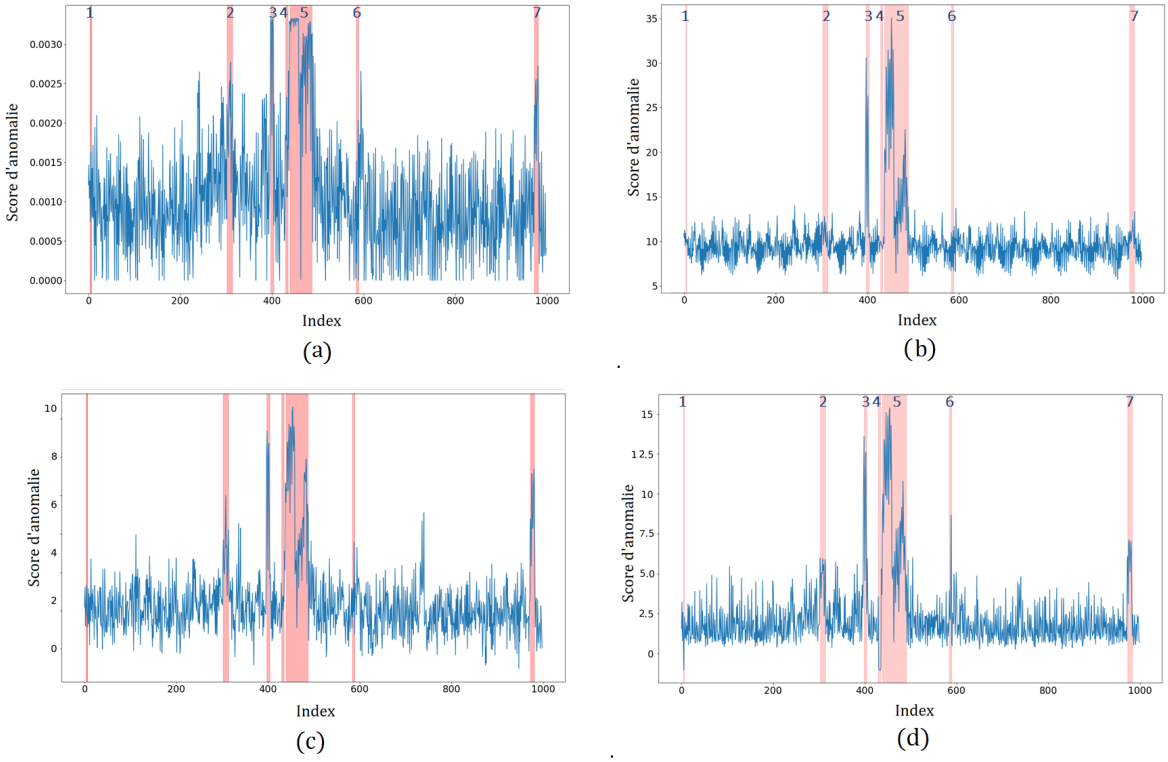

FIGURE 3.10 – Comparaison de scores d’anomalie retournés pour les différents signaux tests de la base d’anomalies par les algorithmes OC-SVM (a), MPCCAD (b), ADDICT (c) et W-ADDICT pour lesquels l’option d’invariance par translation est active avec $\tau_{\max} = 5$. La vérité terrain est représentée par les fonds rouges. Chaque période d’anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1.

Les performances de détection enregistrées pour l’algorithme C-ADDICT sur notre base de test ont été comparées à celles de quatre méthodes que nous avons déjà eu l’occasion d’étudier : l’algorithme OC-SVM appliqué à des signaux mixtes obtenus à l’aide du prétraitement ADDICT (voir section 2.3.1), l’algorithme MPPCAD, l’algorithme ADDICT et sa version pondérée W-ADDICT dont les poids ont été construits à partir des coefficients de corrélation calculés pour chaque paramètre entre le signal testé et sa reconstruction obtenue par décom-

position sur un dictionnaire fixe (voir section 2.4.2). Au regard des scores représentés dans les figures 3.9 (C-ADDICT) et 3.10 (état de l'art), il apparaît que l'algorithme C-ADDICT offre une détection plus efficace des anomalies de notre vérité terrain que les autres méthodes de l'état de l'art. En effet on observe que les scores d'anomalie retournés par C-ADDICT sont positifs uniquement en présence d'anomalies. Par ailleurs, le seuil de détection est plus facile à régler pour la méthode C-ADDICT. Notons que dans ces expérimentations, les hyperparamètres des différentes méthodes ont été déterminés de manière optimale pour ces données à l'aide de la vérité terrain. Un inconvénient de l'algorithme C-ADDICT est que l'anomalie #6, bien détectée par les algorithmes ADDICT et W-ADDICT n'a pas été détectée par C-ADDICT. L'étude des courbes CORs obtenues pour chacune des méthodes à partir de leurs résultats de détection sur les données test confirment ces premières conclusions. En effet, l'algorithme C-ADDICT détecte presque 95% des anomalies pour moins de 2% de fausses alarmes qui correspondent souvent aux fenêtres saines situées juste avant ou juste après les périodes d'anomalies.

Ces résultats encourageants obtenus pour l'algorithme C-ADDICT s'expliquent en grande partie par le fait que la méthode proposée est invariante par translation. Les motifs associés à un fonctionnement normal du satellite ont été appris dans le dictionnaire et tout nouveau signal sain peut alors être estimé par activation des filtres à l'instant exact où les motifs correspondants sont observés dans le signal d'entrée. Nous avons pu observer ce phénomène sur nos données en étudiant les cartes de coefficients représentées dans la figure 3.12 qui sont associées aux dix premiers filtres de notre dictionnaire pour l'estimation d'une journée de télémesure. On constate que les cartes de coefficients sont très parcimonieuses et qu'il est possible de déterminer à quel moment précis les différents filtres ont été activés. En effet, les cartes de coefficients sont composées de très peu de valeurs non nulles (traits de couleurs). La position de ces valeurs dans les cartes indique à quel moment les différents filtres sont observés dans le signal étudié. Notons que les données étudiées ici se composent de peu de motifs différents. L'apprentissage du dictionnaire et l'estimation des signaux sains est donc plus simple. L'efficacité de la méthode sur des données plus complexes mériterait d'être étudiée.

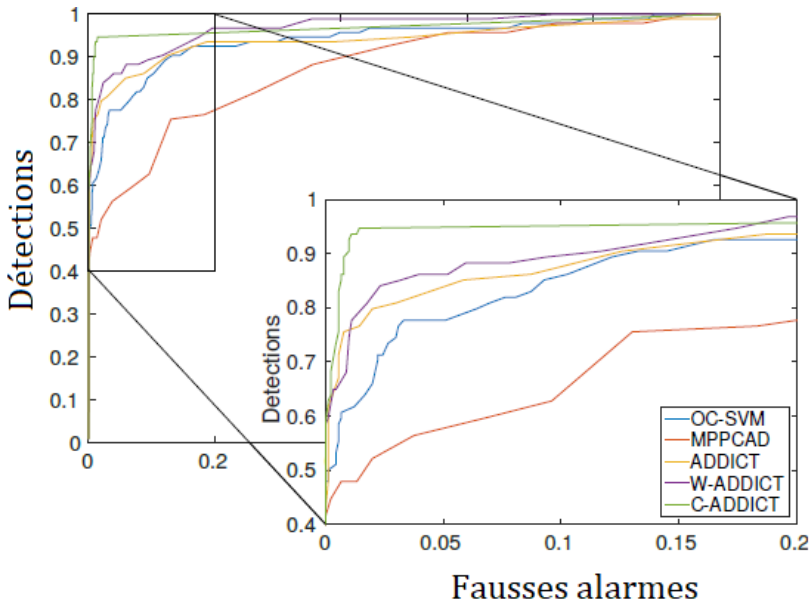

FIGURE 3.11 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par MPPCAD, OC-SVM, C-ADDICT, ADDICT et W-ADDICT (avec l'option d'invariance par translation active et $\tau_{\max} = 5$).

Méthode	Seuil	P_{D}	P_{FA}	AUC
OC-SVM	0.019	80.9%	7%	0.9413
MPPCAD	76	81.9%	25.9%	0.8779
ADDICT	4.1	81%	3%	0.937
W-ADDICT	4.5	85.1%	2.7%	0.9703
C-ADDICT	$\epsilon > 0$	94.7%	1.7%	0.9706

TABLE 3.2 – Résultats de détection en termes de probabilités de détection (P_{D}), de probabilités de fausses alarmes (P_{FA}) et d'aire sous la courbe COR (AUC) pour les algorithmes OC-SVM, MPPCAD, ADDICT, W-ADDICT et C-ADDICT.

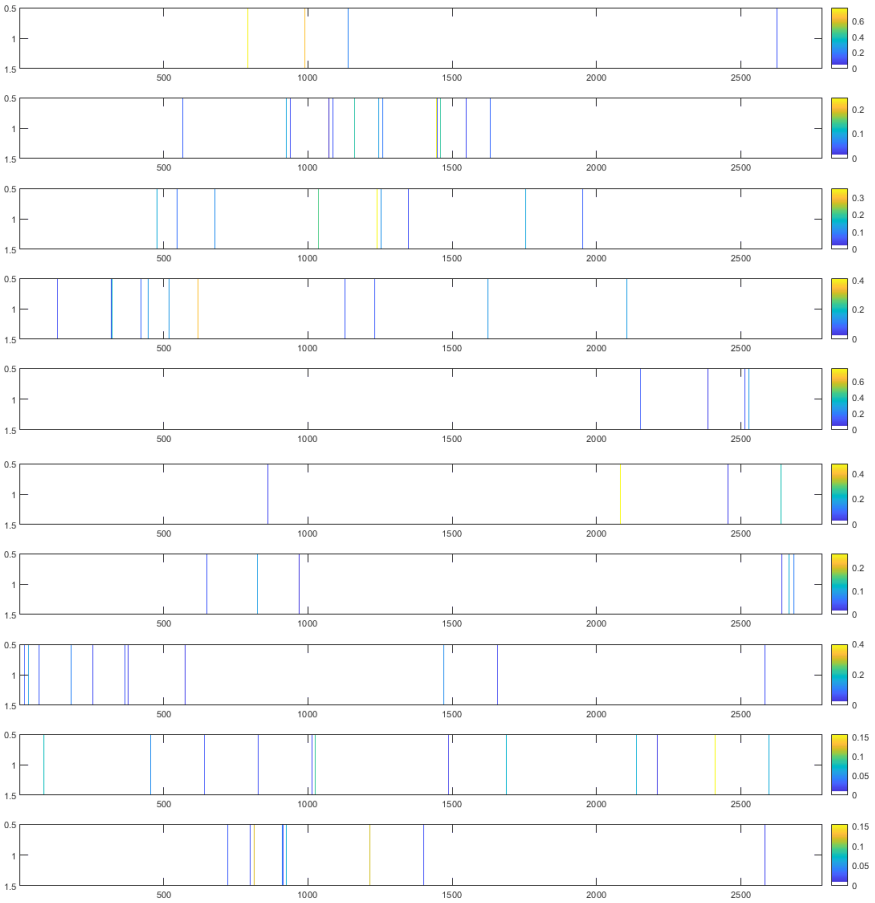

FIGURE 3.12 – Réprésentation des cartes de coefficients associés aux dix premiers filtres du dictionnaire convolutif pour l'estimation convolutive d'une journée de télémesure. Le valeurs non nulles sont représentées par un trait de couleur et leur position dans les cartes indique la position dans le signal étudié où sont observés les différents filtres.

3.4.3.1 Identification des paramètres affectés

L'algorithme C-ADDICT traite les signaux de télémesure mixtes de manière conjointe afin de prendre en compte les relations entre les paramètres et de détecter les anomalies multivariées. Cependant, étant donné la définition du modèle C-ADDICT et la parcimonie imposée aux signaux d'anomalie $\boldsymbol{e}_k, k = 1, ..., L$, la détection peut être réalisée de manière univariée. La figure 3.13 représente les scores d'anomalies obtenus pour chaque signal de la base de test associé à un paramètre de télémesure. Il apparaît que les scores sont positifs presque exclusivement en présence d'anomalies et pour les paramètres affectés. Ces résultats confirment la compétitivité de l'algorithme C-ADDICT qui offre une détection efficace et précise des anomalies en permettant l'identification des paramètres affectés. Ce dernier point est important d'un point de vue opérationnel.

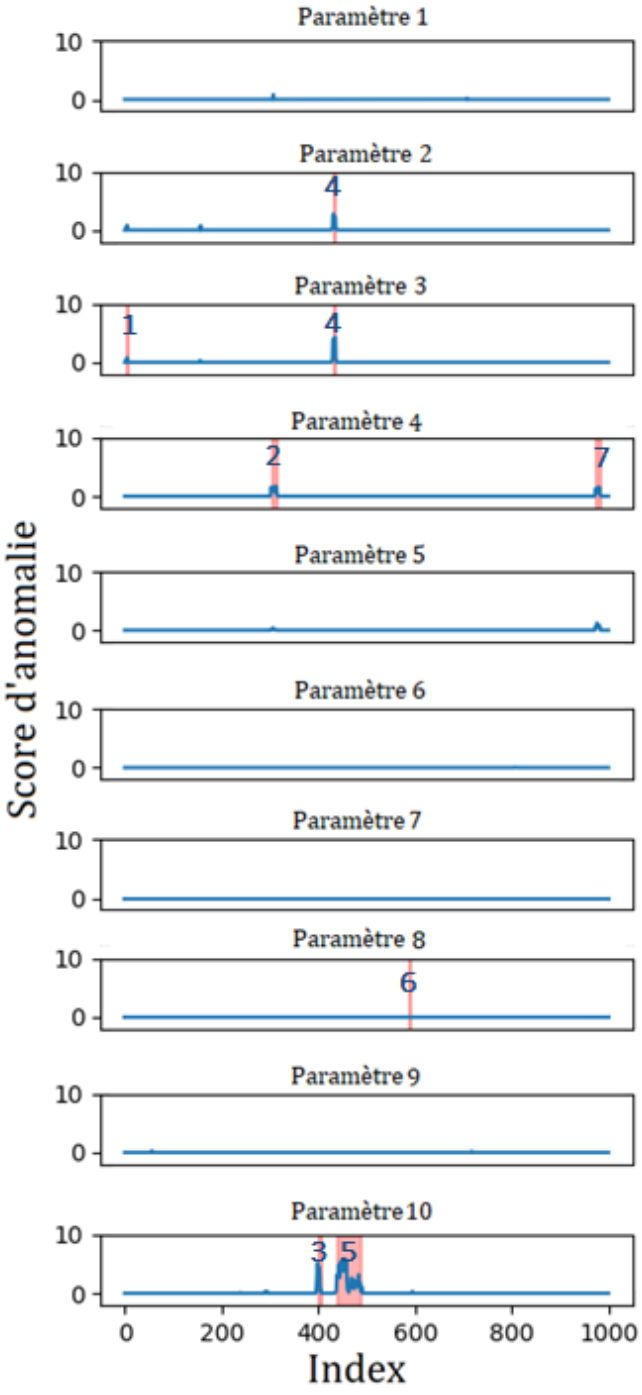

FIGURE 3.13 – Scores d'anomalies univariés retournés pas l'algorithme C-ADDICT. La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1

.

3.4.4 Normalisation des données

De même que pour l'évaluation de l'algorithme ADDICT, la question de la normalisation des données s'est posée. Cette section présente les résultats de détection de l'algorithme C-ADDICT appliqué aux données de test normalisées par la méthode Min-Max. L'étude des scores d'anomalie de la figure 3.14 et des performances de détection représentées sous forme de courbes CORs dans la figure 3.15 et reportées dans le tableau 3.3 mène aux conclusions suivantes : appliqué à des données non normalisées, l'algorithme C-ADDICT détecte la majorité des anomalies de notre vérité terrain avec des scores positifs en présence d'anomalies. Cependant, il retourne beaucoup plus de fausses alarmes que les données sont pas normalisées. En effet, pour un même seuil de détection, il faut accepter 13.7% de fausses alarmes (contre 1.7% avec des données normalisées) pour détecter 83% des signaux affectés (contre 94.5% avec des données normalisées).

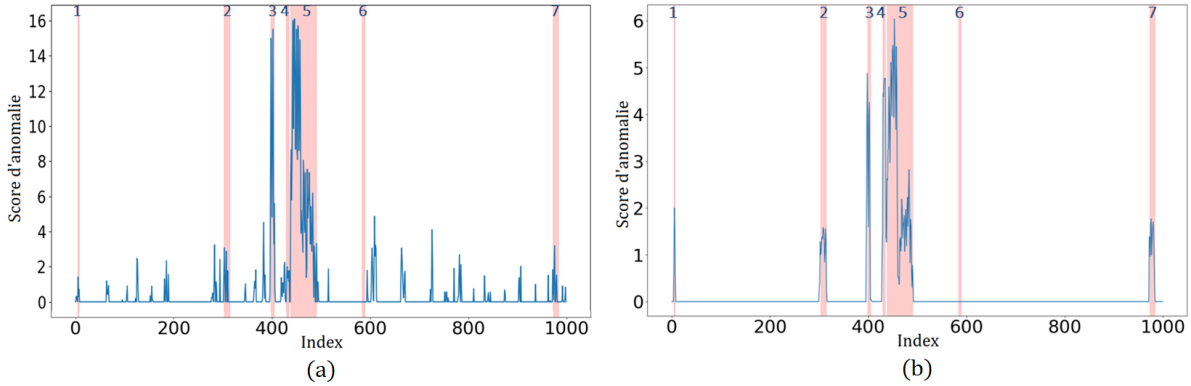

FIGURE 3.14 – Scores d'anomalie obtenus par l'algorithme C-ADDICT appliqué à des données de télémesure non normalisées (a) et normalisées selon la méthode Min-Max (b). La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalies est numérotée de manière à pouvoir être identifiée dans la figure 4.1

Méthode	seuil	P_{D}	P_{FA}
C-ADDICT (Min-Max)	$\epsilon > 0$	94.5%	1.7%
C-ADDICT	$\epsilon > 0$	83%	13.7%

TABLE 3.3 – Résultats de détection en termes de probabilité de détection (P_{D}) et de probabilité de fausse alarme (P_{FA}) pour l'algorithme C-ADDICT appliqué à des données normalisées par la méthode Min-Max et à des données non normalisées.

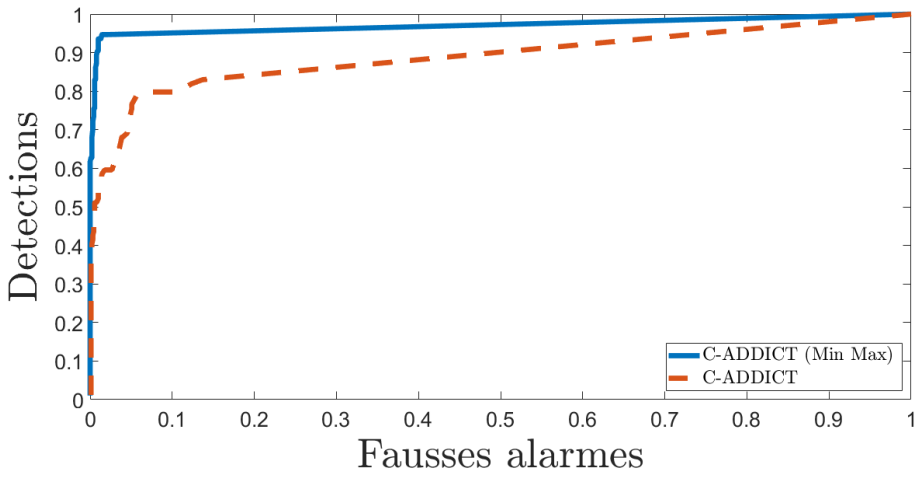

FIGURE 3.15 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par l'algorithme C-ADDICT appliqué à des données de télémesure normalisées par la méthode Min-Max et non normalisées.

3.4.5 Réglage des hyperparamètres

L'utilisation de l'algorithme C-ADDICT nécessite de régler deux hyperparamètres : λ et β qui contrôlent respectivement la parcimonie des cartes de coefficients $\boldsymbol{x}_{k,l}$ et des signaux d'anomalies $\boldsymbol{e}_k$. Nous étudions dans cette section l'intérêt d'un réglage automatique de l'hyperparamètre β à l'aide de la méthode OC-SDT présentée au chapitre précèdent. Appliquée à des signaux sains (sans anomalie), la méthode OC-SDT pour l'estimation de β a mené au résultat suivant : $\widehat{\beta} = 0.3$. Notons que la valeur de β obtenue à l'aide de la vérité terrain est de $\beta = 0.5$. La valeur de l'hyperparamètre λ a été fixée à $\lambda = 0.2$, ce qui correspond à la valeur optimale obtenue à l'aide la vérité terrain. Les scores retournés par l'algorithme C-ADDICT avec ce réglage sont représentés dans la figure 3.16 (a). En comparaison avec les scores retournés par l'algorithme C-ADDICT avec $\beta = 0.5$ (voir figure 3.16 (b)), on observe que le seuil de détection est plus difficile à régler, et ce, malgré une valeur estimée de β très proche de la valeur optimale. Ces résultats témoignent de la sensibilité de l'algorithme C-ADDICT au choix de cet hyperparamètre. Cependant, cela n'entrave en rien la détection des anomalies. En effet, on observe que des scores biens supérieurs à la moyenne ont été enregistrés presque exclusivement en période d'anomalie. Le choix d'un seuil de détection adapté permet alors de détecter les anomalies de manière efficace. L'étude des courbes CORs de la figure 3.17 et la lecture du tableau 3.4 vérifient ces conclusions. En effet, avec une valeur estimée de β et un seuil de détection adapté, l'algorithme C-ADDICT détecte 92.5% des signaux affectés par une anomalie et ne retourne que 3% de fausses alarmes. Ces performances sont très proches de celles obtenues avec un réglage optimal des hyperparamètres. L'intérêt de la méthode OC-SDT pour le réglage de l'hyperparamètre β se confirme pour l'algorithme C-ADDICT.

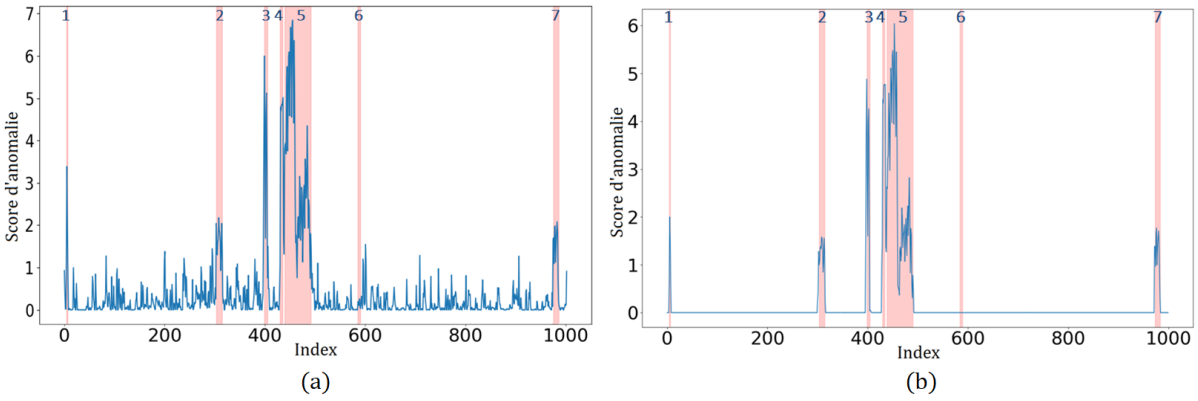

FIGURE 3.16 – Scores d'anomalie obtenus par l'algorithme C-ADDICT avec $\lambda = 0.2$ et β réglé de manière automatique par la méthode OC-SDT (a, $\beta = 0.3$), ou à l'aide de la véritré terrain (b, $\beta = 0.5$). La vérité terrain est représentée par les fonds rouges. Chaque période d'anomalie est numérotée de manière à pouvoir être identifiée dans la figure 4.1

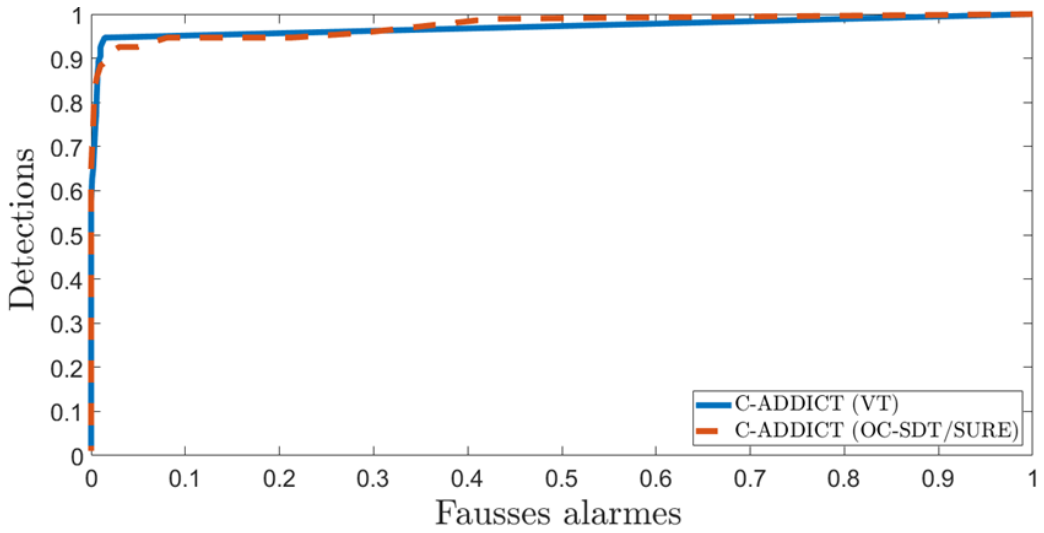

FIGURE 3.17 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par l'algorithme C-ADDICT pour lequel l'hyperparamètre β a été réglé à l'aide de la méthode OC-SDT ou de la vérité terrain.

Méthode	seuil	P_{D}	P_{FA}	AUC
C-ADDICT (VT)	$\epsilon > 0$	94.5%	1.7%	0.9706
C-ADDICT (OCSDT)	1	92.5 %	3%	0.9784

TABLE 3.4 – Résultats de détection en termes de probabilités de détection (P_{D}), de probabilités de fausse alarme (P_{FA}) et d'aire sous la courbe COR (AUC) pour l'algorithme C-ADDICT dont l'hyperparamètre β a été déterminé à l'aide de la méthode OC-SDT ou de la vérité terrain (VT).

3.5 Conclusion

Ce chapitre présente l'algorithme C-ADDICT, une méthode de détection d'anomalies multivariées capable de traiter conjointement des données mixtes. Il s'agit d'une extension de l'algorithme ADDICT qui a l'avantage d'être invariante par translation grâce à l'utilisation d'un dictionnaire convolutif. L'introduction d'un tel dictionnaire a permis de traiter des données mixtes selon une unique stratégie, ce qui est, comme nous avons pu le voir, plus difficile avec un dictionnaire fixe. Par ailleurs, la stratégie d'estimation parcimonieuse développée a pu être étendue afin de répondre au problème d'apprentissage du dictionnaire convolutif de données mixtes. Ce dernier point représente un avantage certain de la méthode vis à vis de l'algorithme ADDICT pour lequel l'apprentissage d'un dictionnaire fixe de données mixtes représente un problème complexe qui n'a pas encore été résolu.
L'algorithme C-ADDICT a été évalué avec une base de données de test composée

de données de télémesure réelles affectées par des anomalies connues et identifiées par des experts. Les premiers résultats obtenus sont très encourageants. En effet, avec 94.7% de détection et 1.7% de fausses alarmes, les performances de l'algorithme C-ADDICT sont très compétitives par rapport à celles des méthodes de l'état de l'art considérées dans cette thèse. Ces premiers résultats ont également mis en évidence la capacité de l'algorithme à identifier les paramètres affectés par les anomalies. Notons toutefois qu'une normalisation des données est nécessaire pour atteindre de telles performances. Par ailleurs, malgré une grande sensibilité au choix des hyperparamètres, et notamment de β, il a été démontré sur ces données qu'un réglage automatique de ce dernier à l'aide des méthodes OC-SDT était possible et ne dégradait que très peu les performances de détection.
Au-delà des aspects quantitatifs, l'algorithme C-ADDICT compte de nombreux avantages comparé à l'algorithme ADDICT. L'atout principal de la méthode est sans doute son algorithme d'apprentissage qui, ayant été développé autour de la stratégie de détection C-ADDICT, permet de construire un dictionnaire de données mixtes de qualité. Un deuxième avantage concerne la taille du dictionnaire. En effet, alors que la surveillance de 10 paramètres de télémesure ($K = 10$) à l'aide d'un dictionnaire fixe nécessite 2000 atomes (détection ADDICT), seuls 100 filtres sont nécessaires pour la détection C-ADDICT. Ce point peut être important en cas de contraintes de stockage souvent présentes pour les systèmes embarqués par exemple. Nous l'aurons compris, l'algorithme C-ADDICT s'impose sur de nombreux points face à l'algorithme ADDICT. Cependant, un point négatif de l'algorithme C-ADDICT est toutefois à noter et concerne son temps d'exécution supérieur à celui de l'algorithme ADDICT. En effet, l'utilisation d'un dictionnaire convolutif complexifie le problème de détection et rend impossible la parallélisation de la détection, qui peut être mise en oeuvre avec un dictionnaire fixe. En conséquence, la détection pour une journée de données et 10 paramètres de télémesure est réalisée en 58 secondes avec l'algorithme C-ADDICT, contre 9 secondes avec l'algorithme ADDICT. Ces temps d'exécution restent raisonnables pour 10 paramètres mais risque de croître de manière exponentielle si l'on considère plusieurs dizaines voir centaines de paramètres de télémesure. La comparaison des algorithmes ADDICT et C-ADDICT d'un point de vue quantitatif et qualitatif est résumée dans le tableau 3.5.

	ADDICT	C-ADDICT
Détection P_{D}	ADDICT : 81% W-ADDICT :85.1% G-ADDICT : 93.6%	94.7%
Fausses alarmes P_{FA}	ADDICT : 3% W-ADDICT : 2.7% G-ADDICT : 3.3%	1.7%
Invariance par translatation	✓ (option)	✓
Algorithme d'apprentissage de dictionnaire adapté	X	✓ (10h pour $\Phi \in \mathbb{R}^{10\times 50\times 100}$)
Algorithme d'apprentissage de dictionnaire	X	✓
Nombre d'atomes/filtres ($K = 10$, $W = 50$)	2000	100
Temps de calcul (Détection 1 jour, en seconde)	9' (parallélisation)	58'

TABLE 3.5 – Comparaison quantitative et qualitative des algorithmes ADDICT et C-ADDICT.

Chapitre 4

Applications industrielles

Sommaire

4.1 Introduction

Les deux chapitres précédents introduisent l'algorithme ADDICT et ses extensions, des méthodes innovantes de détection d'anomalies multivariées qui reposent sur des modèles d'estimation parcimonieuse et d'apprentissage de dictionnaire. Ce dernier chapitre applique ces solutions à de réels cas d'usage industriels de surveillance de la télémesure satellite et reporte les résultats expérimentaux obtenus. Le premier cas d'application étudié considère dix paramètres de télémesure pour lesquels une vérité terrain composées d'anomalies variées identifiées par des experts est disponible. Cet exemple a déjà été traité dans les précédents chapitres et a permis d'évaluer les performances des algorithmes proposés. Les premiers résultats obtenus sont satisfaisants et compétitifs vis à vis des méthodes de l'état de l'art mais la présence de fausses alarmes est à noter. Rappelons que dans un contexte opérationnel chaque détection par l'algorithme entraîne une investigation de la part des experts afin de confirmer la présence de l'anomalie ou d'invalider la détection qui est alors considérée comme une fausse alarme. Cette procédure mobilise des moyens importants ce qui la rend coûteuse et fait de la limitation des fausses alarmes un enjeu important. Dans ce chapitre nous proposons d'adresser ce problème à l'aide d'une approche d'apprentissage séquentiel ou "en ligne" permettant de prendre en compte le retour utilisateur au fur et à mesure des détections, de mettre à jour le modèle de référence, et de limiter le nombre de fausses alarmes retournées par l'algorithme. Cette stratégie d'apprentissage "en ligne" a déjà été étudiée pour plusieurs méthodes de détection d'anomalies dans Raginsky et al. (2012); Laxhammar and Falkman (2014); Ahmed et al. (2007). La version "en ligne" de l'algorithme ADDICT, appelée Online ADDICT, consiste à mettre à jour le dictionnaire par l'actualisation de ses éléments et/ou l'ajout de nouveaux éléments dans le dictionnaire. Par ailleurs, ce premier cas d'usage et sa vérité terrain disponible sont exploités dans ce chapitre afin de réaliser un passage à l'échelle et d'étudier les performances de détection des méthodes proposées dans un contexte plus proche des besoins opérationnels où plusieurs dizaines de paramètres sont traités conjointement.
Les deux cas d'usage suivants décrivent des situations particulières de surveillance de la télémesure satellite durant lesquelles ont eu lieu des manoeuvres de correction d'orbite ou des éclipses. Ces évènements affectent de manière significative la télémesure satellite mais ils sont très peu observés (quelques occurences par an). En conséquence, une quantité suffisante de télémesure associée est rarement disponible en début de vie du satellite et ces évènements ne sont donc pas pris en compte pour lors de l'apprentissage du modèle de référence. En conséquence, chacune de leur réalisation génère de nombreuses fausses alarmes. De manière analogue au premier cas d'usage, une solution intéressante pour limiter la présence des fausses alarmes liées à ce genre d'évènements est d'adopter une stratégie d'apprentissage "en ligne" permettant de les intégrer séquentiellement au modèle et de s'affranchir d'un nouvel apprentissage complet. Ces cas d'usage sont alors deux occasions supplémentaires d'évaluer les performances des algorithmes proposés et d'étudier une version adaptative de l'algorithme ADDICT (également étudiée pour le premier cas d'usage) pour limiter les fausses alarmes liées à l'apparition de nouveaux évènements.
Dans le dernier cas d'usage traité dans ce chapitre, les algorithmes développés dans cette thèse sont appliqués pour la surveillance de la télémesure satellite affectée par un évènement connu mais imprévisible qui peut avoir de lourdes conséquences sur le fonctionnement du satellite

et le bon déroulement de sa mission. L'objectif de ce cas d'application est alors d'étudier sur un exemple réel l'intérêt des approches de détection d'anomalies multivariées proposées pour étudier un évènement particulier et tenter de le comprendre et/ou prévenir son apparition.

4.2 Surveillance de la télémesure satellite

4.2.1 Présentation du cas d'usage

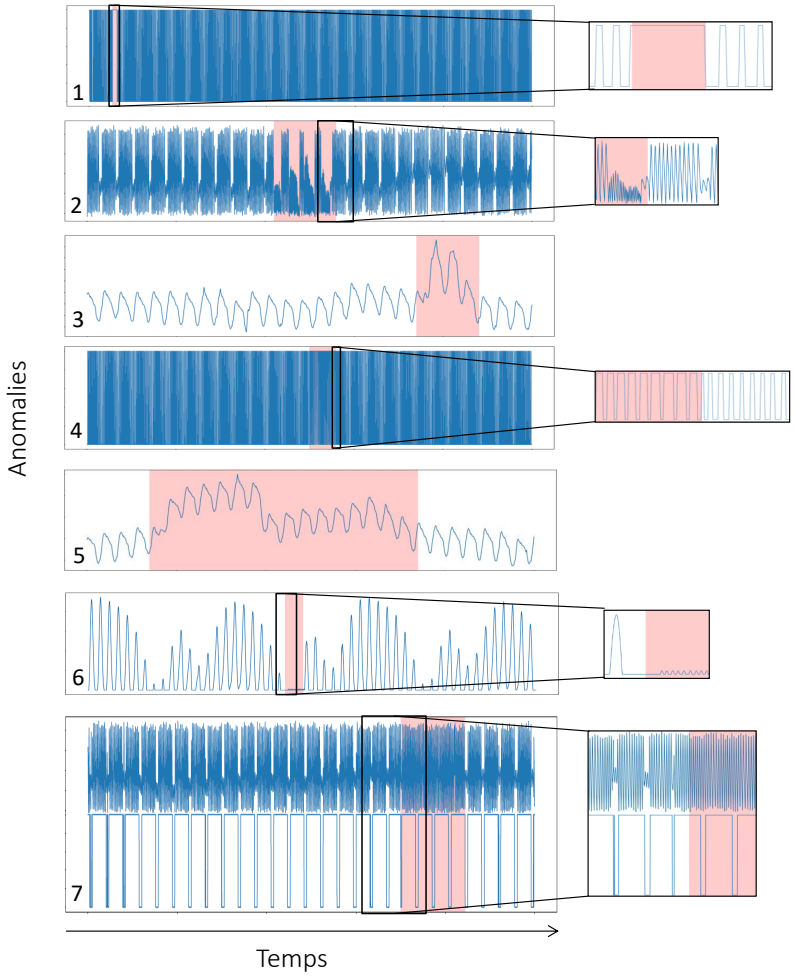

FIGURE 4.1 – Périodes d'anomalies de la vérité terrain

Le premier cas d'application auquel nous nous intéressons dans ce chapitre s'appuie sur une base de données de test composée d'anomalies diverses et munie d'une vérité terrain

permettant de mesurer quantitativement les performances de détection des méthodes testées en termes de probabilités de fausse alarme et de probabilités de détection. Dans cette base d'anomalies, 7 paramètres de télémesure différents connaissent une ou plusieurs périodes d'anomalies. Les périodes d'anomalies de la vérité terrain sont représentées dans la figure 4.1 (encadrés rouges). Notons que parmi ces 7 paramètres, 3 sont discrets et 4 sont continus.
Comme évoqué précédemment, ce cas d'application a été traité dans les précédents chapitres en considérant conjointement 10 paramètres de télémesure (les 7 paramètres de la vérité terrain et 3 paramètres continus supplémentaires pour lesquels aucune anomalie n'a été observée). Ces données ont permis d'évaluer les performances des différentes méthodes de détection d'anomalies proposées ainsi que celles de l'état de l'art et les résultats obtenus sont encourageants. Cependant, la présence de fausses alarmes est à noter. Par ailleurs, rappelons que la télémesure satellite peut compter plusieurs milliers de paramètres de télémesure et que dans un contexte opérationnel de surveillance de la télémesure l'objectif est de pouvoir traiter conjointement au moins quelques dizaines de paramètres. Les résultats obtenus dans les chapitres précédents sur cet exemple avec 10 paramètres traités conjointement ne nous permettent donc pas de conclure quant à la capacité de l'algorithme ADDICT et ses extensions à répondre à un tel besoin. Les anomalies de la vérité terrain sont donc à nouveau considérées dans cette section pour : 1) étudier l'intérêt d'un apprentissage "en ligne“ par les algorithmes ADDICT et W-ADDICT afin de limiter les fausses alarmes, et 2) évaluer les performances de l'algorithme ADDICT avec plusieurs dizaines de paramètres. Notons que les données de ce premier cas d'usage sont majoritairement issues d'un satellite d'observation opéré par le CNES.

4.2.2 Détection d'anomalies séquentielle

Nous reprenons dans cette section les expérimentations menées dans les chapitres précédents et étudions l'intérêt d'une approche d'apprentissage séquentielle pour la prévention des fausses alarmes retournées par l'algorithme ADDICT et ses extensions. Une première stratégie naïve de détection d'anomalies séquentielle par l'algorithme ADDICT consiste à ajouter au dictionnaire les signaux testés identifiés comme fausses alarmes par les experts. Cette piste a été explorée pour ce cas d'application avec l'algorithme ADDICT et sa version pondérée Weighted-ADDICT (W-ADDICT, voir section 2.4.2 pour plus de détails).

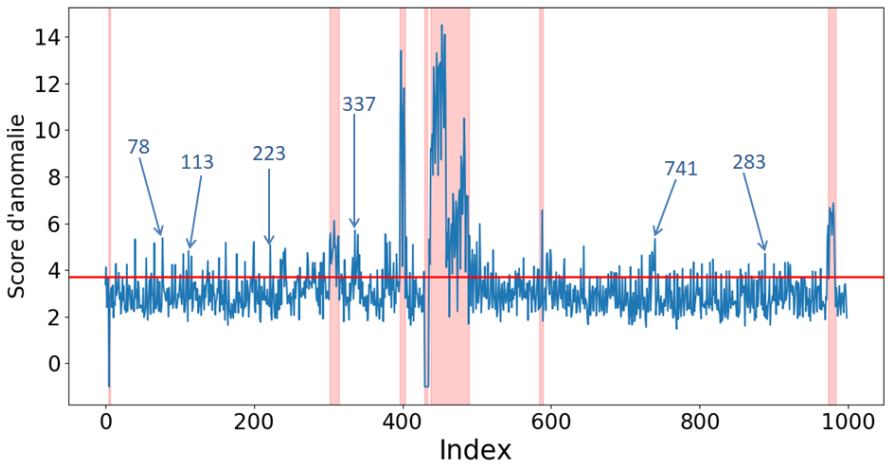

FIGURE 4.2 – Scores d'anomalie retournés par l'algorithme ADDICT pour les signaux de la base d'anomalies. La vérité terrain est représentée par les fonds rouges. Les périodes d'anomalie de la vérité terrain sont représentées dans la figure 4.1. Le seuil de détection S_{PFA} est représenté par une droite horizontale rouge.

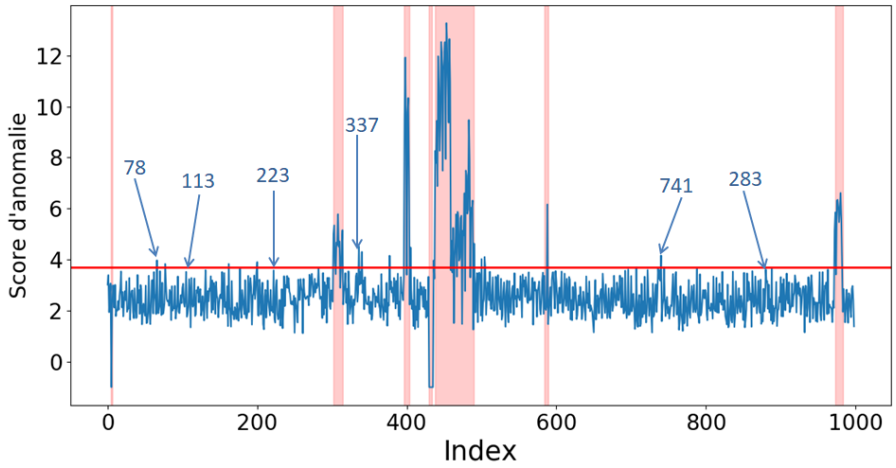

FIGURE 4.3 – Scores d'anomalie retournés pour les signaux de la base d'anomalies par l'algorithme Online ADDICT (i.e., l'algorithme ADDICT après mise à jour du dictionnaire par ajout des fausses alarmes). Les périodes d'anomalie de la vérité terrain sont représentées dans la figure 4.1. Le seuil de détection S_{PFA} est représenté par une droite horizontale rouge.

Cette expérimentation traite 10 signaux dont 7 continus et 3 discrets. Le dictionnaire a été construit à partir de 2 mois de télémesure de ces signaux décrivant un fonctionnement normal du satellite (sans anomalie). Les scores d'anomalies retournés par l'algorithme ADDICT pour les différents signaux de la base de test sont représentés dans la figure 4.2. Le seuil de détection fixé à $S_{\text{PFA}} = 3.7$ est représenté par une droite horizontale rouge. Il s'agit du seuil pour lequel le couple $(P_{\text{D}}, P_{\text{PFA}})$ associé est le plus proche du point optimal (0,1). Une anomalie est détectée par l'algorithme lorsque le score excède ce seuil. Les scores supérieurs au seuil fixé situés en dehors des périodes d'anomalie connues de la vérité terrain (fonds rouges)

correspondent à des fausses alarmes.
Afin d'évaluer la capacité de l'algorithme à intégrer ces fausses alarmes et prévenir leur détection, les fausses alarmes (i.e ;, les signaux test dont le score excède le seuil de détection fixé à 3.7, soit 152 signaux multivariés) ont été ajoutées au dictionnaire et l'algorithme a été appliqué aux mêmes données de test. En comparant les scores d'anomalies avant (figure 4.2, ADDICT) et après (figure 4.3, Online ADDICT) mise à jour du dictionnaire, il apparaît que l'intégration des fausses alarmes dans le dictionnaire a permis d'empêcher de détecter à nouveau certaines fausses alarmes. C'est notamment le cas des signaux tests #113, #223 et #283 par exemple. Les résultats quantitatifs représentés sous formes de courbes CORs dans la figure 4.7 et résumés dans le tableau 4.1 confirment ces premières conclusions avec 10.2% de fausses alarmes retournées initialement contre 7.7% après mise à jour du dictionnaire pour 89% de signaux atypiques détectés. Cependant, à seuil fixé $S_{\text{PFA}} = 3.7$, l'ajout des fausses alarmes au dictionnaire a également fait chuter le taux de bonnes détections qui passe de 89% à 83% après la mise à jour du dictionnaire. Il faut alors abaisser le seuil de détection à $S_{\text{PFA}} = 3.5$ pour atteindre le taux de détection initial de 89% des anomalies. Par ailleurs, on s'aperçoit qu'une proportion importante des fausses alarmes est toujours détectée par l'algorithme. C'est le cas notamment des signaux #78, #337 et #741 par exemple.

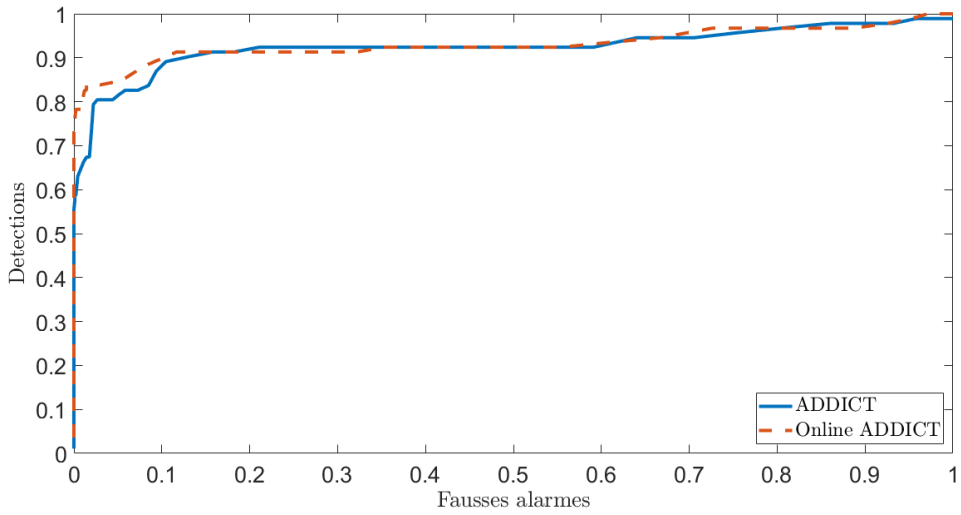

FIGURE 4.4 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par l'algorithme ADDICT avant et après mise à jour du dictionnaire.

Méthode	seuil	P_{D}	P_{FA}	AUC
ADDICT	3.7	89%	10.2%	0.937
Online ADDICT	3.7	83%	1.2%	0.947
Online ADDICT	3.5	89%	7.7%	0.947

TABLE 4.1 – Résultats de détection en termes de probabilités de détection (P_{D}), de probabilités de fausse alarme (P_{FA}) et d'aire sous la courbe COR (AUC) pour les algorithmes ADDICT et Online ADDICT (i.e., l'algorithme W-ADDICT après mise à jour du dictionnaire par ajout des fausses alarmes.

La même expérimentation a été réalisée avec la version pondérée de l'algorithme ADDICT, W-ADDICT (voir section 2.4.2 pour plus de détails). Les scores d'anomalie retournés par l'algorithme avant et après l'ajout des fausses alarmes au dictionnaire sont représentés dans les figures 4.5 et 4.6 et témoignent de l'intérêt de la mise à jour du dictionnaire pour la prévention des fausses alarmes. C'est notamment le cas des signaux #113, #741 et #883 par exemple. Les résultats quantitatifs représentés sous forme de courbes CORs et résumés dans le tableau 4.2 appuient ces conclusions. On y découvre que pour un seuil de détection fixé à $S_{\text{PFA}} = 3.5$, l'ajout des fausses alarmes dans le dictionnaire a fait chuter la probabilité de fausses alarmes qui passe de 9.9% à 4.4% et que la probabilité de bonne détection reste stable (88.3%). Rappelons que dans la première expérimentation menée avec l'algorithme ADDICT, l'ajout des fausses alarmes au dictionnaire avait fait chuter le taux de fausse alarme mais également le taux de bonne détection. La pondération semble assurer une certaine robustesse de l'algorithme à la mise à jour du dictionnaire et au choix du seuil de détection.

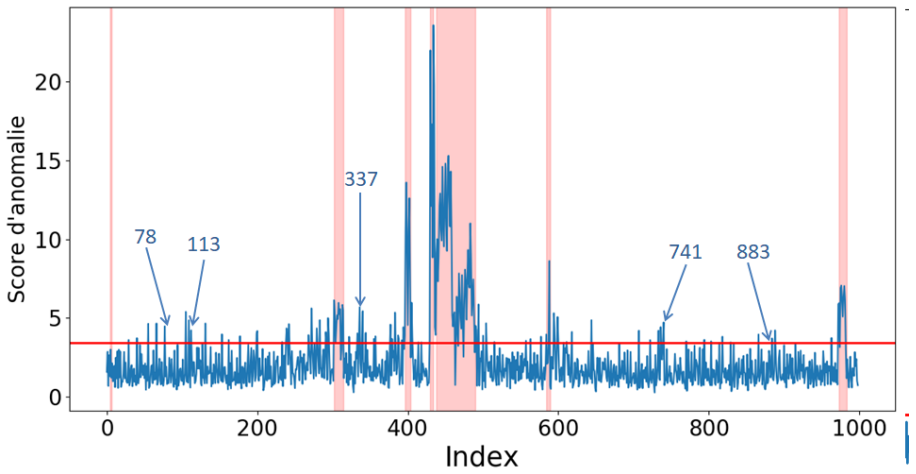

FIGURE 4.5 – Scores d'anomalie retournés par l'algorithme W-ADDICT pour les signaux de la base d'anomalies. La vérité terrain est représentée par les fonds rouges. Les périodes d'anomalie de la vérité terrain sont représentées dans la figure 4.1. Le seuil de détection S_{PFA} est représenté par une droite horizontale rouge.

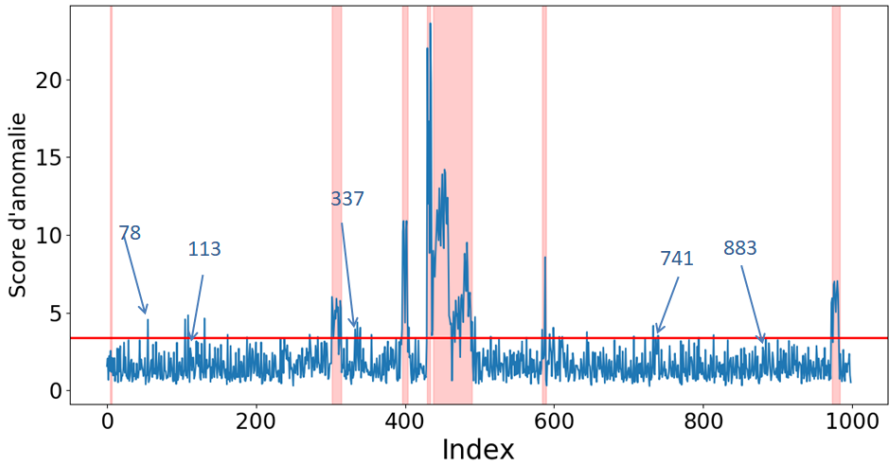

FIGURE 4.6 – Scores d'anomalie retournés pour les signaux de la base d'anomalies par l'algorithme Online W-ADDICT (i.e., l'algorithme W-ADDICT après mise à jour du dictionnaire par ajout des fausses alarmes). La vérité terrain est représentée par les fonds rouges. Les périodes d'anomalie de la vérité terrain sont représentées dans la figure 4.1. Le seuil de détection S_{PFA} est représenté par une droite horizontale rouge.

Méthode	seuil	P_{D}	P_{FA}	AUC
W-ADDICT	3.5	88.3%	9.9%	0.96
Online W-ADDICT	3.5	88.3%	4.4%	0.9766

TABLE 4.2 – Résultats de détection en termes de probabilités de détection (P_{D}) et de probabilités de fausses alarmes (P_{FA}) pour les algorithmes W-ADDICT et Online W-ADDICT (i.e., l'algorithme W-ADDICT après mise à jour du dictionnaire par ajout des fausses alarmes).

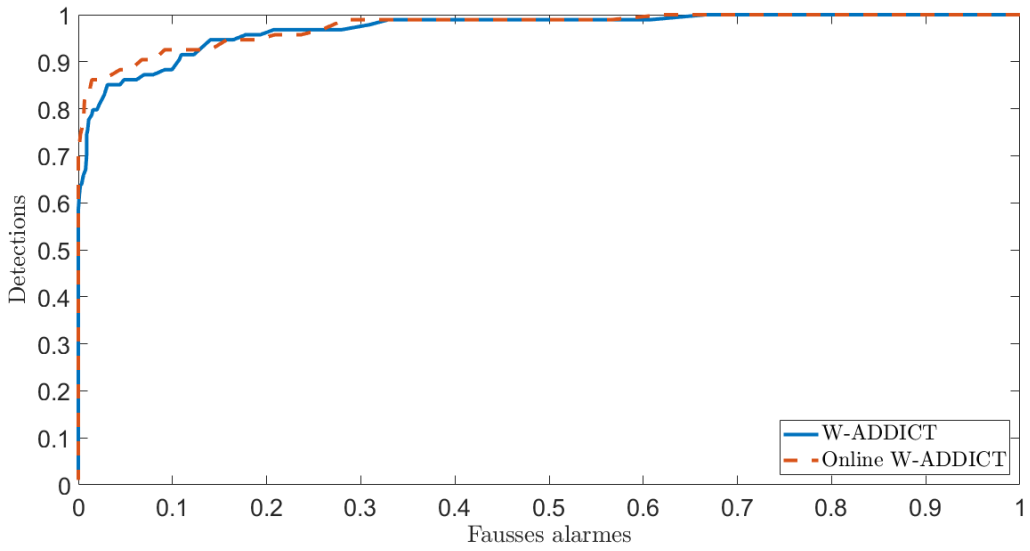

FIGURE 4.7 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par l'algorithme ADDICT avant et après mise à jour du dictionnaire.

Les résultats présentés dans cette section mettent en évidence l'intérêt d'une stratégie de détection séquentielle par l'algorithme ADDICT. En effet, l'approche naïve d'intégration des fausses alarmes dans le dictionnaire a permis dans de nombreux cas d'empêcher une nouvelle détection de fausses alarmes observées par le passé. Cependant une proportion importante de fausses alarmes est toujours détectée malgré leur intégration dans le dictionnaire. L'intérêt de l'approche testée dans cette section reste donc limitée et la stratégie de mise à jour du dictionnaire est une piste qui mérite d'être approfondie afin de développer une méthode de détection plus efficace.

4.2.3 Passage à l'échelle

Cette section teste la capacité de l'algorithme ADDICT à détecter des anomalies qui affectent un ou plusieurs paramètres de télémesure parmi plusieurs dizaines de paramètres traités conjointement. L'algorithme ADDDICT a ainsi été appliqué à 10, 20 et 30 paramètres en ajoutant aux 7 paramètres de la vérité terrain 3, 13 et 23 paramètres continus issus du même satellite et pour lesquels aucune anomalie n'a été observée. Les performances de détection en termes de probabilités de fausse alarme et de probabilités de détection sont représentées sous forme de courbes CORs dans la figure 4.8.

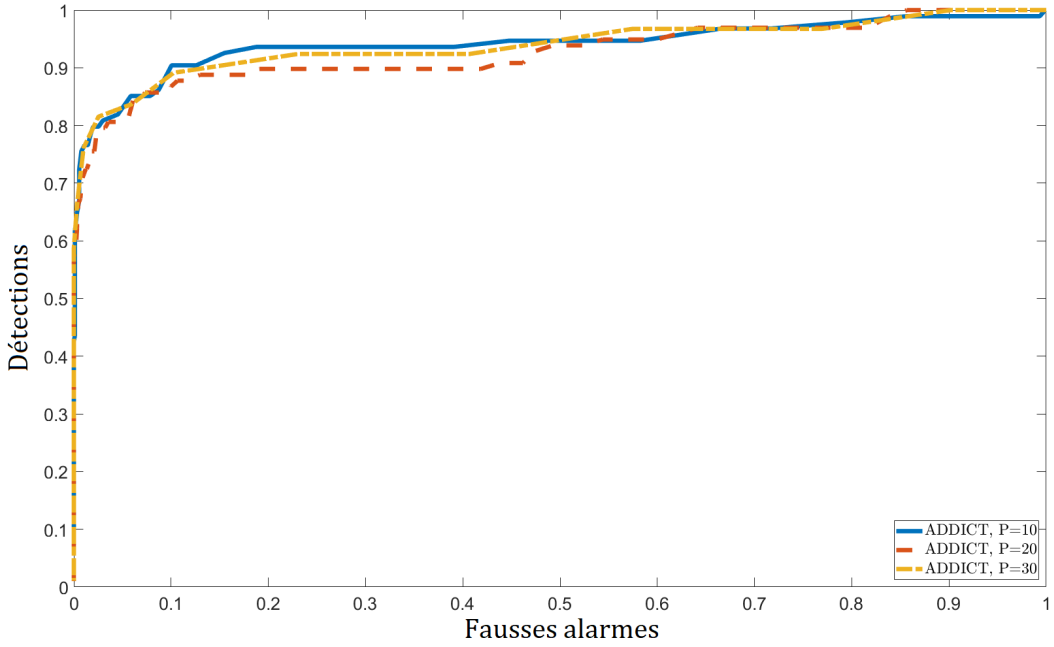

FIGURE 4.8 – Comparaison des courbes CORs obtenues à partir des résultats de détection retournés par l'algorithme ADDICT considérant conjointement 10, 20 et 30 paramètres de télémesure.

Paramètres	Atomes	Temps d'exécution (en secondes)
10	2000	3
20	8000	134
30	15000	205

TABLE 4.3 – Comparaison en termes de taille du dictionnaire et de temps de calcul de l'algorithme ADDICT appliqué à 10, 20 et 30 paramètres de télémesure.

Au regard de ces résultats, il apparaît que jusqu'à 30 paramètres les performances de détection de l'algorithme ADDICT sont équivalentes. Il s'agit ici d'un passage à l'échelle limité étant donné les moyens importants en termes de capacité de calcul et de mémoire que requiert un passage à l'échelle de plus grande envergure. En effet, augmenter le nombre de paramètres à traiter conjointement n'est pas sans conséquences. La principale conséquence concerne la taille du dictionnaire et plus précisément le nombre d'atomes et leur taille. Rappelons que selon la taille W des fenêtres temporelles étudiées et le nombre K de paramètres surveillés, la taille des signaux multivariés et des atomes est de $N = W \times K$ (voir le pré-traitement ADDICT décrit dans la section 2.3.1 pour plus de précisions). Par ailleurs, comme nous avons pu le constater précédemment, une détection d'anomalies efficace à l'aide d'une méthode d'estimation parcimonieuse n'est possible que si le dictionnaire compte un nombre suffisant d'atomes (voir section 2.3.5.5). Pour l'exemple étudié ici, les chiffres reportés dans le tableau 4.3 ci-dessus illustrent ce phénomène : un traitement conjoint de 20 signaux de télémesure nécessite de construire un dictionnaire de 8 000 atomes pour atteindre les performances de détection obtenues avec 10 signaux et un dictionnaire de 2 000 atomes. Avec 30 signaux, 15 000 atomes permettent d'atteindre ces performances. Notons que dans ces expérimentations, la taille des fenêtres temporelles est $W = 50$. La taille des atomes (nombre de points de chaque élément du dictionnaire) est alors de 500, 1000 et 1500 pour 10, 20 et 30 signaux. Nécessairement, de tels volumes de données complexifient les calculs et rallongent le temps d'exécution de l'algorithme qui passe de 3 secondes pour traiter 10 signaux à 134 secondes pour 20 signaux et à 205 secondes pour 30 signaux. Faute de temps et de ressources disponibles cette étude n'a pas pu être étendue à plus que 30 paramètres de télémesure ni réalisée pour les extensions de l'algorithme ADDICT.

4.3 Détection de manoeuvre de correction d'orbite

4.3.1 Présentation du cas d'usage

Ce deuxième cas d'application s'intéresse à la surveillance de la télémesure en période de manoeuvre de correction d'orbite et soulève l'intérêt de : 1) l'utilisation d'une méthode de détection d'anomalies multivariées pour prendre en compte le contexte opérationnel des satellites et limiter les fausses alarmes liées aux manoeuvres, et 2) la mise à jour séquentielle de l'algorithme afin d'intégrer les nouvelles manoeuvres réalisées. Cet exemple traite 5 pa-

ramètres de télémesure dont 4 continus et 1 discret issus d'un satellite d'observation opéré par le CNES. Un extrait de télémesure de ces paramètres est représenté dans la figure 4.9. L'information concernant la réalisation des manoeuvres est portée par le paramètre discret (signal vert) qui prend la valeur 1 en période de manoeuvre et 0 sinon. Comme nous pouvons le voir, la télémesure enregistre en périodes de manoeuvre (fond gris) des comportements atypiques comparés à ceux associés à un fonctionnement courant du satellite (hors manoeuvre). Dans un contexte de manoeuvre ces signatures ne représentent pas une anomalie. Cependant, si elles sont enregistrées dans un contexte de fonctionnement courant du satellite (hors période de manoeuvre), elles peuvent traduire la présence d'une anomalie (fond rouge). L'intérêt d'une méthode de détection d'anomalies multivariées pour cet exemple est donc de pouvoir détecter ces comportements atypiques si aucune manoeuvre n'a été programmée et d'éviter la détection de fausses alarmes en période de manoeuvre. Cela suppose de traiter conjointement plusieurs paramètres de télémesures discrets et continus afin de prendre en compte le contexte opérationnel du satellite. Rappelons que l'algorithme de détection d'anomalies NOSTRADAMUS développé par le CNES est un algorithme de détection d'anomalies univariées qui traite les paramètres de télémesure indépendamment les uns des autres. En conséquence, aucune information de contexte n'est prise en compte et les comportements atypiques liés aux manoeuvres déclenchent systématiquement la détection d'une anomalie par l'algorithme. Notons que l'apprentissage des ces signatures par l'algorithme est risqué car une anomalie qui affecte la télémesure de manière analogue à une manoeuvre pourrait ne pas être détectée. Par ailleurs, comme évoqué précédemment, cet exemple s'intéresse à l'apprentissage séquentiel des manoeuvres considérées comme de nouveaux évènements à prendre en compte au cours de la vie du satellite. L'enjeu de ce cas d'application est alors de déterminer dans quelle mesure les méthodes proposées dans cette thèse sont capables d'intégrer les manoeuvres au fur et à mesure de leur réalisation pour limiter la détection des prochaines manoeuvres tout en assurant la détection des potentielles anomalies.

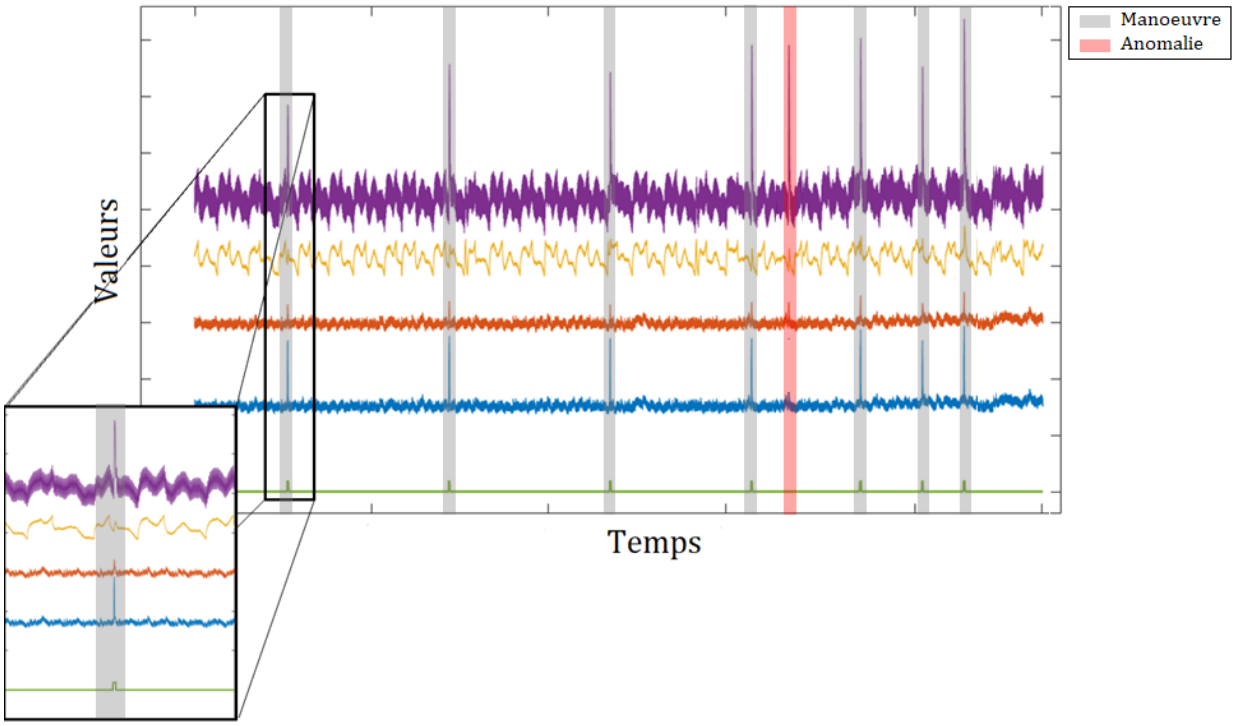

FIGURE 4.9 – Télémesure de 5 paramètres dont 4 continus et un discret (signal vert). L'information concernant la programmation des manoeuvres est portée par le paramètre discret binaire prenant la valeur 1 en périodes de manoeuvre et 0 sinon. Les périodes de manoeuvre sont marquées par les fonds gris et la période d'anomalie par un fond rouge.

Les données présentées dans la figure 4.9 constituent une base de données de test dont la vérité terrain recense les périodes de manoeuvre (fond gris) et une période d'anomalie multivariée (fond rouge). Les périodes en dehors des périodes de manoeuvre ou d'anomalies sont associées à un fonctionnement courant du satellite. L'anomalie considérée ici est fictive et se définie par des comportements atypiques observés pour certains paramètres continus et analogues à ceux enregistrés en périodes de manoeuvre. Cependant le paramètre discret, prenant des valeurs nulles sur cette période, n'indique pas qu'une manoeuvre a été réalisée. Il s'agit donc d'une anomalie multivariée. La détection d'une telle anomalie suppose de traiter conjointement plusieurs paramètres de télémesure discrets et continus. Les algorithmes ADDICT et C-ADDICT ont été appliqués à ces données et les résultats obtenus selon plusieurs scénarios sont présentés à la section suivante.

4.3.2 Résultats expérimentaux

Afin d'évaluer les algorithmes ADDICT et C-ADDICT pour ce cas d'application, ces derniers ont été testés selon plusieurs scénarios. Un premier scénario consiste à construire le dictionnaire à partir de données associées à un fonctionnement courant du satellite. Les manoeuvres sont considérées comme de nouvelles opérations encore jamais menées et la té-

lémesure associées exclut ces périodes de manoeuvre pour l'apprentissage. Pour ce premier scénario, il s'agit de s'assurer qu'une anomalie est bien détectée par les différentes méthodes pour toutes les périodes de manoeuvre ainsi que pour la période d'anomalie. Dans les scénarios suivants, les manoeuvres ont été progressivement intégrées au dictionnaire. Les scores d'anomalie retournés pour chacun des scénarios par les algorithmes ADDICT et C-ADDICT sont représentés dans les figures 4.10 et 4.11. Nos commentaires et conclusions concernant les résultats obtenus par l'algorithme ADDICT et C-ADDICT pour les différents scénarios sont les suivants :

- Scénario 1 :
  Les scores retournés par les algorithmes ADDICT et C-ADDICT pour le premier scénario sont bien supérieurs à la moyenne exclusivement en période de manoeuvre ou d'anomalie. Ces résultats témoignent de la capacité des méthodes proposées à détecter tout nouveau comportement atypique dans les données de télémesure.

- Scénario 2 :
  Pour le deuxième scénario, la première période de manoeuvre a été intégrée au dictionnaire construit pour le premier scénario. Au regard des scores d'anomalie retournés à partir du dictionnaire actualisé il apparaît que cette manoeuvre n'est plus détectée. Par ailleurs, d'après les résultats de détection Online C-ADDICT (i.e, C-ADDICT avec un dictionnaire mis à jour), il est possible de fixer un seuil de détection permettant de détecter l'anomalie multivariée sans retourner de fausses alarmes pour les manoeuvres suivantes. Ce n'est pas le cas avec l'algorithme Online ADDICT étant donné des scores encore élevés retournés en périodes de manoeuvre. Ces résultats traduisent le fait que la prise en compte d'une seule manoeuvre n'est pas suffisant pour empêcher la détection des manoeuvres suivantes par l'algorithme ADDICT.

- Scénario 3 :
  Dans ce troisième scénario réalisé pour l'algorithme ADDICT, la première et la deuxième période de manoeuvre ont été intégrées au dictionnaire. On observe que ces deux périodes de manoeuvre ne sont plus détectées par l'algorithme. Par ailleurs, il apparaît que l'intégration de ces deux périodes de manoeuvre a permis de faire chuter les scores d'anomalie associés aux périodes de manoeuvres suivantes. Il est alors possible de fixer un seuil de détection permettant de détecter uniquement l'anomalie multivariée.

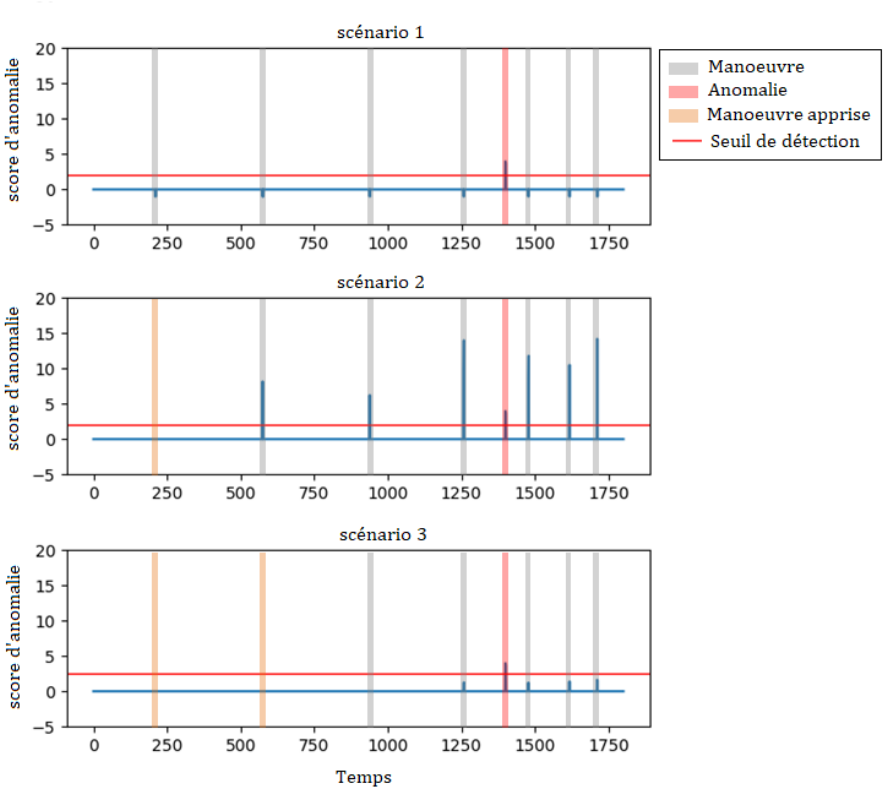

FIGURE 4.10 – Comparaison des scores d'anomalie retournés par l'algorithme ADDICT et sa version "en ligne", Online ADDIC,T appliqués selon plusieurs scénarios aux données de télémesure associées au cas d'application des manoeuvres de correction d'orbite.

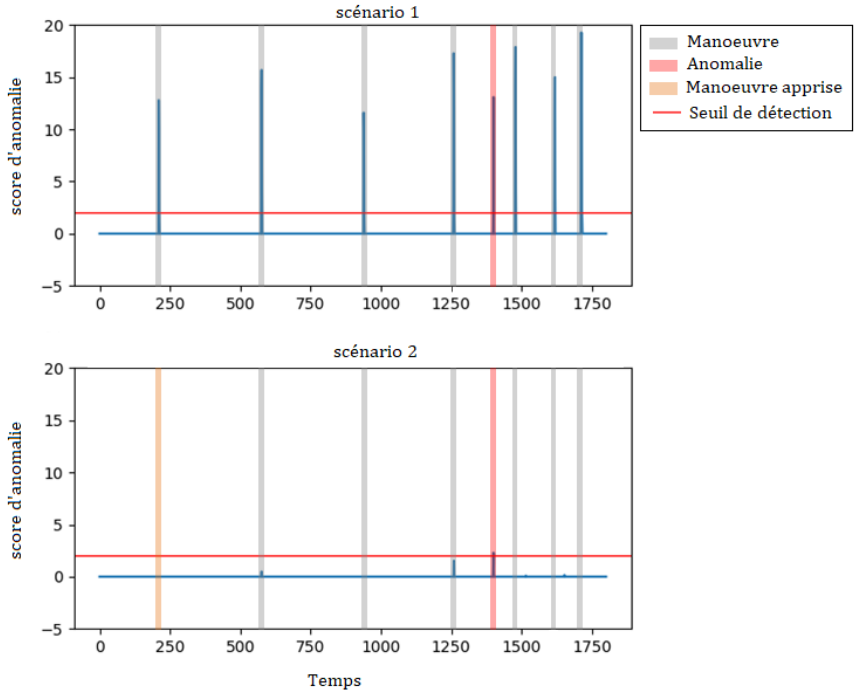

FIGURE 4.11 – Comparaison des scores d'anomalie retournés par l'algorithme C-ADDICT et sa version "en ligne", Online C-ADDICT, appliqués selon plusieurs scénarios aux données de télémesure associées au cas d'application des manoeuvres de correction d'orbite.

Les résultats présentés dans cette section témoignent de l'intérêt de traiter conjointement plusieurs paramètres de télémesure discrets et continus et de prendre en compte le contexte opérationnel des satellites afin de limiter les fausses alarmes liées à des évènements ou à un contexte opérationnel connu. Par ailleurs, ces résultats démontrent l'intérêt des méthodes d'estimation parcimonieuse pour la détection séquentielle. En effet, notons qu'avec l'algorithme ADDICT l'intégration des deux premières manoeuvres seulement a permis d'empêcher la détection des 5 manoeuvres suivantes. Pour l'algorithme C-ADDICT, cet objectif a été atteint dès l'intégration de la première manoeuvre.

4.4 Surveillance de la télémesure en période d'éclipse

4.4.1 Présentation du cas d'usage

Ce troisième cas d'application étudie l'intérêt de l'apprentissage séquentiel de nouveaux modes de fonctionnement des satellites. Plus précisément, l'exemple traité dans cette section concerne l'intégration du mode de fonctionnement en période d'éclipse d'un satellite géostationnaire. Cette étude est réalisée à partir des données de télémesure de deux paramètres continus acquises sur une durée totale ayant connue 19 périodes d'éclipse. Les données

traitées dans cette section sont représentées dans la figure 4.12 et sont issues d'un satellite géostationnaire conçu par Airbus. Les périodes associées au nouveau mode de fonctionnement (mode éclipse) sont repérables par leur fond gris dans la figure 4.12. De même que pour le cas d'application précédent, nous avons fait le choix pour cet exemple de tester les algorithmes ADDICT et C-ADDICT capables de traiter conjointement plusieurs paramètres de télémesure et d'intégrer facilement de nouveaux évènements par l'ajout d'éléments dans le dictionnaire. Ces deux méthodes ont été appliquées aux données représentées dans la figure 4.12 selon plusieurs scénarios. Les résultats obtenus sont présentés dans la section suivante.

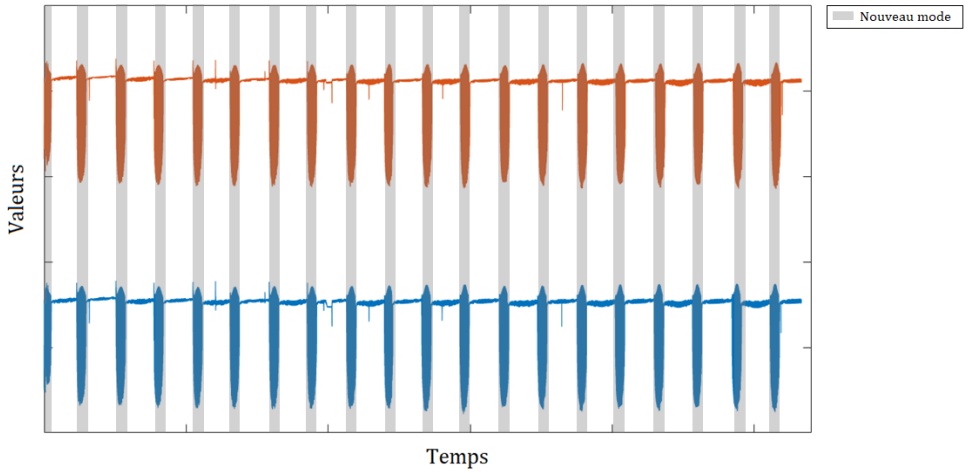

FIGURE 4.12 – Télémesure de 2 paramètres continus représentant deux modes de fonctionnement du satellite.

4.4.2 Résultats expérimentaux

Le protocole de test suivi pour ce cas d'application est le même que celui du cas d'application précédent. Dans un premier temps, un dictionnaire a été appris à partir de données de télémesure associées à un unique mode de fonctionnement connu du satellite (hors éclipse). Ce premier scénario consiste à évaluer la capacité des méthodes testées à détecter le nouveau mode de fonctionnement. Par la suite, le fonctionnement du satellite en mode éclipse a été intégré au dictionnaire.
Les scores d'anomalie retournés par les algorithmes ADDICT et C-ADDICT pour les différents scénarios sont présentés dans les figures 4.13 et 4.14. Ces résultats expérimentaux mènent aux conclusions suivantes :

- Les scores d'anomalie retournés pour le premier scénario témoignent de la capacité des deux méthodes proposées à détecter le nouveau mode de fonctionnement satellite. En effet des scores supérieurs à la moyenne ont été retournés par les deux méthodes presque exclusivement en période d'éclipse ce qui traduit une détection efficace.

- Les scores d'anomalie retournés pour les deuxième et troisième scénarios démontrent qu'une détection "en ligne" à l'aide d'une méthode d'estimation parcimonieuse est possible.

- Comme pour le cas précédant la détection "en ligne" est plus efficace avec l'algorithme C-ADDICT qu'avec l'algorithme ADDICT. En effet, la prise en compte d'une seule période d'éclipse suffit à empêcher la détection des prochaines occurrences tandis que deux périodes d'éclipse sont nécessaires pour atteindre cet objectif avec l'algorithme ADDICT.

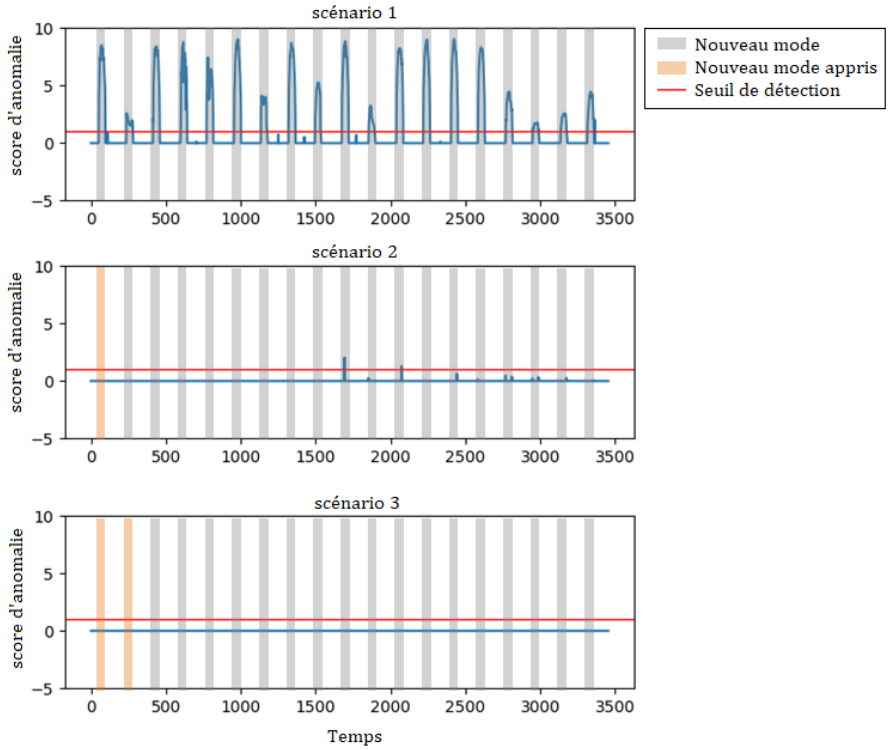

FIGURE 4.13 – Comparaison des scores d'anomalie retournés par l'algorithme ADDICT et sa version "en ligne", Online ADDICT appliqués selon plusieurs scénarios aux données de télémesure associées au cas d'application des nouveaux mode de fonctionnement satellite.

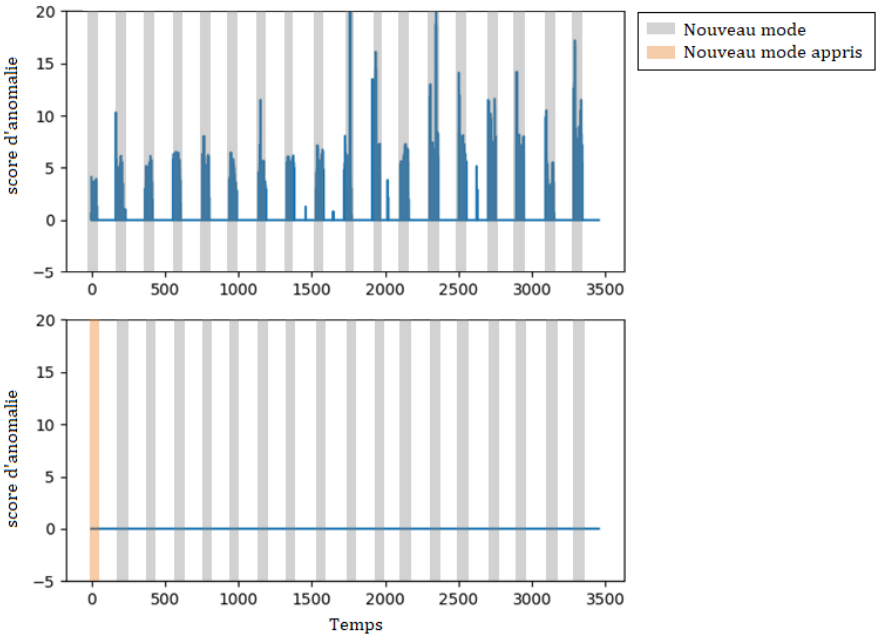

FIGURE 4.14 – Comparaison des scores d'anomalie retournés par l'algorithme C-ADDICT et sa version "en ligne", Online C-ADDICT appliqués selon plusieurs scénarios aux données de télémesure associées au cas d'application des nouveaux mode de fonctionnement satellite

4.5 Détection d'évènements dans la télémesure satellite

4.5.1 Présentation du cas d'usage

Ce dernier cas d'usage étudie l'intérêt des méthodes de détection d'anomalies multivariées développées pendant la thèse pour détecter, et pourquoi pas anticiper, l'apparition d'un évènement particulier, connu, mais imprévisible, et qui peut affecter le fonctionnement d'un satellite. Cette étude a été menée à partir de données réelles de télémesure issues de 3 satellites conçus par Airbus. Pour chacun d'eux, la télémesure d'une vingtaine de paramètres sélectionnés par des experts décrivant un fonctionnement courant du satellite sur 100 orbites consécutives a été mise à disposition pour l'apprentissage du dictionnaire. L'évaluation des algorithmes a été réalisée à partir de la télémesure de ces mêmes paramètres pour 100 orbites consécutives durant lesquelles l'évènement étudié s'est produit et a été identifié par les experts.
Un extrait de la télémesure des paramètres considérés pour chacun des trois satellites est représenté dans les figures 4.15, 4.16 et 4.17. Notons que les paramètres traités ne sont pas nécessairement les mêmes pour chacun des trois satellites. Au regard de ces figures, l'apparition de l'évènement, marqué par un fond rouge, se traduit par un changement de mode de fonctionnement de quelques équipements (paramètres discrets) et à son maintien prolongé

(e.g., figure 4.15 paramètres 1 à 9, figure 4.16 paramètre 1, figure 4.17 paramètre 5), ou par une chute brutale et/ou une stabilisation (variance nulle) des valeurs enregistrées pour certains paramètres continus (e.g., figure 4.15 paramètres 13 à 19, figure 4.16 paramètres 11 à 19, figure 4.17 paramètres 11 à 19).

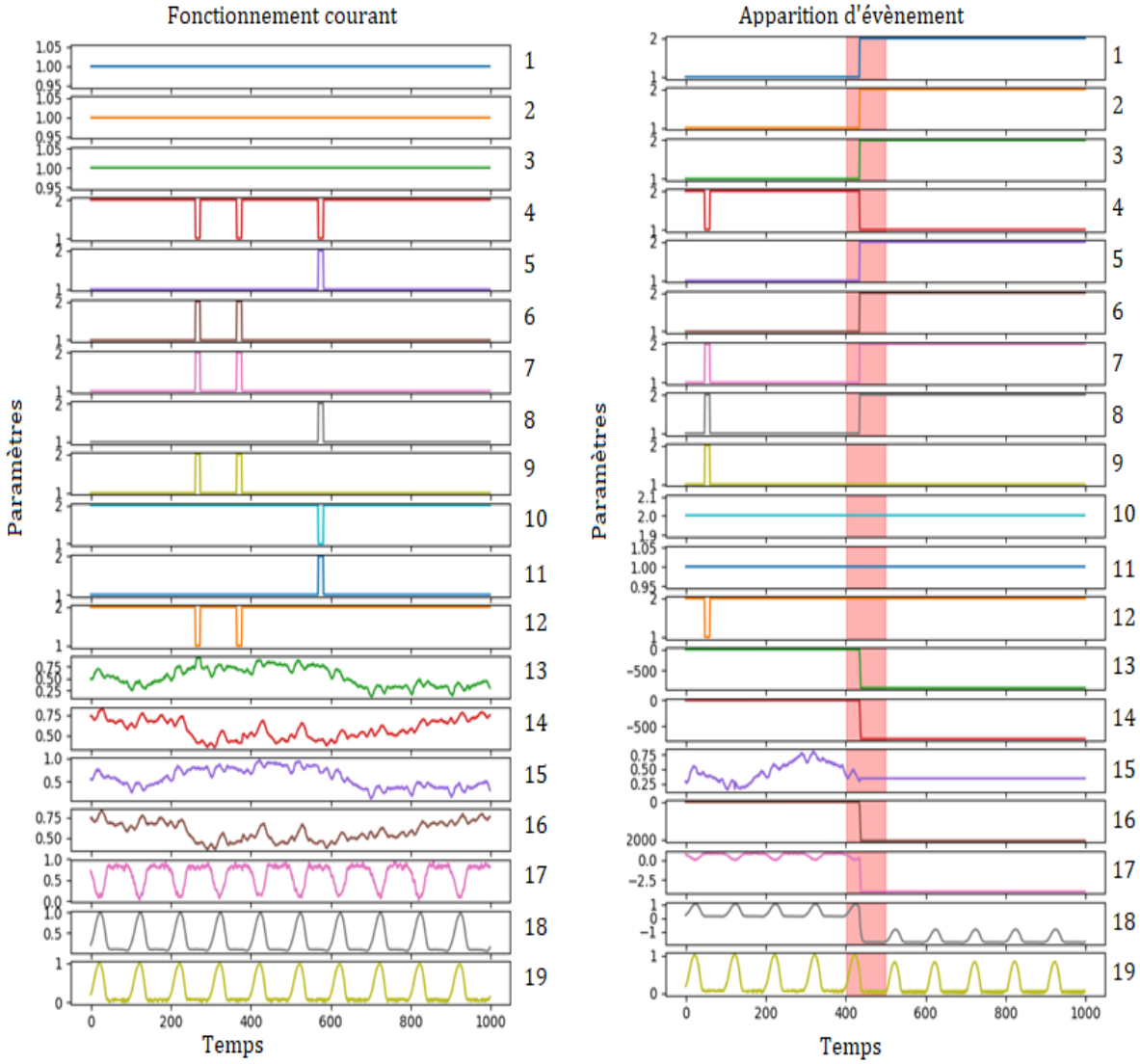

FIGURE 4.15 – Télémesure de 19 paramètres dont 12 discrets et 7 continus associée au premier satellite étudié. La télémesure décrit le fonctionnement courant du satellite sur 10 orbites (gauche) et le fonctionnement du satellite sur 10 orbites autour de l'évènement (droite). Le début de l'évènement est marqué par un fond rouge.

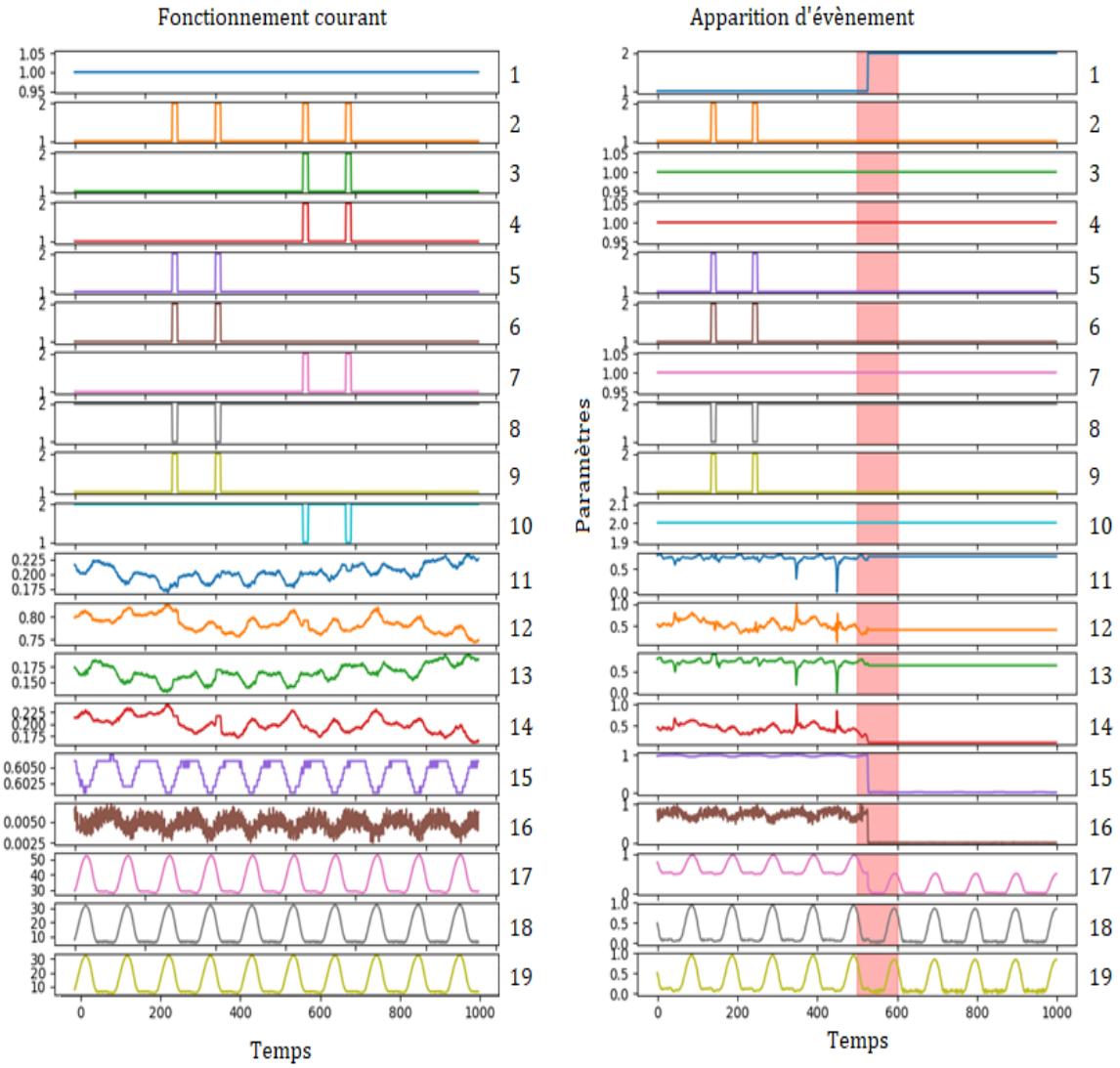

FIGURE 4.16 – Télémesure de 19 paramètres dont 10 discrets et 9 continus associée au deuxième satellite étudié. La télémesure décrit le fonctionnement courant du satellite sur 10 orbites (gauche) et le fonctionnement du satellite sur 20 orbites autour de l'évènement (droite). Le début de l'évènement est marqué par un fond rouge.

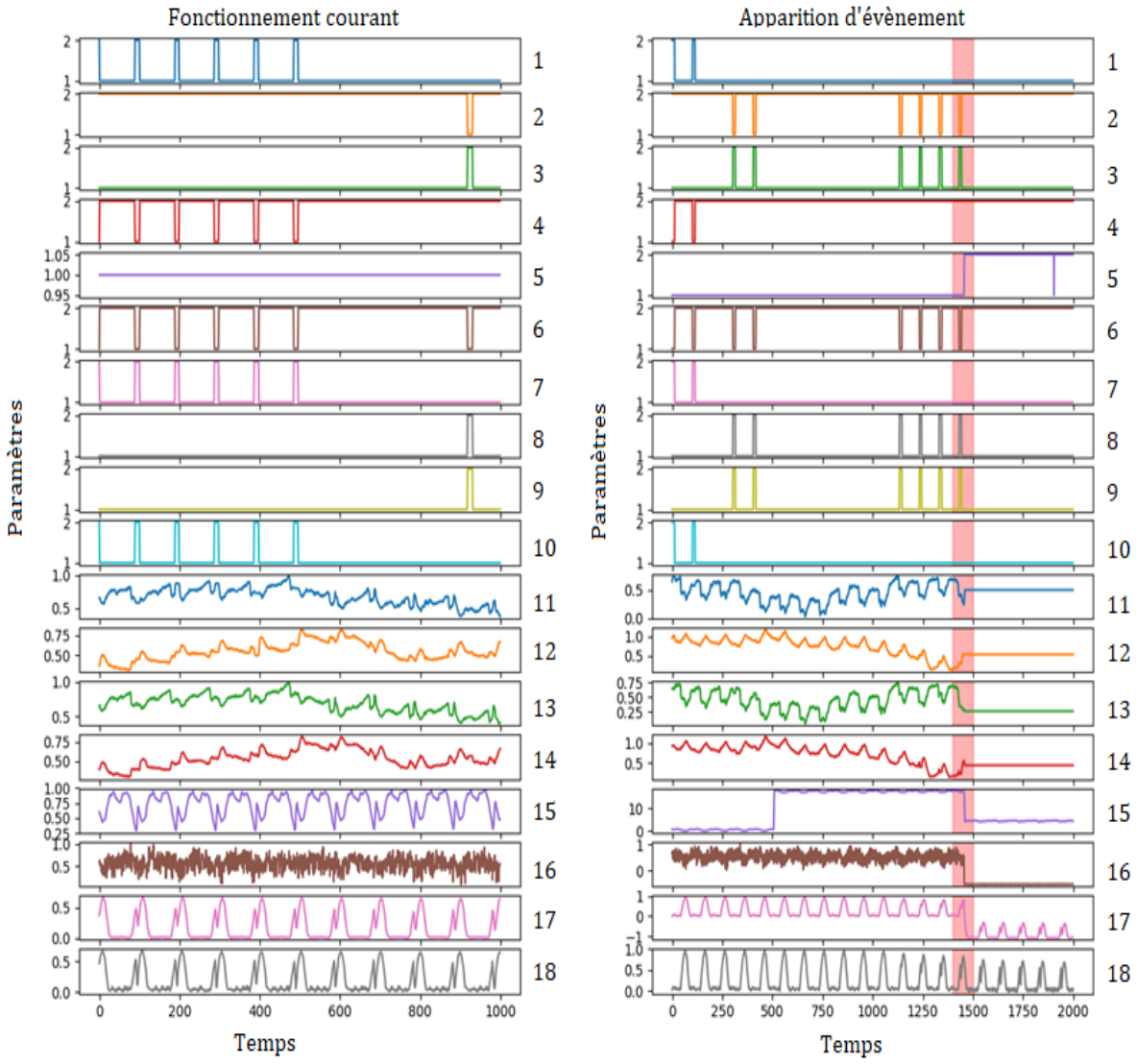

FIGURE 4.17 – Télémesure de 18 paramètres dont 10 discrets et 8 continus associée au troisième satellite étudié. La télémesure décrit le fonctionnement courant du satellite sur 10 orbites (gauche) et le fonctionnement du satellite sur 20 orbites autour de l'évènement (droite). Le début de l'évènement est marqué par un fond rouge.

4.5.2 Détection d'évènement

Pour ce cas d'usage, les algorithmes ADDICT et C-ADDICT ont été appliqués aux données présentées dans la section précédente qui ont été rééchantillonnées afin de limiter la taille des signaux multivarié à 100 mesures par paramètre et par orbite. Pour l'algorithme ADDICT la télémesure a été découpée en fenêtres de taille $W = 100$ (voir section 2.3.1 pour plus de détails sur le pré-traitement) et les dictionnaires construits pour chaque satellite comptent 8000 atomes de taille $N = 100K$ (où K est le nombre de paramètres étudiés). Le dictionnaire

C-ADDICT compte 200 filtres de taille $M = 100$. Rappelons que les différents dictionnaires ont été construits à partir de données de télémesure associées à un fonctionnement courant du satellite observé pendant 100 orbites consécutives. Pour chaque orbite, un score d'anomalie ADDICT et un score d'anomalie C-ADDICT est retourné et représenté dans les figures 4.18, 4.19 et 4.20. L'orbite lors de laquelle a débuté l'évènement selon les experts est marquée par un fond rouge. Dans les trois cas, on observe que les scores retournés par les deux méthodes à l'apparition de l'évènement et pour les orbites suivantes sont bien supérieurs à ceux retournés pour les premières orbites ce qui témoigne d'une détection efficace de l'apparition de l'évènement. Par ailleurs, on s'aperçoit que des scores plus élevés que la moyenne des précédents ont été retournés avant le début de l'évènement pour le deuxième et le troisième satellites. Ces scores témoignent peut-être d'un changement de fonctionnement léger et précoce du satellite du fait de l'apparition imminente de l'évènement. Pour aller plus loin dans l'investigation et tenter de mieux comprendre l'évènement étudié, il convient de s'intéresser aux scores univariés et d'identifier les paramètres qui ont pu influencer les scores d'anomalies multivariés des figures 4.19 et 4.20.

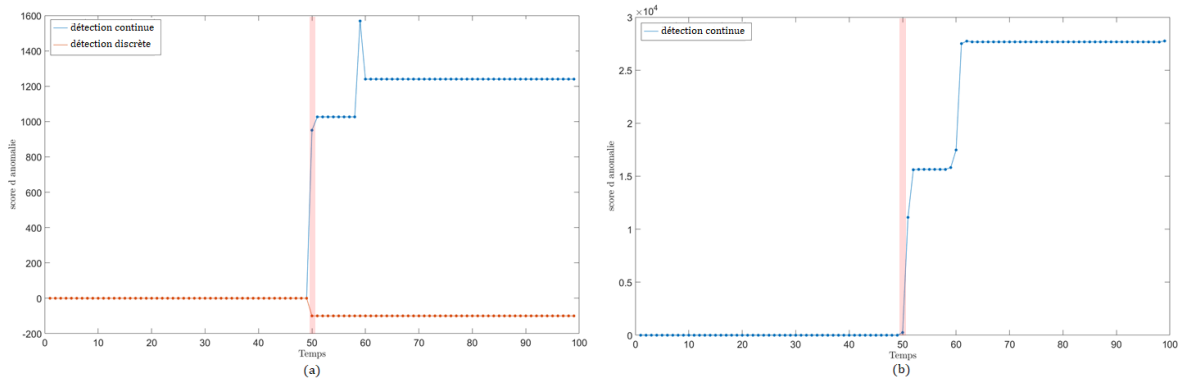

FIGURE 4.18 – Comparaison des scores retournés par ADDICT (a) et C-ADDICT (b) pour les signaux de télémesure issus du premier satellite étudié.

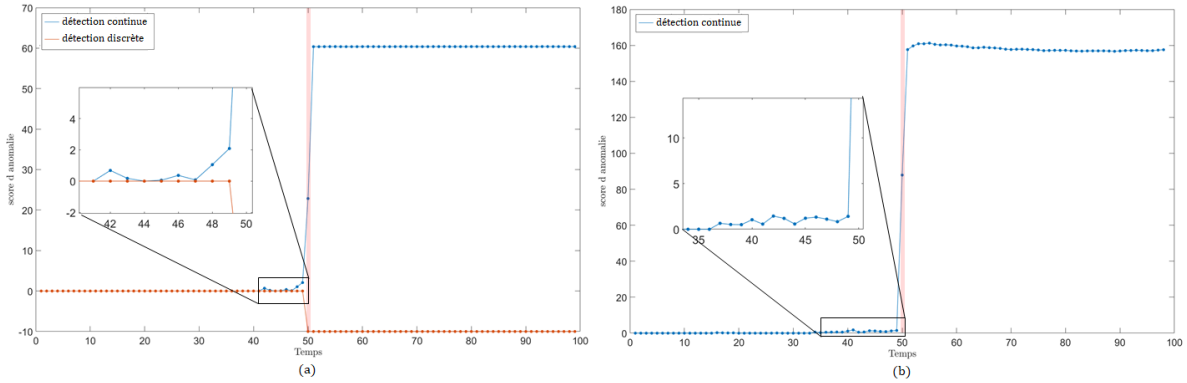

FIGURE 4.19 – Comparaison des scores retournés par ADDICT (a) et C-ADDICT (b) pour les signaux de télémesure issus du deuxième satellite étudié.

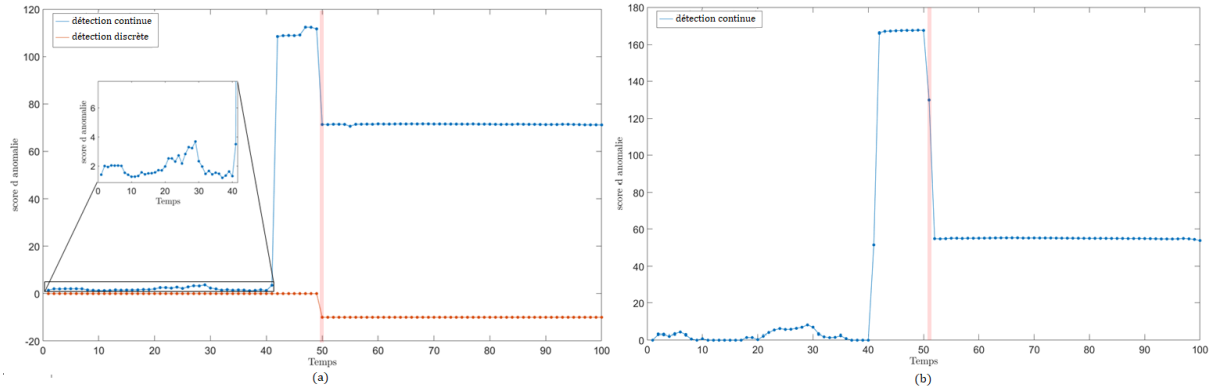

FIGURE 4.20 – Comparaison des scores retournés par ADDICT (a) et C-ADDICT (b) pour les signaux de télémesure issus du troisième satellite étudié.

4.5.3 Détection univariée

L'étude des scores univariés permet d'identifier immédiatement les paramètres affectés par une anomalie ou un évènement particulier. Les scores d'anomalies univariées retournés par l'algorithme ADDICT pour les signaux de test issus du deuxième et du troisième satellite sont représentés dans les figures 4.21 et 4.22. Il apparaît que des scores supérieurs à la moyenne sont observés pour certains paramètres à l'apparition de l'évènement (fonds rouges) et sur les orbites suivantes. Ces paramètres correspondent aux paramètres pour lesquels un changement de comportement est visible dans les figures 4.16 et 4.17 (droite, apparition d'évènement). Ces résultats témoignent une fois de plus de la capacité de l'algorithme ADDICT à identifier les paramètres affectés. Par ailleurs, on observe que pour le deuxième satellite des scores élevés sont enregistrés pour les paramètres #11 et #13. L'étude de la reconstruction des signaux de test par décomposition sur le dictionnaire a mis en évidence le fait que ces résultats étaient dus aux valeurs extrêmes mesurées pour ces deux paramètres sur les orbites qui précèdent l'évènement (voir figure 4.16, paramètres #11 et #13, droite, apparition d'évènement). Pour le troisième satellite, des scores supérieurs à la moyenne sont enregistrés pour tous les paramètres continus, quelques orbites avant l'apparition de l'évènement. Cependant les scores les plus élevés sont retournés pour le paramètre #15. Le signal de test de ce paramètre est représenté dans la figure 4.17 et décrit une hausse brutale de ses valeurs juste avant l'apparition de l'évènement ce qui explique les scores retournés. Les scores élevés retournés pour les autres paramètres continus sont dus à l'estimation parcimonieuse réalisée conjointement pour les différents paramètres continus et/ou peuvent témoigner du fait que les différents paramètres sont corrélés ou liés entre eux.
Ces expérimentations ont permis d'identifier quelques paramètres susceptibles d'indiquer l'apparition de l'évènement étudié. Pour confirmer ces résultats, une étude plus approfondie et une investigation de la part des experts est nécessaire. Par ailleurs, notons que les changements détectés dans les signaux de test de ces paramètres sont à priori univariés. Ces expérimentations n'ont pas permis de démontrer l'intérêt d'une méthode de détection d'anomalies multivariées pour ce cas d'usage.

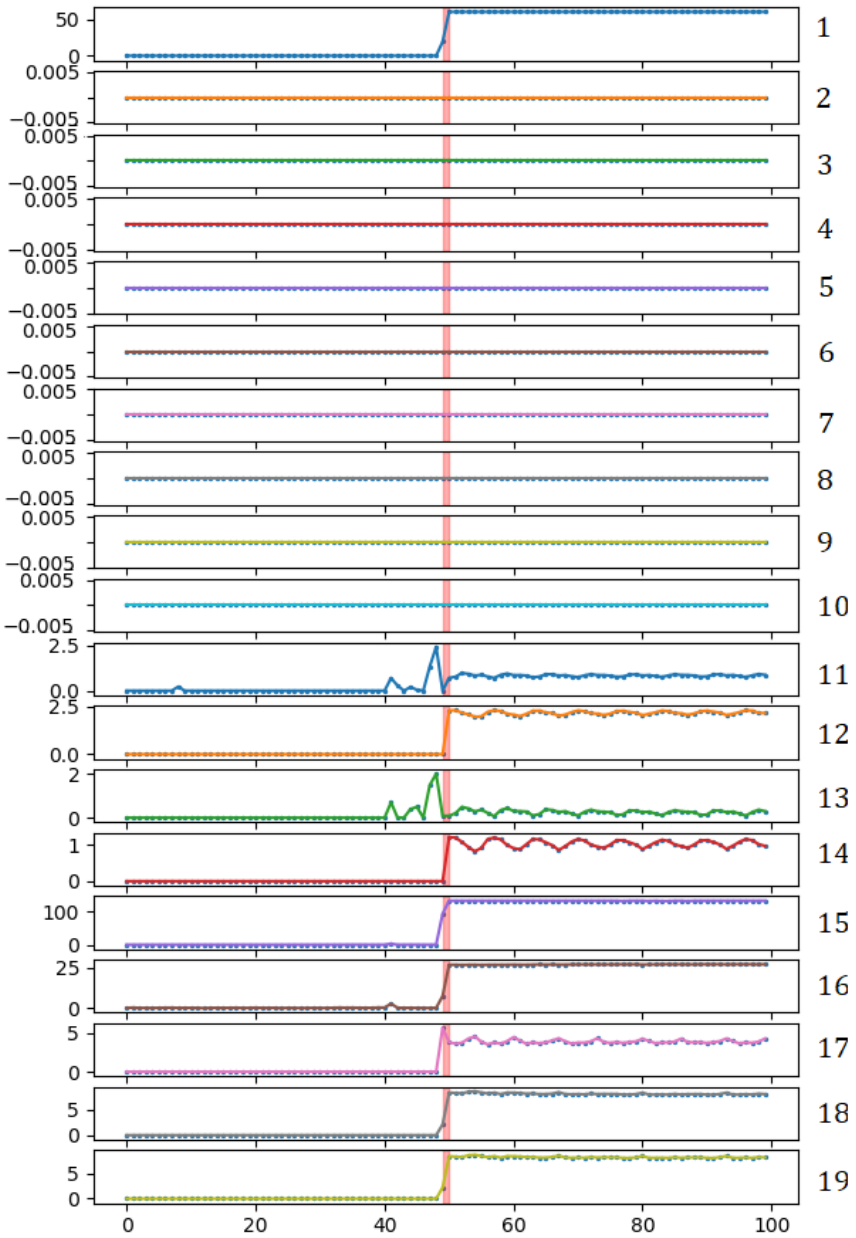

FIGURE 4.21 – Scores d'anomalies univariés retournés par l'algorithme ADDICT pour les signaux de télémesure issus du deuxième satellite.

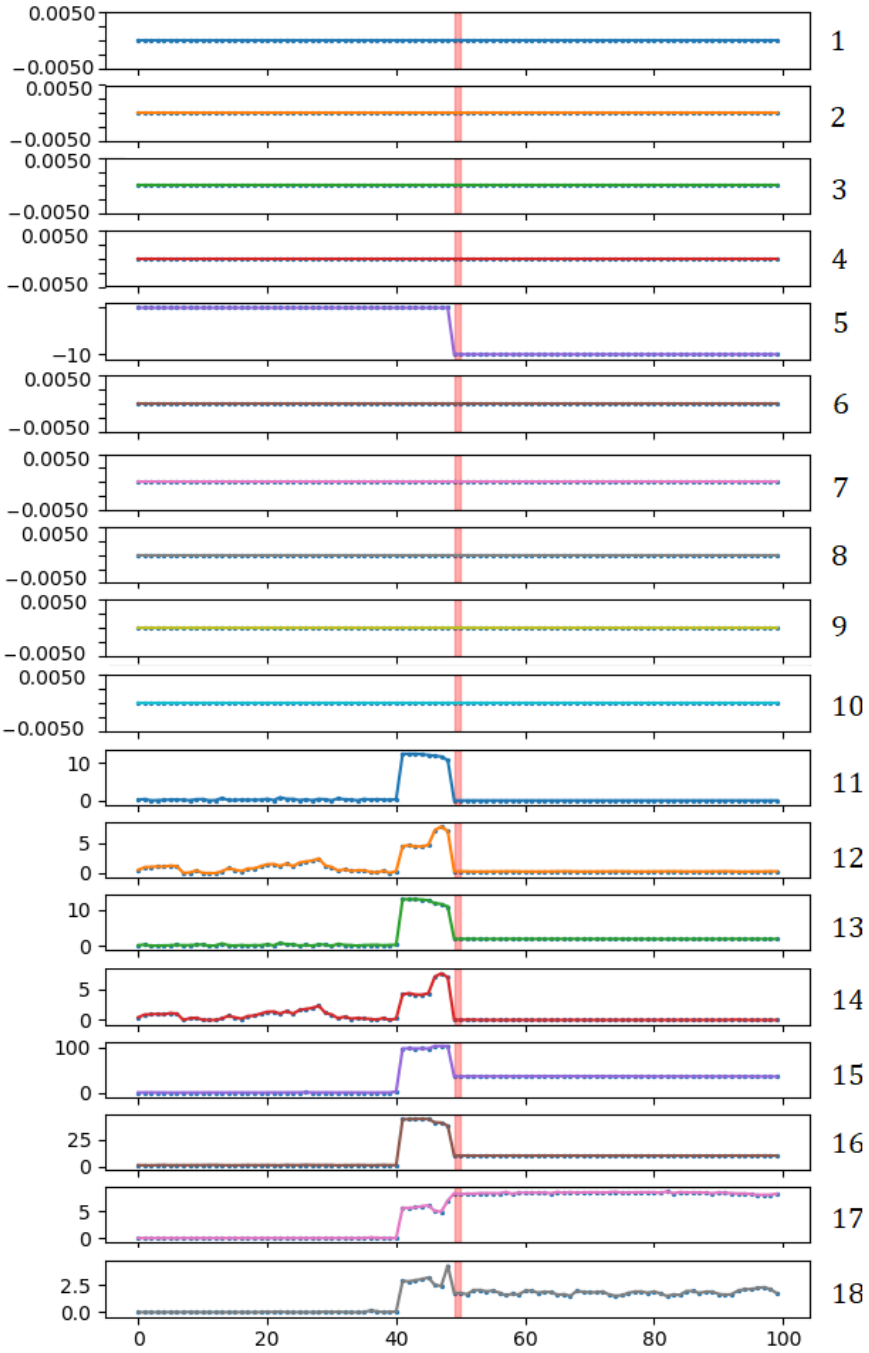

FIGURE 4.22 – Scores d'anomalies univariés retournés par l'algorithme ADDICT pour les signaux de télémesure issus du troisième satellite.

4.6 Conclusion

Ce chapitre nous plonge au coeur du contexte opérationnel dans lequel s'inscrit cette thèse et soulève les enjeux métier qui y sont liés telles que la détection d'anomalies multivariés avec un nombre important de signaux de télémesure ou la limitation des fausses alarmes qui engendrent des vérifications très couteuses. Les différents cas d'application traités dans ce chapitre sont autant d'occasions d'évaluer les différentes méthodes proposées dans cette thèse et les premiers résultats expérimentaux obtenus confirment l'intérêt des méthodes d'estimation parcimonieuse pour la détection d'anomalies, et plus particulièrement de l'algorithme ADDICT et de ses extensions. En effet, les performances enregistrées sur une base composée d'anomalies diverses munie de sa vérité terrain, ou sur un ensemble de données décrivant un contexte opérationnel de manoeuvres satellite, témoignent de la force de ces approches à traiter conjointement de nombreux paramètres de télémesure pour détecter différents types d'anomalies et limiter les fausses alarmes. Par ailleurs, ces expérimentations mettent en lumière un réel avantage de ces méthodes, à savoir leur capacité à intégrer le retour utilisateur de manière séquentielle pour prévenir les fausses alarmes. La stratégie naïve de détection "en ligne" ADDICT testée ici consiste à ajouter de nouveaux éléments au dictionnaire suite à l'identification de fausses alarmes. Les premiers résultats obtenus à l'aide de cette méthode sont satisfaisants mais quelques limites apparaissent. En effet, l'ajout des fausses alarmes dans le dictionnaire ne garantit pas qu'une nouvelle occurrence de ces dernières ne sera pas détectée. Par ailleurs, cette stratégie fait croître le nombre d'éléments du dictionnaire et la complexité de calcul de l'algorithme. Pour faire suite à ces travaux, plusieurs perspectives pourraient alors être envisagées :

- Faute de temps, la recherche d'une stratégie de mise à jour du dictionnaire réellement efficace n'a pu être menée mais mériterait d'être étudiée pour permettre une meilleure prévention des fausses alarmes connues et un contrôle de la taille du dictionnaire. Des stratégies d'apprentissage "en ligne" de dictionnaire telle que celle proposée dans Mairal et al. (2009) pourraient être adaptées pour répondre à ce problème.

- En plus de la mise à jour des éléments du dictionnaire, on pourrait envisager un réglage séquentiel des hyperparamètres du modèle et/ou du seuil de détection comme cela a pu être étudié dans Raginsky et al. (2012).

Conclusion et perspectives

Cette thèse avait pour objectif de proposer une méthode de détection d'anomalies permettant d'améliorer le suivi de l'état de santé de satellites et plus précisément la surveillance de la télémesure. Le premier chapitre décrit la télémesure satellite et présente une méthode classique de surveillance par seuil ainsi que quelques algorithmes récents de détection d'anomalies univariées basés sur des techniques d'apprentissage semi-supervisé. L'objectif des chapitres suivants est d'étudier des méthodes de détection d'anomalies multivariées capables de traiter conjointement plusieurs signaux de télémesure mixtes (discrets et continus), et de prendre en compte les corrélations et relations qui peuvent exister entre eux.
Le deuxième chapitre formule le problème de détection d'anomalies multivariées comme un problème d'estimation parcimonieuse et présente l'algorithme ADDICT développé pour le résoudre. La stratégie de détection adoptée consiste en un apprentissage semi-supervisé et repose sur l'hypothèse que la nouvelle télémesure décrivant un fonctionnement normal du satellite peut être estimée en utilisant quelques signaux sains de la télémesure passée (sans anomalie). Plus précisément, l'idée est de se constituer une base, appelée dictionnaire et construit à partir de données saines, et de formuler le problème de détection d'anomalies comme un problème d'estimation parcimonieuse qui décrit la nouvelle donnée par une combinaison linéaire d'éléments du dictionnaire dont le nombre de coefficients non nuls du vecteur à estimer est contraint par une norme ℓ_1. Une seconde contrainte de parcimonie est également introduite pour traduire le fait que les anomalies satellites sont rares et affectent peu de paramètres en même temps. Cette parcimonie structurée, assurée par une norme ℓ_2, permet notamment de pouvoir identifier les paramètres affectés par une anomalie et ceux qui ne le sont pas. L'algorithme ADDICT proposé a été testé sur des données réelles de télémesure satellite munies d'une vérité terrain composée d'anomalies diverses. Les résultats obtenus démontrent l'intérêt et la compétitivité de la méthode proposée par rapport aux approches de l'état de l'art. Ce chapitre présente également plusieurs extensions telles qu'une généralisation du problème posé afin de s'adapter au contexte multivarié et à l'hétérogénéité des données traitées, et une version pondérée permettant d'intégrer de l'information externe au modèle et d'améliorer les performances de détection. La pondération étudiée a été construite à partir du coefficient de corrélation calculé entre le signal testé et sa reconstruction et a permis de limiter le nombre de fausses alarmes retournées. Enfin, ce chapitre s'intéresse au réglage des hyperparamètres du modèle et présente une méthode d'estimation qui semble prometteuse.
Le deuxième chapitre propose de remplacer le dictionnaire fixe du modèle ADDICT par un dictionnaire convolutif afin d'exploiter ses propriétés d'invariance par translation. La combinaison linéaire peut alors être réécrite comme une somme de produits de convolution entre des cartes de coefficients et les éléments du dictionnaire convolutif. Cette extension convolutive de l'algorithme ADDICT, appelée C-ADDICT, a permis d'améliorer de manière significative les performances de l'algorithme ADDICT. Par ailleurs, le problème d'estimation parcimonieuse convolutive proposé a pu être étendu au problème d'apprentissage de dictionnaire. Cependant, l'introduction d'un dictionnaire convolutif a complexifié le problème rendant sa résolution plus coûteuse.

Le quatrième chapitre met en application les différentes méthodes proposées dans cette thèse pour répondre à de réels besoins industriels de détection d'anomalies multivariées avec un nombre important de signaux de télémesure, ou de limitation des fausses alarmes. Pour ce faire, un passage à l'échelle a été réalisé à partir de données de télémesure satellite réelles associées à une vérité terrain composée d'anomalies connues et identifiées par des experts. Ces expérimentations ont permis de démontrer la capacité de l'algorithme ADDICT proposé à traiter conjointement plusieurs dizaines de signaux de télémesure et y détecter efficacement les anomalies. En ce qui concerne la limitation des fausses alarmes, l'intérêt d'une version séquentielle ou "en ligne" a été étudié. Une méthode naïve de mise à jour du dictionnaire par ajout séquentiel des fausses alarmes a été testée pour limiter la présence de fausses alarmes liées à des opérations de manoeuvre de correction d'orbite ou à l'entrée du satellite en mode éclipse par exemple. Les premiers résultats obtenus sont encourageants mais certaines fausses alarmes sont toujours détectées et la complexité de l'algorithme n'est pas maîtrisée et augmente au fur et à mesure des mises à jour, ce qui peut, à terme, être problématique.

Plusieurs perspectives peuvent être envisagées pour faire suite à ces travaux. Le premier chapitre formule le problème de détection d'anomalies multivariées comme un problème d'estimation parcimonieuse dans lequel l'estimation des données discrètes et des données continues est réalisée selon deux stratégies distinctes. Le modèle proposé pour l'estimation parcimonieuse discrète mène à un problème combinatoire dont la résolution est coûteuse. Il serait alors intéressant d'essayer de se ramener à un problème de minimisation plus simple à résoudre. Les signaux de télémesure discrets peuvent être vus comme des signaux "clippés" ou numérisés pour lesquels des modèles d'estimation parcimonieuse et d'apprentissage de dictionnaires ont été étudiés Rencker et al. (2019, 2018). Il serait alors intéréssant de pouvoir adapter ces derniers pour le traitement de données mixtes. Une autre piste qui pourrait également être explorée consiste à adopter une stratégie d'estimation parcimonieuse discrète à l'aide d'un modèle probabiliste permettant de contraindre les coefficients de la décomposition à prendre des valeurs discrètes Hennings et al. (2010); Exarchakis and Lücke (2017). Par ailleurs, s'affranchir d'une résolution combinatoire pourrait permettre de traiter conjointement les données continues et les données discrètes et de résoudre un unique problème de minimisation. Le problème de détection d'anomalies pourrait alors être plus facilement étendu au problème d'apprentissage de dictionnaire qui n'a pas été adressé dans ce chapitre et représente un point important à traiter.

Les algorithmes proposés dans les chapitres 2 et 3 reposent sur des modèles d'estimation parcimonieuse qui imposent de régler deux hyperparamètres. Une méthode basée sur la théorie des machines à vecteurs support a été proposée et permet d'estimer efficacement l'un d'entre eux. Pour aller plus loin, on pourrait envisager d'étudier d'autres techniques d'estimation d'hyperparamètres, telles les approches d'estimation Bayesienne, et d'essayer de trouver une méthode capable d'estimer conjointement plusieurs hyperparamètres Mohammad-Djafari (1993). Par ailleurs, la détection d'anomalies suppose de fixer un seuil de détection auquel sont comparés les scores retournés par l'algorithme de manière à détecter une anomalie lorsque la score excède le seuil fixé. Ce seuil représente un compromis entre le taux de bonnes détections

exigé et le taux de fausses alarmes accepté. Cependant, le score d'anomalie ADDICT n'est pas borné ce qui le rend difficile à interpréter (e.g., un score d'anomalie compris entre 0 et 1 peut être interprété comme la probabilité qu'une anomalie se produise) et donc à régler de manière intuitive dans le respect des contraintes métiers. Si une vérité terrain est disponible, les performances de l'algorithme peuvent être calculées pour différentes valeurs du seuil et il est alors possible d'en déduire sa valeur optimale pour les données traitées. Cependant, dans un contexte opérationnel, il est rare qu'une vérité terrain existe. Afin de contourner ce problème de réglage du seuil de détection, on pourrait envisager adopter une approche de "prédiction conforme" (conformal prediction en anglais) Laxhammar and Falkman (2011); Shafer and Vovk (2008); Gammerman and Vovk (2006) qui s'inscrit dans une logique de tests statistiques et consiste à estimer une p-valeur (i.e., la probabilité d'observer un score égal ou supérieur à celui retourné, sous l'hypothèse que le signal testé est associé à un fonctionnement normal du satellite) pour chaque signal testé. La détection est alors réalisée à partir de cette p-valeur estimée (comprise entre 0 et 1) qui est comparée à un seuil de sorte qu'une anomalie est détectée si elle est supérieure à ce seuil. Ce nouveau seuil peut alors être interprété comme une probabilité de fausses alarmes autorisées et peut être fixé selon les contraintes opérationnelles. La mise en place d'une stratégie de prédiction conforme permet alors de s'affranchir de la construction d'une vérité terrain pour fixer la valeur du seuil de détection et de maitriser le nombre de fausses alarmes retournées ce qui représente un enjeux important.

Le dernier chapitre évalue les algorithmes proposés dans de réels contextes industriels de surveillance de la télémesure. Ces applications ont mis en évidence les limites des méthodes proposées qu'il serait intéressant d'essayer de surmonter. Tout d'abord, le passage à l'échelle a permis de quantifier en termes de taille du dictionnaire et de temps de calcul le niveau de complexité de l'algorithme lorsque le nombre de signaux de télémesure considérés augmente. Afin de pouvoir traiter conjointement de nombreux paramètres de télémesure tout en limitant la compléxité, une première idée serait de réaliser un pré-traitement adapté pour réduire la dimension de séries temporelles Barreyre (2018). Par ailleurs, ce chapitre a testé une approche naïve de mise à jour séquentielle du dictionnaire permettant de limiter les fausses alarmes. Cependant, la stratégie adoptée ne permet pas d'éliminer toutes les fausses alarmes connues ni de contrôler la taille du dictionnaire. La recherche d'une méthode plus efficace de mise à jour du dictionnaire mériterait donc d'être menée afin d'actualiser les éléments du dictionnaire et/ou d'en ajouter de nouveaux lorsque cela est nécessaire. Certains algorithmes d'apprentissage de dictionnaire de la littérature pourraient être adaptés dans ce sens Mairal et al. (2009); Aharon et al. (2006); Barthélemy et al. (2013). De plus, un réglage séquentiel des hyperparamètres du modèle et/ou du seuil de détection pourrait également être envisagé Raginsky et al. (2012) pour limiter la présence des fausses alarmes, prendre en compte l'évolution du fonctionnement du satellite, le vieillissement des équipements etc.

Annexe A

Algorithmes d'apprentissage de dictionnaires

A.1 L'algorithme K-SVD

L'algorithme K-SVD, proposé par Aharon et al. (2006), est un algorithme d'apprentissage de dictionnaire inspiré du célèbre algorithme de classification K-Means. L'objectif de l'algorithme K-Means est de former K groupes homogènes en affectant tous les éléments d'un nuage de points à un groupe. Une fois placés dans un groupe, chaque élément est alors représenté par l'élément moyen du groupe auquel il appartient. L'algorithme K-SVD peut être vu comme une généralisation du K-Means dans le sens où l'on cherche ici à représenter des signaux multivariés, non pas par un élément (l'élément moyen), mais par plusieurs éléments (les atomes du dictionnaire). A chaque itération du K-Means, les éléments moyens qui représentent chaque groupe sont actualisés selon les réaffectations. Par analogie, à chaque itération de K-SVD, les atomes sont actualisés selon leur intervention dans la décomposition des signaux d'apprentisssage.
Nous noterons $\mathbf{Y} \in \mathbb{R}^{N \times V}$ la matrice des signaux d'apprentissage, et $\boldsymbol{X} \in \mathbb{R}^{L \times V}$ les vecteurs de coefficients des combinaisons linéaires d'atomes d'un dictionnaire $\boldsymbol{\Phi}$.
La première étape de l'algorithme qui consiste à résoudre le problème d'estimation parcimonieuse et déterminer la matrice $\boldsymbol{X}$ est réalisée par l'algorithme OMP décrit par l'algorithme 8. Pour la seconde étape, les atomes sont actualisés les uns après les autres comme décrit par l'algorithme 7.
L'algorithme K-SVD réalise l'apprentissage du dictionnaire en reformulant la décomposition comme suit

$$\| \mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{X} \|_2^2 = \| \mathbf{Y} - \sum_{i=1}^{L} \boldsymbol{\Phi}_{.j}\mathbf{X}_{j.} \|_F^2 \tag{A.1}$$

$$= \| (\mathbf{Y} - \sum_{i \neq l} \boldsymbol{\Phi}_{.j}\boldsymbol{X}j.) - \boldsymbol{\Phi}_{.l}\boldsymbol{X}_{l.} \|_F^2 \tag{A.2}$$

$$= \| \mathbf{E}_l - \boldsymbol{\Phi}_{.l}\boldsymbol{X}_{l.} \|_F^2 \tag{A.3}$$

Où $x_{j.}$ correspond à la $j^{\text{ème}}$ ligne de la matrice des coefficients $\boldsymbol{X}$ (la ligne contenant les coefficients relatifs au $l^{\text{ème}}$ atome) et $\boldsymbol{\Phi}_{.l}$ le $l^{\text{ème}}$ atome de $\boldsymbol{\Phi}$. On note $\mathbf{E}_l$ la matrice d'erreur

lorsque l'atome l à actualiser n'est pas pris en compte dans la représentation du signal, et $\mathbf{E}_l^R$ la matrice d'erreur correspondant à la matrice $\mathbf{E}_l$ dans laquelle on ne conserve que les signaux d'apprentissage dont la représentation fait intervenir l'atome l considéré. Les indices de ces derniers sont stockés dans un vecteur noté ω_l. Les mises à jour de $\mathbf{\Phi}_{.l}$ ainsi que de $\mathbf{X}_{l.}$ sont obtenues par une décomposition en valeur singulière de la matrice d'erreur réduite : $\mathbf{E}_l^R = \mathbf{U} \triangle \mathbf{V}^T$. L'actualisation de $\mathbf{\Phi}_{.l}$ est donnée par la première colonne de $\mathbf{U}$ notée $\mathbf{U}_{.1}$ et celle de $\mathbf{X}_{l.}$ est donnée par la première colonne de $\mathbf{V}$ multiplié par le premier élément de la diagonale de $\triangle$, notée $\triangle_{1,1}$, i.e., $\mathbf{V}_{.1}\triangle_{1,1}$.

Algorithm 7 $\mathbf{\Phi}$=K-SVD

Require: k=1, $\mathbf{\Phi}^0$
  repeat
    Actualisation de $\boldsymbol{x}$
    $\boldsymbol{x}^k \leftarrow \text{OMP}(\mathbf{Y}, \mathbf{\Phi}^{k-1})$
    Actualisation de $\mathbf{\Phi}$ et $\boldsymbol{X}$
    for l = 1 to L **do**
      $\omega_l^k \leftarrow \{i | 1 \leq i \leq V, \mathbf{X}_{j,i}^k \neq 0\}$
      $\mathbf{E}_{.l} \leftarrow \mathbf{Y} - \sum_{l \neq j} \mathbf{\Phi}_{.j} \mathbf{X}_{j.}^k$
      $\mathbf{E}_{.l}^R \leftarrow \{\mathbf{E}_{l.i} | i \in \omega_l^k\}$
      Décomposition SVD : $\mathbf{E}_l^R = \mathbf{U} \triangle \mathbf{V}^T$
      $\mathbf{\Phi}_{.l}^k \leftarrow \mathbf{U}_{.1}$
      $\mathbf{X}_{l.}^k \leftarrow \mathbf{V}_{.1} \triangle_{11}$
    end for
    $k \leftarrow k + 1$
  until critère d'arrêt

Algorithm 8 $\mathbf{X}$= OMP$(\mathbf{Y}, \mathbf{\Phi})$

Require: k=1, $\epsilon^0 = \mathbf{Y}, \mathbf{\Phi}^0 = \oslash$
  repeat
    for l = 1 to L **do**
      Corrélation : $\mathbf{C}_l^k \leftarrow \sum_{v=1}^{V} \langle \epsilon_{.v}^{k-1}, \mathbf{\Phi}_{.l} \rangle$
    end for
    Sélection : $l^k \leftarrow \arg\max_l |\mathbf{C}_{.l}^k|$
    Dictionnaire : $\mathbf{\Phi}^k \leftarrow [\mathbf{\Phi}^{k-1}, \mathbf{\Phi}_{.l}]$
    Coefficients : $\boldsymbol{x}^k \leftarrow \arg\min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{\Phi}^k \mathbf{X}\|_2^2$
    Résidu : $\epsilon^k \leftarrow \mathbf{Y} - \mathbf{\Phi}^k x^k$
    $k \leftarrow k + 1$
  until critère d'arrêt

A.2 L'algorithme ODL

L'algorithme ODL (Online Dictionary Learning) proposé par Mairal et al. (2009) est un algorithme d'apprentissage dit « en ligne », c'est-à-dire que, à chaque itération, le dictionnaire est actualisé en prenant en compte les informations des itérations précédentes. Par ailleurs l'algorithme ODL a été developpé pour permettre l'apprentissage de dictoinnaire à partir de gros volume de données.
A chaque itération, un élément de l'ensemble d'apprentissage est selectionné de manière aléatoire. Nous noterons noté $\mathbf{y} \in \mathbb{R}^N$ le signal sélectionné à l'itération k. Dans une première étape, le problème d'estimation parcimonieuse du signal selectionné sur le dictionnaire (A.4) est résolu à l'aide de l'algorithme LARS Efron et al. (2004); Osborne et al. (2000).
Le caractère « en ligne » de la méthode intervient dans la deuxième étape d'actualisation du dictionnaire qui consiste à mettre à jour le dictionnaire en résolvant le problème de minimisation suivant

$$\mathbf{\Phi}^k = \arg\min_{\mathbf{\Phi}} \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2} \parallel \mathbf{y}^k - \mathbf{\Phi}^{k-1} \boldsymbol{x}^k \parallel_2^2 + a \parallel \boldsymbol{x}^k \parallel_1 \tag{A.4}$$

Où k correspond à la $k^{\text{ème}}$ itération. On introduit les matrices $\mathbf{A}$ et $\mathbf{B}$

$$\mathbf{A}^k \leftarrow \mathbf{A}^{k-1} + \frac{1}{2} \boldsymbol{x}^k \boldsymbol{x}_T^k$$
$$\mathbf{B}^k \leftarrow \mathbf{B}^{k-1} + \mathbf{y}^k \boldsymbol{x}_T^k$$

Avec $\boldsymbol{x}_T^k$ la matrice transposée de $\boldsymbol{x}$ à l'itération k. Nous pouvons reformuler à partir de ces matrices le problème de minimisation (A.4) . L'algorithme ODL est décrit ci-dessous par l'algorithme 4.

$$\mathbf{\Phi}^k = \arg\min_{\mathbf{\Phi}} \frac{1}{k} \sum_{i=1}^{k} \frac{1}{2} \parallel \mathbf{y}^k - \mathbf{\Phi}^{k-1} \boldsymbol{x}^k \parallel_2^2 + a \parallel \boldsymbol{x}^k \parallel_1 \tag{A.5}$$
$$= \arg\min_{\mathbf{\Phi}} \frac{1}{k} (Tr(\mathbf{\Phi}_T^{k-1} \mathbf{\Phi}^{k-1} \mathbf{A}^k) - 2Tr(\mathbf{\Phi}_T \mathbf{B}^k)) \tag{A.6}$$

Algorithm 9 $\mathbf{\Phi}$=ODL($\mathbf{Y}, \mathbf{\Phi}$)

Require: $k = 1, \mathbf{Y}, \mathbf{\Phi}^0$
  for $k = 1$ to K **do**
    Actualisation de $\boldsymbol{x}$:
    $\boldsymbol{x} \leftarrow \text{LARS}(\mathbf{y}^k, \mathbf{\Phi}^{k-1})$
    Actualisation de $\mathbf{\Phi}$
    $\mathbf{A}^k \leftarrow \mathbf{A}^{k-1} + \frac{1}{2}\boldsymbol{x}^k \boldsymbol{x}_T^k$
    $\mathbf{B}^k \leftarrow \mathbf{B}^{k-1} + \mathbf{y}^k \boldsymbol{x}_T^k$
    for $l = 1$ to L **do**
      $\mathbf{u}_l \leftarrow \frac{1}{\mathbf{A}_{ll}}(\mathbf{B}_{.l} - \mathbf{\Phi}\mathbf{A}_{.l}) + \mathbf{D}_{.l}$
      $\mathbf{d}_{.l} \leftarrow \frac{1}{\max(\|\mathbf{u}_l\|_2, 1)}\mathbf{u}_l$
    end for
    $k \leftarrow k + 1$
  end for

Annexe B

Détails de calculs

B.1 Mise à jour de la variable auxilaire $\boldsymbol{z}$

Cette section détaille la mise à jour de la variable auxiliaire $\boldsymbol{z}$ introduite avec une contrainte d'égalité afin de reformuler le problème LASSO (voir section 2.2) (2.9) sous la forme

$$\min_{\boldsymbol{x},\boldsymbol{z}} \frac{1}{2} \parallel \boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x} \parallel_2^2 + \lambda \parallel \boldsymbol{z} \parallel_1 \text{ s.c } \boldsymbol{z} = \boldsymbol{x}. \tag{B.1}$$

La mise à jour de $\boldsymbol{z}$ à l'itération $k+1$ est donnée par

$$\begin{aligned}
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \left[\alpha \parallel \boldsymbol{z} \parallel_1 + \langle \boldsymbol{m}^k, \boldsymbol{z} - \boldsymbol{x}^{k+1} \rangle + \frac{\mu^k}{2} \parallel \boldsymbol{z} - \boldsymbol{x}^{k+1} \parallel_2^2 \right] \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \left[\frac{\alpha}{\mu} \parallel \boldsymbol{z} \parallel_1 + \frac{1}{\mu} \langle \boldsymbol{m}^k, \boldsymbol{z} - \boldsymbol{x}^{k+1} \rangle + \frac{1}{2} \parallel \boldsymbol{z} - \boldsymbol{x}^{k+1} \parallel_2^2 \right] \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \left[\frac{\alpha}{\mu} \parallel \boldsymbol{z} \parallel_1 + \frac{1}{\mu} \langle \boldsymbol{m}^k, \boldsymbol{z} - \boldsymbol{x}^{k+1} \rangle + \frac{1}{2} \parallel \boldsymbol{z} - \boldsymbol{x}^{k+1} \parallel_2^2 + \frac{1}{2\mu^2} \langle \boldsymbol{m}, \boldsymbol{m} \rangle \right] \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \left[\frac{\alpha}{\mu} \parallel \boldsymbol{z} \parallel_1 + \frac{1}{\mu} \langle \boldsymbol{m}^k, \boldsymbol{z} - \boldsymbol{x}^{k+1} \rangle + \frac{1}{2} \langle \boldsymbol{z} - \boldsymbol{x}^{k+1}, \boldsymbol{z} - \boldsymbol{x}^{k+1} \rangle + \frac{1}{2\mu^2} \langle \boldsymbol{m}, \boldsymbol{m} \rangle \right] \\
\boldsymbol{z}^{k+1} &= \arg\min_{\boldsymbol{z}} \left[\frac{\alpha}{\mu} \parallel \boldsymbol{z} \parallel_1 + \frac{1}{2} \parallel \boldsymbol{x}^{k+1} - \frac{1}{\mu} \boldsymbol{m}^k - \boldsymbol{z} \parallel_2^2 \right]
\end{aligned}$$

On pose $\mathbf{g} = \boldsymbol{x}^{k+1} - \frac{1}{\mu^k} \boldsymbol{m}^k$ et $\gamma = \frac{\alpha}{\mu^k}$ pour se ramener au problème suivant

$$\boldsymbol{z} = \arg\min_{\boldsymbol{z}} \left[\frac{1}{2} \parallel \mathbf{g} - \boldsymbol{z} \parallel_2^2 + \gamma \parallel \boldsymbol{z} \parallel_1 \right].$$

Bibliographie

Adler, A., Elad, M., Hel-Or, Y., Rivlin, E., 2015. Sparse Coding With Anomaly Detection. J. Signal Process. Syst. 79, 179–188.

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD : An Algorithm for Designing Overcomplete Dictionaries for Sparse Decomposition. IEEE Trans. Signal. Process. 54, 4311–4322.

Ahmed, T., Coates, M., Lakhina, A., 2007. Multivariate Online Anomaly Detection Using Kernel Recursive Least Squares, in : Proc. IEEE conf. Comput. Commun. (INFOCOM'2007), Anchorage, Alaska, USA. pp. 625–633.

Amsaleg, L., Chelly, O., Furon, T., Girard, S., Houle, M.E., Kawarabayashi, K.I., Nett, M., 2015. Estimating Local Intrinsic Dimensionality, in : Proc. Int. Conf. Knowledge Discovery and Data Mining (SIGKDD'15), Sydney, Australia. pp. 29–38.

Barreyre, C., 2018. Statistiques en grande dimension pour la détection d'anomalies dans les données fonctionnelles issues des satellites. PhD Thesis. Université de Toulouse. Toulouse, France. URL : `https://tel.archives-ouvertes.fr/tel-01689819`.

Barreyre, C., Laurent, B., Loubes, J.M., Boussouf, L., Cabon, B., 2019. Multiple Testing for Outlier Detection in Space Telemetries. IEEE Trans. Big Data (Early Access) .

Barreyre, C., Laurent, B., Loubes, J.M., Cabon, B., Boussouf, L., 2017. Detecting Abnormal Events in Multivariate Telemetries thanks to Covariance Analysis, in : Proc. Int. Conf. Big Data from Space (BiDS'17), Toulouse, France. pp. 118–121.

Barthélemy, Q., Gouy-Pailler, C., Isaac, Y., Souloumiac, A., Larue, A., Mars, J.I., 2013. Multivariate temporal dictionary learning for EEG. Journal of Neuroscience Methods 215, 19–28.

Beck, A., Teboulle, M., 2009a. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM Journal on Imaging Sciences 2, 183–202.

Beck, A., Teboulle, M., 2009b. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM J. Sci. Comput. 1, 183–202.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed Optimization and Statistical Learning via Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning 3, 1–222.

Breunig, M.M., Kriegel, H.P., t. Ng, R., Sander, J., 2000. LOF : Identifying Density-based Local Outliers, in : Proc. Int. Conf. Management of data (MOD'00), Dallas, TX, USA. pp. 93–104.

Candès, E., Wakin, M.B., Boyd, S., 2008. Enhancing Sparsity by Reweighted ℓ_1 Minimization. Journal of Fourier Analysis and Applications 14, 877–905.

Carrera, D., Boracchi, G., Foi, A., Wohlberg, B., 2015. Detecting Anomalous Structures by Convolutional Sparse Models, in : Proc. IEEE Int. Joint Conf. Neur. Net. (IJCNN'15), Killarney, Ireland.

Chalasani, R., Principe, J.C., Ramakrishnan, N., 2013. A fast proximal method for convolutional sparse coding, in : Proc. Int. Joint Conf. Neural Networks (IJCNN'13), Dallas, TX, USA. pp. 1–5.

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly Detection : A Survey. ACM Computing Survey 43, 1–72.

Chandola, V., Banerjee, A., Kumar, V., 2012. Anomaly Detection for Discrete Sequences : A Survey. ACM Computing Survey 24, 823–839.

Chen, S.S., Donoho, D.L., Saunders, M.A., 1996. Atomic Decomposition by Basis Pursuit. SIAM J. Sci. Comput. 20, 33–61.

Chen, S.S., Donoho, D.L., Saunders, M.A., 1998. Atomic Decomposition by Basis Pursuit. SIAM Journal on Scientific Computing 20, 33–61.

Cheng, H., Liu, Z., Yang, L., Chen, X., 2013. Sparse Representation and Learning in Visual Recognition : Theory and Applications. Signal Process. 93, 1408–1425.

Clauser, F., Peebles, G., Lagerstrom, P., Lipp, J., Krueger, R., Graham, E., Shevell, R., Sturdevant, V., Grimminger, G., Klemperer, W., Luskin, H., Baker, B., Bradshaw, E., Wheaton, E., Liepmann, H., 1946. Preliminary Design of an Experimental World-Circling Spaceship. RAND Corporation.

Dantzig, G., 1990. Origins of the Simplex Method. ACM , 141–151.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1997. Maximum Likelihood from Incomplete Data via the EM Algorithm. J. Roy. Statist. Soc. Series B (Methodological) 39, 1–38.

Dong, W., Li, B., 2010. Sparsity-Based Image Denoising via Dictionary Learning and Structural Clustering, in : Proc. IEEE conf. Comput. Vis. Pattern Recogni. (CVPR'10), San Siego, CA, USA. pp. 457–464.

Donoho, D.L., Johnstone, I.M., 1994. Adapting to Unknown Smoothness via Wavelet Shrinkage. J. Amer. Statist. Assoc. 90, 1200–1224.

Dubois, C., Avignon, M., Escudier, P., 2014. Observer la terre depuis l'espace - Enjeux des données spatiales pour la société. DUNOD.

Eckstein, J., Bertsekas, D., 1992. On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. Mathematical programming (Series A and B) 55, 293–318.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least Angle Regression. Ann. Statist. 32, 407–499.

Elad, M., Aharon, A., 2006. Image Denoising via Sparse and Redundant Representation over Learned Dictionaries. IEEE Trans. Image Process. 14, 3736–3745.

Eldar, Y., 2009. Generalized SURE for Exponential Families : Applications to Regularization. IEEE Trans. Signal Process. 57, 471–481.

Engan, K., Aase, S.O., Husoy, J.H., 1999. Method of Optimal Directions for Frame Design, in : IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP'99), pp. 2443–2446.

Engelen, J.E.V., Hoos, H.H., 2019. A survey on Semi-Supervised Learning. Machine Learning 109, 373–440.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in : Proc. Int. Conf. Knowledge Discovery and Data Mining (KDD'96), Portland, Oregon, USA. pp. 226–231.

Exarchakis, G., Lücke, J., 2017. Discrete SParse Coding. Neural Computation 29, 2979–3013.

Figueiredo, M.A.T., Nowak, R.D., 2003. An EM Algorithm for Wavelet-Based Image Restoration. IEEE Trans. Image Process. 12, 906–916.

Fu, W.J., 1998. Penalized Regressions : The Bridge versus the LASSO. Journal of Computational and Graphical Statistics 7, 397–416.

Fuertes, S., Picard, G., Tourneret, J.Y., Chaari, L., Ferrari, A., Richard, C., 2016. Improving Spacecraft Health Monitoring with Automatic Anomaly Detection Techniques, in : Proc. Int. Conf. Space Operations (SpaceOps'16), Daejeon, South Korea.

Gammerman, A., Vovk, V., 2006. Hedging Predictions in Machine Learning. The Computer Journal 50, 151–163.

Garcia-Cardona, C., Wohlberg, B., 2018. Convolutional Dictionary Learning : A Comparative Review and New Algorithms. IEEE Trans. Comput. Imag. 4, 366–381.

Giryes, R., Elad, M., Eldar, Y., . Appl. Comput. Harmon. Anal. , 407–422.

Grosse, R., Raina, R., Kwong, H., Ng, A.Y., 2007. Shift-Invariant Sparse Coding for Audio Classification, in : Proc. Int. Conf. Uncertainty in Artificial Intelligence, Vancouver, BC, Canada. pp. 149–158.

Heide, F., Heidrich, W., Wetzstein, G., 2010. Fast and Flexible Convolutional Sparse Coding, in : Proc. IEEE conf. Comput. Vis. Pattern Recogni. (CVPR'15), Boston, MA, USA. pp. 5135–5143.

Hennings, M., Puertas, G., Bornschein, J., Eggert, J., Lücke, J., 2010. Binary Sparse Coding, in : Proc. Int. Conf. Latent Variables Analysis and Signal Separation, St. Malo, France. pp. 450–457.

Houle, M.E., 2013. Dimensionality, Discriminability, Density and Distance Distributions, in : Proc. Int. Conf. Data Mining Workshops, Dallas, TX, USA. pp. 468–473.

Huang, K., Aviyente, S., 2006. Sparse Representation for signal classification, in : Proc. Neur. Inf. Process. Syst. (NIPS'06), Whistler, B C, Canada. pp. 609–616.

Hundman, K., Constantinou, V., Laporte, C., Colwell, I., Soderstrom, T., 2018. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding, in : Proc. Int. Conf. Knowledge Data Mining (KDD'18), London, United Kingdom. pp. 387–395.

Kavukcuoglu, K., Sermanet, P., Boureau, Y.L., Gregor, K., Michael, M., Cun, Y.L., 2010. Learning Convolutional Feature Hierarchies for Visual Recognition, in : Proc. Neur. Inf. Process. Syst. (NIPS'10), pp. 1090–1098.

Kriegel, H.P., Kröger, P., Schubert, E., Zimek, A., 2009. LoOP : Local Outlier Probabilities, in : Proc. Int. Conf. Information and Knowledge Management (CIKM'09), Hong Kong, China. pp. 1649–1652.

Laxhammar, R., Falkman, G., 2011. Sequential Conformal Anomaly Detection in Trajectories based on Hausdorff Distance, in : IEEE Int. Conf. Information Fusion (Fusion'2011), Chicago, Illinois, USA. pp. 1–8.

Laxhammar, R., Falkman, G., 2014. Online Learning and Sequential Anomaly Detection in Trajectories. IEEE Trans. Pattern Anal. Mach. intell. 36, 1158–1173.

Lehot, F., Clervoy, J., 2015. Histoire de la conquête spatiale. VUIBERT.

Lewiski, M.S., Sejnowski, T.J., 1999. Coding Time-Varying Signals Using Sparse,Shift-Invariant Representations, in : Proc. Neur. Inf. Process. Syst. (NIPS'99), Cambridge, MA, USA. pp. 730–736.

Luisier, F., Blu, T., Unser, M., 2011. Image Denoising in Mixed Poisson-Gaussian Noise. IEEE Trans. Image Process. 20, 696–708.

Luus, R., Jaakola, T.H.I., 1973. Optimization by Direct Search and Systematic Reduction of the Size of the Search Region. J. AIChE 19, 760–766.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2009. Online Dictionary learning for Sparse Coding, in : Proc. Int. Conf. Mach. Learn. (ICML'09), Montreal, QC, Canada. pp. 689–696.

Mallat, S.G., Zhang, Z., 1993. Matching Pursuits with Times-Frequency Dictionaries. IEEE Trans. Signal Process. 41, 3397–3415.

Markou, M., Singh, S., 2003a. Novelty Detection : A Review - part 1 : statistical approaches. Signal Process. 83, 2481–2497.

Markou, M., Singh, S., 2003b. Novelty Detection : A review - part 2 : Neural Network Based Approaches. Signal Process. 83, 2499–2521.

Martínez-Heras, J.A., Donati, A., Kirksch, M., Schmidt, F., 2012. New telemetry Monitoring Paradigm with Novelty Detection, in : Proc. Int. Conf. Space Operations (SpaceOps'12), Stockholm, Sweden.

Mei, X., Ling, H., 2011. Robust Visual Tracking and Vehicle Classification via Sparse Representation. IEEE Trans. Patt. Anal. Mach. intell. 33, 2259–2272.

Mohammad-Djafari, A., 1993. On the Estimation of Hyperparameters in Bayesian Approach of Solving Inverse Problems, in : IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP'93), Minneapolis, Minnesota USA. pp. 495–498.

Moreau, T., 2017. Convolutional Sparse Representations – Application to Physiological Signals and Interpretability for Deep Learning. PhD Thesis. Université Paris-Saclay. URL : `https://tel.archives-ouvertes.fr/tel-01689819`.

Nesterov, Y., 1983. A Method of Solving a Convex Programming Problem with Convergence Rate $O(1/k)^2$. Soviet Mathematics Doklady 27, 372–376.

Nesterov, Y., Nemirovski, A., 2013. On First-Order Algorithms for l1/nuclear Norm Minimization. Cambridge University Press 22, 509–575.

Oktar, Y., Turkan, M., 2018. A Review of Sparsity-Based Clustering Methods. Signal Process. 148, 20–30.

O'Meara, C., Schlag, L., Faltenbacher, L., Wickler, M., 2016. ATHMoS : Automated Telemetry Health Monitoring System at GSOC using Outlier Detection and Supervised Machine Learning, in : Proc. Int. Conf. Space Operations (SpaceOps'16), Daejeon, South Korea.

O'Meara, C., Schlag, L., Wickler, M., 2018. Applications of Deep Learning Neural Networks to Satellite Telemetry Monitoring, in : Proc. Int. Conf. Space Operations (SpaceOps'18), Marseille, France.

Osborne, M., Presnell, B., Turlach, B., 2000. A New Approach to Variable Selection in Least Squares Problems. IMA J. Num. Analysis 20, 389–403.

Pati, P.K.Y., Rezaiifar, R., 1993. Orthogonal Matching Pursuits : Recursive Function Approximation with Application to Wavelet decomposition, in : Proc. Asilomar Conf. Signals, Syst. Comput. (ACSSC'93), Pacific Grove, CA, USA. pp. 40–41.

Pendse, G.V., 2011. A tutorial on the LASSO and the shooting algorithm. Technical Report. Harvard Medical School.

Pesquet, J.C., Benazza-Benyahia, A., Chaux, C., 2009. Approach for Digital Signal/image Deconvolution Problems. IEEE Trans. Signal Process. 12, 4616–4632.

Pilastre, B., Boussouf, L., D'Escrivan, S., Tourneret, J.Y., 2019a. Multivariate Anomaly Detection in Mixed Telemetry Time-Series Using a Sparse Decomposition, in : Proc. of 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP'19), Le gosier, Guadeloupe. pp. 430–434.

Pilastre, B., Boussouf, L., D'Escrivan, S., Tourneret, J.Y., 2019b. Représentation Parcimonieuse pour la Détection d'Anomalies dans des Signaux Multivariés, in : Groupe d'Etude du Traitement du Signal et des Images (GRETSI'2019), Lille, France.

Pilastre, B., Boussouf, L., D'Escrivan, S., Tourneret, J.Y., 2019c. Spacecraft Health Monitoring using a Weighted Sparse Decomposition, in : World Congress on Condition Monitoring (WCCM'19), Singapore.

Pilastre, B., Boussouf, L., D'Escrivan, S., Tourneret, J.Y., 2020a. Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning. Signal Process. 168, 107320.

Pilastre, B., Silva, G., Boussouf, L., D'Escrivan, S., Rodriguez, P., Tourneret, J.Y., 2020b. Anomaly Detection in Mixed Time-Series using a Convolutional Sparse Representation with Application to Spacecraft Health Monitoring, in : Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP'19), Barcelona, Spain. pp. 3242–3246.

Pimentel, M., Clifton, D., Clifton, L., Tarassenko, L., 2014. A review of Novelty Detection. Signal Process. 99, 215–249.

Raginsky, M., Willet, R.M., Horn, C., Silva, J., Marcia, R.F., 2012. Sequential Anomaly Detection in the Presence of Noise and Limited Feedback. IEEE Trans. Inform. Theory 58, 5544–5562.

Rencker, L., Bach, F., Wang, W., M.D.Plumbley, 2018. Consistent Dictionary Learning for Signal Declipping, in : Proc. Int. Conf. Latent Variable Analysis and Signal Separation (LVA/ICA'2018), Guildford, UK.

Rencker, L., Bach, F., Wang, W., Plumbley, M., 2019. Sparse recovery and dictionary learning from nonlinear compressive measurements. IEEE Transactions on Signal Processing 67, 5659–5670.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C., 2001. Estimating the Support of a High-Dimensional Distribution. Neural Computation 3, 1443–1471.

Shafer, G., Vovk, V., 2008. A Tutorial on Conformal Prediction. The Journal of Machine Learning Research 9, 371–421.

Silva, G., Rodríguez, P., 2018. Efficient Algorithm for Convolutional Dictionary Learning via Accelerated Proximal Gradient Consensus, in : IEEE Int. Conf. Imag. Process. (ICIP'18), Athens, Greece. pp. 3978–3982.

Silva, G., Rodriguez, P., 2018. Efficient Convolutional Dictionary Learning Using Partial Update Fast Iterative Shrinkage-Thresholding Algorithm, in : IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP'18), Calgary, Canada. pp. 4674–4678.

Soubies, E., Blanc-Féraud, L., Aubert, G., 2015. A Continuous Exact ℓ_0 Penalty (CEL0) for Least Squares Regularized Problem. SIAM Journal on Imaging Sciences 8, 1607–1639.

Stein, C., 1981. Estimation of the Mean of a Multivariate Normal Distribution. Ann. Statis. 9, 1135–1511.

Takeishi, N., Yairi, T., 2014. Anomaly Detection from Multivariate Times-Series with Sparse Representation, in : Proc. IEEE Int. Conf. Syst. Man and Cybernetics, San Siego, CA, USA. pp. 2651–2656.

Tax, D., Duin, R., 2004. Support vector Data Description. Machine Learning 54, 45–66.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the LASSO. J. Roy. Statist. Soc. Series B (Methodological) 58, 267–288.

Tibshirani, R., Wasserman, L., 2015. Steins unbiased risk estimate, in : Course, notes from "Statistical Machine Learning, Spring 2015". Chapman and Hall/CRC, pp. 1–12.

Tipping, M., Bishop, C., 1999. Mixtures of Probabilistic Principal Component Analyzers. Neural Comput. 11, 443–482.

Tosic, I., Frossard, P., 2011. Dictionary Learning. IEEE Signal Process. Mag. 28, 27–38.

Dupre la Tour, T., 2018. Nonlinear Models for Neurophysiological Time Series. PhD Thesis. Université Paris-Saclay. URL : `https://pastel.archives-ouvertes.fr/tel-01990746`.

Verzola, I., Donati, A., Heras, J.A.M., Schubert, M., Somodi, L., 2016. Project Sybil : A Novelty Detection System for Human Spaceflight Operations, in : Proc. Int. Conf. Space Operations (SpaceOps'16), Daejeon, South Korea.

W. Dong, X. Li, L.Z., Shi, G., 2011. Sparsity-Based Image Denoising via Dictionary Learning and structural Clustering, in : Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR'11), pp. 457–464.

Wohlberg, B., 2014. Efficient Convolutional Sparse Coding, in : Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc. (ICASSP'14), pp. 7173–7177.

Wohlberg, B., 2016a. Efficient Algorithms for Convolutional Sparse Representations. IEEE Trans. Image Process. 25, 301–315.

Wohlberg, B., 2016b. Efficient Algorithms for Convolutional Sparse Representations. IEEE Trans. Imag. Process. 25, 301–315.

Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust Face Recognition via Sparse Representation. IEEE Trans. Patt. Anal. Mach. intell. 31, 1–18.

Xu, S., Yang, X., Jiang, S., 2017. A Fast Nonlocally Centralized Sparse Representation Algorithm for Image Denoising. Signal Process. 131, 99–112.

Yairi, T., Takeishi, N., Oda, T., Nakajima, Y., Nishimura, N., Takata, N., 2017. A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Pobabilistic Clustering and Dimensionality Reduction. IEEE Trans. Aerosp. Electron. Syst. 53, 1384–1401.

Yellin, F., Haeffele, B., Vidal, R., 2017. Blood Cell Detection and Counting in Holographic Lens-Free Imaging by Convolutional Sparse Dictionary Learning and Coding, in : Proc. IEEE Int. Symp. Biomed. Imag. (ISBI'17), Melbourne, Australia. pp. 650–653.

Yuan, M., Lin, Y., 2006. Model Selection and Estimation in Regression with Grouped Variables. J. Roy. Stat. Soc. Ser. B 68, 49–67.

Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R., 2010. Deconvolutional Networks, in : Proc. IEEE conf. Comput. Vis. Pattern Recogni. (CVPR'10), San Siego, CA, USA. pp. 2528–2535.

Zhang, Q., Li., B., 2011. Discriminative K-SVD for Dictionary Learning in Face Recognition, in : Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR'10), pp. 2691–2698.

Zhu, X., 2008. Semi-Supervised Learning Literature Survey. Technical Report. University of Wisconsin – Madison.