Statistiques en grande dimension pour la détection d'anomalies dans les données fonctionnelles issues des satellites

Clementine Barreyre

AIRBUS
DEFENCE & SPACE

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
l'Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Présentée et soutenue le *18/05/2018* **par :**

Clémentine Barreyre

Statistique en grande dimension pour la détection d'anomalies dans les données fonctionnelles issues des satellites

---

JURY

Bertrand Cabon	Ingénieur Airbus	Encadrant industriel
Béatrice Laurent	Professeur d'Université	Directrice de Thèse
Jean-Michel Loubes	Professeur d'Université	Co-directeur de Thèse
Mathilde Mougeot	Professeur d'Université	Rapporteure
Adeline Samson	Professeur d'Université	Examinatrice
Jérome Saracco	Professeur d'Université	Rapporteur

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur(s) de Thèse :
Béatrice LAURENT, *Jean-Michel LOUBES* et *Bertrand CABON*
Rapporteurs :
Mathilde MOUGEOT et *Jérome SARACCO*

À ma famille

Toujours du plus grand des soutiens

Remerciements

Je souhaite remercier en premier lieu ma directrice de thèse Béatrice, qui a été toujours d'une grande disponibilité pour m'aider dans mes travaux de recherche. Cela a été très agréable de travailler avec toi, notamment quand on a eu l'occasion de se creuser la tête ensemble. C'est au final pour cela que j'ai commencé à aimer les mathématiques, et tu as su faire perdurer cette passion, de mon cursus à l'INSA jusqu'à maintenant. Merci beaucoup pour cela. Tu es de plus une personne avec de grandes qualités humaines, toujours de bon conseil. Je te dois beaucoup concernant l'aboutissement de cette thèse.

Je souhaite également remercier Jean-Michel, mon co-directeur de thèse, pour son engagement, pour toutes ces idées que tu m'as soufflées tout au long de la thèse. Tu m'as également laissé une grande liberté pour aller au bout de mes idées, et su me rediriger vers le droit chemin quand je commençais à un peu trop divaguer. Merci pour tes précieux conseils, pour avoir partagé ta vaste connaissance des mathématiques, et pour ta bonne humeur.

Il serait impossible de ne pas remercier Bertrand, sans qui cette thèse n'aurait peut-être jamais vu le jour. Tu m'as d'abord fait confiance en tant que stagiaire, et tu n'as jamais cessé de croire en moi, au point d'aller défendre le montage de cette thèse bec et ongles, avec l'acharnement qui te ressemble. Je ne pensais pas qu'il était possible de travailler autant dans la bonne humeur avant de t'avoir comme chef. Pour ces très bons moments de discussions complètement loufoques, je te remercie !

Je souhaite aussi évidemment remercier Loïc, mon binôme et collègue de bureau depuis 3 ans. Je sais que tu as passé un temps fou à me venir en aide, à y passer même des heures, sans que tu aies à le faire, et toujours de bon coeur. Je te serai également toujours reconnaissante pour avoir vanté les mérites de mes algorithmes auprès de (presque) toute la boîte. Je sais que si je suis embauchée aujourd'hui, c'est parce que tu as pu tâter le terrain pour moi. Un énorme merci pour cela. Enfin, c'était très agréable de constater qu'on avait le même sens de l'humour de qualité (parfois) approximative. C'est un réel plaisir de continuer à travailler avec toi.

Je remercie également mes clients et sponsors, qui m'ont permis de réaliser cette thèse: Guillaume et Alexandre côté validation des satellites télécoms, Celestino, Antoine, Grégory et Sophie pour le suivi des satellites en vol. Merci pour le temps que vous m'avez accordé, pour toutes vos idées et pour m'avoir fait confiance au cours de ces trois années.

Je tiens également à remercier mes chefs qui m'ont supporté tout au long de la thèse et se sont battus pour mon embauche. Merci à Arnaud, Jean-Michel, Hervé.

Le travail reste le travail, mais s'il s'effectue dans la bonne humeur, c'est toujours plus agréable. Et pour cela j'ai eu une chance inouïe. Je peux remercier le pôle pose dans son intégralité : Mathilde, Lalla, Théo, Eliott, Hugo, Fred, Fawzi, Gwenaël, Sylvain, mais aussi tous les collègues croisés en salle de pause ou autour d'un repas, qui ont pu rendre ces années si agréables. Pour m'avoir soutenu en dehors des heures de travail, je remercie également tout mon entourage, ma famille pour son soutien sans faille, mon groupe de musique, mes amis pour tous ces bons moments et pour m'avoir tant encouragé tout au long de ces trois années.

Je souhaite enfin remercier Mathilde Mougeot et Jérome Saracco d'avoir accepté de rapporter ma thèse.

Table des matières

Table des figures

Liste des tableaux

Résumé

Ce travail de thèse consiste au développement de méthodes statistiques pour détecter des comportements anormaux dans les données fonctionnelles que produit le satellite tout au long de sa vie. Les satellites sont des systèmes complexes qui, une fois en vol, ne peuvent être maintenus que par l'utilisation d'équipements redondants et l'utilisation de certains modes de fonctionnements. L'analyse des données produites par le satellite est par conséquent l'unique moyen de se rendre compte de son état de santé afin d'anticiper les pannes, et de les détecter dès qu'elles surviennent. De plus, les données produites par le satellite sont très nombreuses et ne peuvent être revues au cas par cas par les ingénieurs. En effet, un satellite peut enregistrer plus de 10 000 paramètres en continu, tout au long de sa vie, le plus souvent à raison d'une mesure toutes les trente secondes.
Les données sont essentiellement des mesures de paramètres au cours du temps. Par conséquent, ce sont en grande partie des données fonctionnelles. Dans cette thèse, nous abordons des données provenant de deux sources différentes : les données de télémesures et de tests.
Les télémesures sont des indicateurs de bonne santé que produit le satellite continuellement, de son lancement jusqu'à sa désorbitation. Ce sont majoritairement des mesures physiques telles que des valeurs de températures, angles, intensités de courant. Les données de notre étude sont par conséquent des séries temporelles échantillonnées régulièrement tout au long de la vie du satellite. Les données de tests sont des mesures de performances réalisées sur le satellite avant qu'il ne soit envoyé dans l'Espace, et sont ainsi majoritairement des données fonctionnelles.
Pour ces deux sources de données, les méthodes que nous développons sont destinées à identifier les individus les plus aberrants parmi un jeu de courbes, ce dernier pouvant contenir les résultats d'une phase de test ou une archive de télémesures, par exemple.
Au cours de ce travail de thèse, nous avons développé des méthodes statistiques univariées et multivariées pour détecter des comportements atypiques dans un jeu de courbes. Ce sont par conséquent des données de grande dimension. De plus, une grande partie de l'information acquise est fortement corrélée, de par la continuité des courbes considérées. Pour cette raison, nous avons développé des indicateurs définis grâce à des projections sur des bases de fonctions afin de réduire la dimension des données. Nous réalisons la sélection des coefficients en privilégiant les indicateurs contenant de l'information pertinente sur

les anomalies.

Un premier travail au cours de cette thèse a été d'identifier les différentes anomalies que l'on pouvait observer sur les données et de comprendre comment les mettre en évidence grâce à des projections sur des bases de fonctions. Nous avons testé un grand nombre de bases de fonctions différentes : Fourier, ondelettes, bases à noyaux, analyse en composantes principales, analyse en composantes principales à noyaux. Nous avons également étudié les données recalées dans le temps, ainsi que le périodogramme des fonctions. Visuellement, nous avons remarqué que les anomalies locales étaient plus difficiles à mettre en lumière en retenant uniquement des premiers coefficients issus d'une projection. Les hauts niveaux d'ondelettes et les coefficients issus de la projection du périodogramme sont les plus efficaces pour mettre en lumière les disparités locales dès les premiers coefficients. En complément de cette revue des projections, nous avons également appliqué plusieurs méthodes de détection d'anomalies, telles que la One-Class SVM et le Local Outlier Factor (LOF). En plus de ces deux méthodes, nous avons développé notre propre méthode pour prendre en compte la saisonnalité des courbes que nous considérons dans le cas des télémesures, que nous avons appelée TNND (Temporal Nearest Neighbors Dissimilarity). Nous avons pu identifier que la One-Class SVM n'était pas adaptée aux télémesures spatiales, et que le Local Outlier Factor ainsi que notre nouvelle méthode renvoyaient de très bons résultats, que ce soient sur des données simulées comme sur des données réelles.

En se basant sur cette étude, nous avons développé une nouvelle procédure pour sélectionner automatiquement les coefficients les plus intéressants pour la détection d'anomalies dans un cadre semi-supervisé. Notre méthode est une procédure de tests multiples où nous appliquons, à tous les niveaux de coefficients, un test à deux échantillons réalisé à partir du calcul de la distance de Wasserstein. Ici, nous comparons les coefficients de deux sous-ensembles, dont l'un d'entre eux est connu comme ne comportant pas d'anomalies. Nous contrôlons le taux de faux positifs grâce à la procédure de Benjamini-Hochberg. Cette nouvelle méthode nous renvoie donc, à partir d'une projection donnée, un ensemble de coefficients jugés intéressants pour la détection de données aberrantes, sur lesquels nous pouvons appliquer une méthode de détection d'anomalies au choix. Nous avons choisi pour illustrer les résultats de cette procédure de calculer les valeurs du LOF, étant donné ses résultats satisfaisants obtenus dans l'étude précédente.
Nous testons cette procédure sur les coefficients issus de l'ACP et de la décomposition en ondelettes de Haar. Les résultats nous montrent que cette procédure améliore très largement la détection d'anomalies dans les télémesures, et permet également de diminuer le taux de fausses détections. De plus, nous avons pu montrer par un exemple que cette procédure restait efficace même si des anomalies subsistaient dans l'ensemble supposé nominal. Ce dernier résultat nous permet d'appliquer la procédure dans un environnement non labélisé.

Enfin, nous nous sommes intéressés à l'aspect multivarié des données. En effet, en cas de panne avérée sur le satellite, il est fort probable que plus d'une télémesure soit impactée par un changement majeur. Pour cette raison, nous nous sommes intéressés aux covariances des télémesures entre elles. Pour cela, nous cherchons à comparer les covariances entre un groupe de télémesures pour deux journées, ou périodes consécutives.
Nous avons appliqué trois tests statistiques ayant des angles d'approche différents. Le premier test a été développé par Fremdt, et se base sur la différence entre les projections des deux matrices de covariances sur la base des vecteurs propres d'une matrice de référence. Le deuxième test, introduit par Ilea, est construit à partir des valeurs propres du produit matriciel entre l'inverse de la matrice de covariance de la deuxième journée, et de la matrice de covariance de la première journée. Enfin, le troisième test, développé par Cai, est un test robuste dans des conditions sparse, et repose sur la plus grande différence entre les deux matrices. Nous avons également développé un nouveau test asymptotique, inspiré des deux premiers tests. Outre la démonstration de la convergence de notre test, nous démontrons par des exemples que ce test est dans la pratique le plus puissant sur les données dont nous disposons. Nous abordons également la problématique de réduction de dimension dans le calcul de covariance.
Pour chacune des méthodes, nous avons pu détecter les anomalies les plus significatives, améliorant sensiblement le taux de fausses alarmes. Ces méthodes sont en train d'être intégrées aux outils internes à Airbus pour faciliter le suivi des satellites.

Chapitre 1

Introduction

1.1 Qu'est-ce qu'un satellite ?

Avant de vous présenter les axes de recherche de cette thèse, il est important de présenter en premier lieu le contexte de l'industrie spatiale afin de mieux cerner les problématiques autour desquelles s'inscrit cette thèse, et la provenance des données que nous traitons.

1.1.1 Description générale

1.1.1.1 Historique et applications

Selon une définition générale, un satellite est un corps en orbite circulant autour d'un objet de plus grande taille, en particulier une étoile ou une planète. Les satellites peuvent être naturels, à l'exemple de la Terre qui est un satellite du Soleil, ou la Lune qui est un satellite de la Terre. Mais les satellites que nous aborderons dans cette étude sont des satellites artificiels.
Le premier satellite artificiel, Sputnik, a été lancé en 1957 par l'URSS, en plein contexte de conquête spatiale, voir [51] pour plus d'éléments de contexte historique. Les premiers satellites avaient pour but de démontrer le progrès scientifique grâce à l'accomplissement de missions symboliques. Au cours de cette période, les scientifiques ont réalisé de grandes découvertes sur l'Espace et son environnement, et ont notamment réalisé les premiers vols habités dans l'Espace.
Au-delà de ces missions, une industrie spatiale s'est bâtie. Aujourd'hui, les satellites artificiels ont un réel poids économique dans de nombreux domaines : télécommunications, positionnement (GPS), météorologie, défense... Certains satellites sont encore purement dédiés à l'exploration scientifique, dont les missions sont toujours plus ambitieuses et à la pointe de la technologie. A titre d'exemple, le programme Juice initié en 2012 par l'ESA, l'agence spatiale européenne, a pour but d'explorer les lunes glacées de Jupiter.
Comme nous venons de le constater, il existe un grand nombre d'applications possibles

pour les satellites, par conséquent beaucoup de satellites orbitent actuellement au-dessus de nos têtes. Selon l'association UCS (Union of Concerned Scientists), il y avait mi-2016 environ 1400 satellites en opération autour de la Terre.
Dans les prochaines sections, nous nous intéressons aux propriétés des satellites qui vont impacter les données produites tout au long de la vie du satellite, que nous prendrons en compte dans notre étude.

1.1.1.2 Orbites

En fonction de leur mission, les satellites peuvent avoir des propriétés très variées, impactant directement les données qu'ils produisent, et qui entrent dans notre périmètre d'étude. Parmi les paramètres ayant une forte influence, le choix de l'orbite impacte directement l'environnement dans lequel se place le satellite. En effet, la vitesse d'un satellite et sa période de révolution sont directement liés à l'orbite sur laquelle il circule. Une orbite se caractérise par son altitude, son inclinaison et son excentricité. En voici quelques exemples, parmi les orbites les plus connues.

- Les orbites basses (souvent appelées LEO pour Low-Earth-Orbit) sont des orbites situées entre 180 et 2000km de la Terre. Les satellites chargés d'observer la Terre sont donc prioritairement envoyés sur une orbite basse. En effet, être plus proche de la Terre permet logiquement d'obtenir une meilleure résolution optique. Parmi les satellites en orbite basse, on trouve les satellites météorologiques, certains satellites de télécommunications, ainsi que la station spatiale internationale (ISS).

- Les orbites géostationnaires (souvent appelées GEO) sont des orbites caractérisées par une inclinaison orbitale nulle et une vitesse orbitale identique à celle du corps autour duquel il gravite. Cela signifie qu'un satellite géostationnaire orbite dans le plan de l'équateur, et reste situé au-dessus de la même position sur Terre. Ces satellites se situent à une distance de 35 786km de la Terre. De nombreux satellites de télécommunications ainsi que certains satellites météorologiques circulent sur cette orbite. En totalité, on dénombre plus de 250 satellites actuellement placés sur l'orbite géostationnaire. Cette orbite est souvent choisie pour ses propriétés de stabilité géographique au cours du temps.

D'autres orbites peuvent être choisies, comme les orbites héliosynchrones, cas particulier des orbites basses où le satellite passe toujours au-dessus d'une position donnée à la même heure locale, ou encore les points de Lagrange pour des besoins plus spécifiques d'exploration scientifique. Voir [12] pour plus d'informations.
A titre d'exemple, on peut remarquer que l'orbite influence les données produites par le satellite puisque l'ensoleillement du satellite est directement lié à sa période de révolution. L'ensoleillement joue un rôle considérable sur l'environnement du satellite, notamment sur

les températures à bord et la production d'énergie par les panneaux solaires. En conséquence, beaucoup de paramètres à bord du satellite en dépendent.

1.1.1.3 Charge utile et plateforme

Chaque satellite, quelle que soit sa mission, est composé d'une charge utile et d'une plateforme. Dans le cadre de la thèse, il est important de noter que les données relatives à la plateforme et à la charge utile sont, comme pour le choix de l'orbite, sensiblement différentes.

- La charge utile est définie spécifiquement pour la mission du satellite. Elle comporte les antennes émettrices et réceptrices, ainsi que tous les équipements électroniques qui vont servir à traiter le signal entre les deux antennes pour mener à bien la mission. Pour les satellites d'observation, elle comporte également les senseurs optiques ou radars nécessaires à l'observation de la Terre. Ces équipements sont reliés les uns avec les autres grâce à des guides d'ondes et des câbles coaxiaux.

- La plateforme a plusieurs fonctions majeures, annexes aux besoins spécifiques de la mission mais indispensables. Tout d'abord, elle permet de fournir de l'énergie au satellite, grâce notamment aux panneaux solaires et aux batteries que l'on trouve à bord. Ensuite, elle permet le maintien de la position et de l'attitude du satellite, grâce à des propulseurs. Ceux-ci sont également indispensables à la mise à poste du satellite après qu'il se soit désolidarisé du lanceur. Ensuite, elle comporte également le contrôle thermique, la gestion du bord qui pilote le satellite. Les plateformes sont souvent redondantes d'un satellite à un autre, ce qui permet aux constructeurs de proposer des plateformes génériques. Les données provenant des plateformes sont donc directement comparables d'un satellite à un autre, contrairement aux données issues de la charge utile. Dans le contexte de la thèse liée à l'analyse de données satellites, cette similarité entre les plateformes permet de comparer des données provenant de plusieurs satellites, afin par exemple d'anticiper le comportement de ces plateformes relativement à ce qui a été observé sur les précédentes.

Le schéma 1.1 nous montre une représentation éclatée d'un satellite de télécommunications, où chaque sous-système est classé en fonction de son appartenance à la charge utile ou la plateforme.
Pour cette étude, on considèrera deux familles de satellites : les satellites de télécommunications et d'observation. Ces satellites ont des caractéristiques différentes qui seront détaillées dans le prochain paragraphe. Les données issues de ces deux types de satellites sont une fois de plus assez différentes, et bénéficieront de traitements différents lors du travail de thèse.

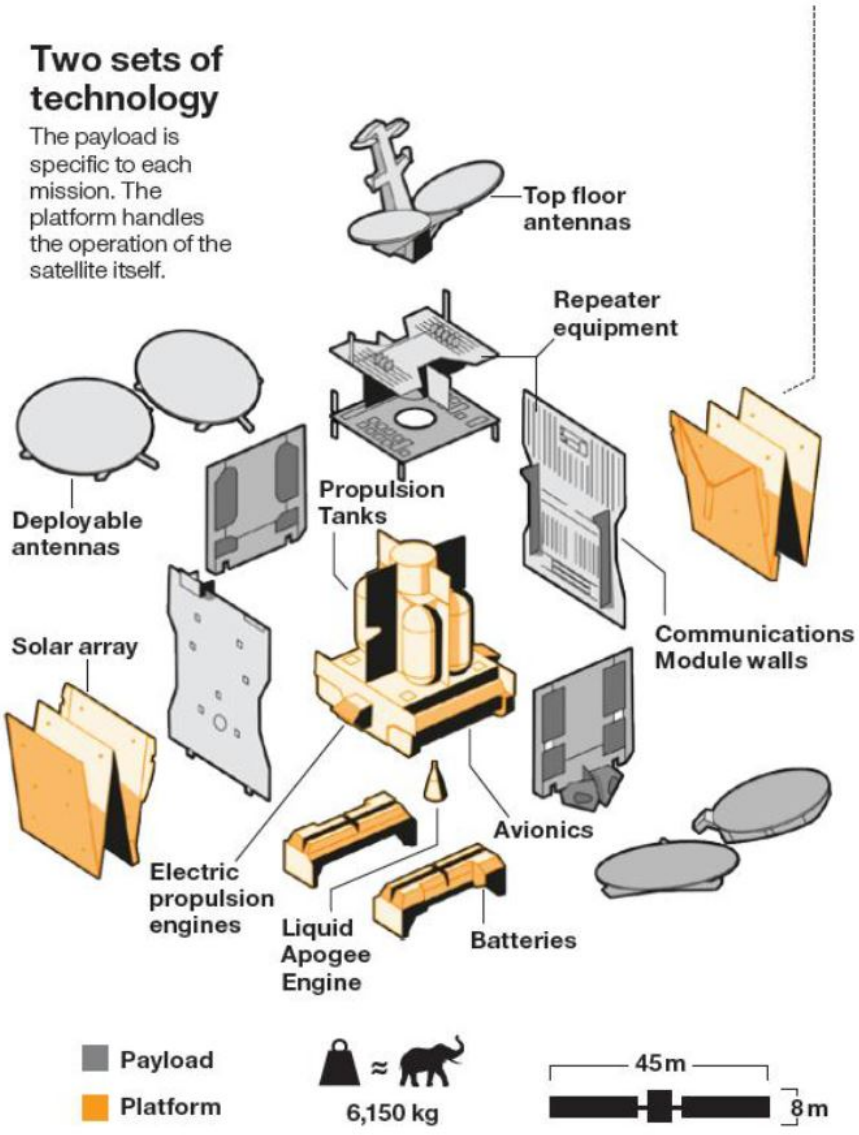

FIGURE 1.1 – Charge utile (payload) et plateforme (platform) dans le satellite.

1.1.2 Satellite de télécommunications

Maral et al. [55] ont fourni une description très détaillée des satellites de télécommunications et de leur fonctionnement. Ces satellites peuvent être utilisés pour assurer des liaisons téléphoniques, la diffusion de programmes télévisuels, internet et le positionnement GPS.
Le premier satellite de télécommunications, INTELSAT I, a été lancé en 1965. Comme expliqué précédemment, la plupart des satellites de télécommunications sont géostationnaires. Les fonctions assurées par la plateforme étant identiques d'un satellite de télécommunications à un autre, les constructeurs proposent des plateformes génériques à tous les satellites de télécommunications géostationnaires. La plateforme de référence pour les satellites de télécommunication fabriqués par Airbus Defence and Space depuis 2004 s'appelle Eurostar 3000. Une cinquantaine de satellites munis de cette plateforme sont aujourd'hui en vol, tel que le satellite SES-10, représenté en figure 1.2. Voir [64] pour plus de détails sur cette plateforme.
La fonction de la charge utile dans le cas d'un satellite de télécommunications consiste en la réception d'un signal de très faible amplitude provenant de la Terre, ou éventuellement d'un autre satellite, son amplification, et son renvoi vers la Terre. Le signal passe par des équipements comme les multiplexeurs d'entrée et de sortie, souvent désignés IMUX, OMUX. Parmi les équipements constituant la charge utile, on trouve également des amplificateurs (ou LNA), des filtres et oscillateurs. Cette variété d'équipements dans la charge utile a pour conséquence de produire une grande variété de données, qui feront partie des données étudiées dans la thèse.
Le client, futur propriétaire du satellite, spécifie certaines fonctionnalités qui vont affecter directement l'architecture de la charge utile, comme par exemple les zones sur Terre à cibler, le choix d'antennes multifaisceaux, le nombre d'antennes, les bandes passantes requises, l'utilisation d'équipements redondants dans le cas d'éventuelles pannes des équipements principaux. En conséquence, la charge utile change considérablement d'un satellite à un autre, elle est donc pensée sur mesure pour chaque nouveau satellite : cette phase porte le nom particulier de la personnalisation, fréquemment désignée comme la customisation.

1.1.3 Satellite d'observation

Les satellites d'observation peuvent répondre à de nombreux besoins : météorologie, reconnaissance militaire, modélisation du climat, géodésie, suivi de parcelles d'agriculture, recherche scientifique. De nombreux satellites d'observation sont connus du grand public, à l'image des satellites Pléiades et Spot, dont fait partie le satellite Spot-6, représenté Figure 1.3. Ces satellites sont destinés à l'observation de la Terre. On peut également citer quelques satellites d'exploration scientifique, tel que le télescope Hubble, Gaia et Rosetta.

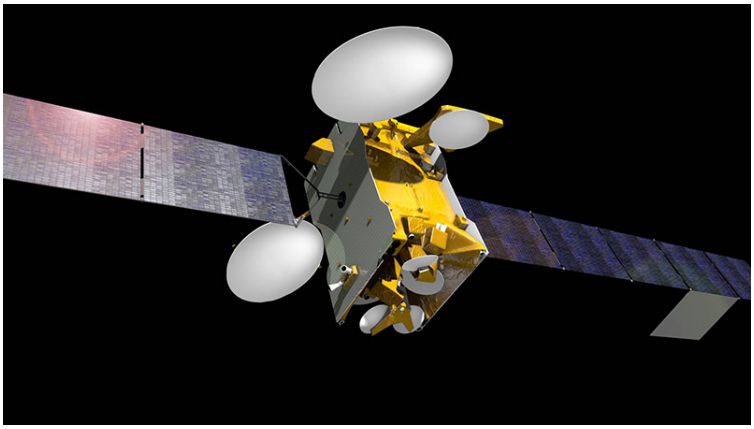

FIGURE 1.2 – SES-10, un satellite conçu par Airbus.

FIGURE 1.3 – Spot 6, un satellite conçu par Airbus.

Les applications étant très variées, les satellites sont très différents les uns des autres. La charge utile, comme pour les satellites de télécommunications, est pensée sur mesure pour chaque satellite, et les plateformes peuvent aussi varier considérablement. Il existe néanmoins quelques plateformes redondantes d'un satellite à un autre, telle que la plateforme Astrobus proposée par Airbus, intégrée notamment aux satellites Spot [22].
La plupart de ces satellites comportent des instruments de télédétection, ou instruments optiques, qui peuvent être des senseurs optiques, des radars, des spectromètres. Ils participent à l'analyse d'ondes électromagnétiques : lumière visible, lumière infrarouge, rayons X. La plupart de ces satellites circulent en orbite basse, et ont une période de révolution autour la Terre de l'ordre de 100 minutes. Cette variation de périodicité et l'absence de généricité des satellites d'observation est également à prendre en compte dans la thèse.

1.1.4 Tendances et innovations actuelles

Il est important de mettre en évidence les tendances et innovations majeures qui entrent en jeu actuellement dans l'industrie spatiale. Ces innovations impactent considérablement la manière dont les satellites sont conçus. Ces évolutions du contexte ont joué en faveur de l'élaboration du cadre d'étude de cette thèse, pour anticiper les challenges à venir autour de l'analyse des données spatiales.

1.1.4.1 De la propulsion chimique à la propulsion électrique

Encore aujourd'hui, la plupart des satellites en orbite utilisent la propulsion chimique pour atteindre leur position finale. Pour les satellites plus récents, actuellement en conception ou en attente de lancement, la mise en orbite est planifiée pour se faire grâce à la propulsion électrique. Cette nouvelle manière de procéder a des répercussions sur les conditions de mise à poste du satellite.
Il est important de comprendre en premier lieu les nombreux avantages apportés par la propulsion électrique. Elle est tout d'abord moins coûteuse car elle permet de réduire le poids du satellite et la taille de ses réservoirs. L'utilisation de la propulsion est en effet limitée par la quantité d'ergol contenue dans les réservoirs initialement. Dans le cas de la propulsion électrique, le Xénon est ejecté en l'ionisant et non plus en le brulant, nécessitant une plus faible quantité d'ergol que la propulsion chimique. Cette dernière propriété va permettre d'allonger considérablement la durée de vie des satellites. En contrepartie, la mise à poste du satellite avec la propulsion électrique peut prendre plusieurs mois, contre quelques jours pour la propulsion chimique. Pour avoir de plus amples informations, l'article [11] détaille le cas du satellite Eutelsat 172B, premier satellite à propulsion électrique construit par Airbus. En complément, l'encadré 1.4 issu du même article nous offre une visualisation des différences entre les deux types de propulsions.
La mise en orbite du satellite est une étape cruciale, et le suivi de cette phase devra donc s'adapter à l'allongement de sa durée. Il est plus difficile de mobiliser les ingénieurs à temps plein pendant plusieurs mois que pendant quelques jours pour assurer le suivi de cette phase. En conséquence, l'automatisation du processus de surveillance de la mise en orbite permet d'apporter des solutions et d'alléger la charge de travail de ces équipes. Le travail réalisé dans cette thèse permet de pouvoir répondre partiellement à cette problématique.

1.1.4.2 Naissance de constellations de satellites

Une autre nouveauté qui stimule l'industrie spatiale est l'émergence de constellations de satellites. Une constellation est un ensemble de satellites ayant une mission commune. Afin de se répartir la couverture géographique équitablement, les satellites se situent en orbite

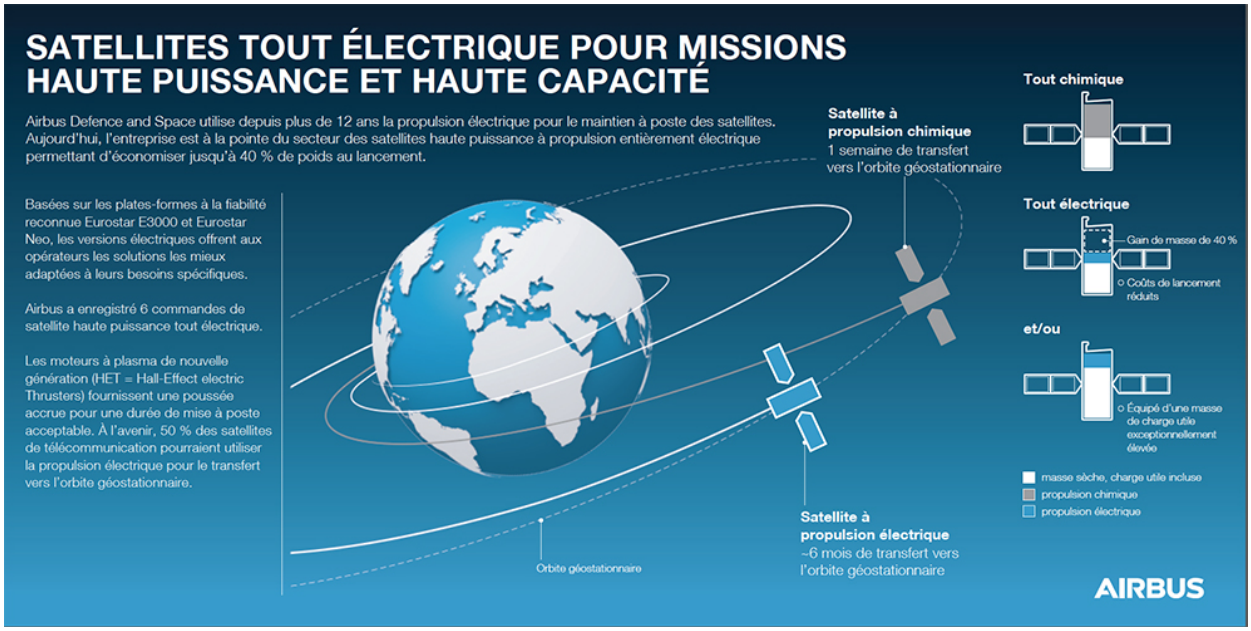

FIGURE 1.4 – Eutelsat 172B et les avantages de la propulsion chimique.

basse et leurs mouvements sont synchronisés. Pour les satellites de télécommunications, les constellations permettent de réduire les retards dans les conversations téléphoniques. En effet, la plupart des satellites de télécommunications sont géostationnaires, donc très éloignés de la Terre. En orbite basse, les délais sont allégés par la plus grande proximité avec la Terre tout en assurant une stabilité de la couverture, grâce à la synchronisation des satellites de la constellation.
Le positionnement est également assuré par des constellations de satellites, comme par exemple les satellites Galileo. Les constellations de satellites sont nécessaires pour pouvoir être précisément localisé à n'importe quel endroit de la planète.
Enfin, les constellations de satellites d'observation permettent déjà de pouvoir produire de l'imagerie satellite à plusieurs endroits de manière simultanée, à l'image des constellations Pléiades et Spot. Une nouvelle constellation construite par Airbus [22] va permettre d'améliorer encore une fois la réactivité et la résolution de l'imagerie produite par les satellites, afin d'obtenir rapidement des images de grande qualité lors de situations urgentes telles que les catastrophes naturelles.
Enfin, les constellations de satellites permettront de diffuser le réseau internet sur toute la surface du globe, grâce notamment au projet de OneWeb satellites, qui va envoyer pas moins de 900 satellites autour du globe (voir [7], ainsi que l'encadré 1.5).
La naissance de ces constellations de satellites est un vrai défi pour le futur, et impacte directement la manière de concevoir les satellites et de les opérer. Par exemple, la production des satellites doit être repensée, car ces satellites seront désormais dans un mode de

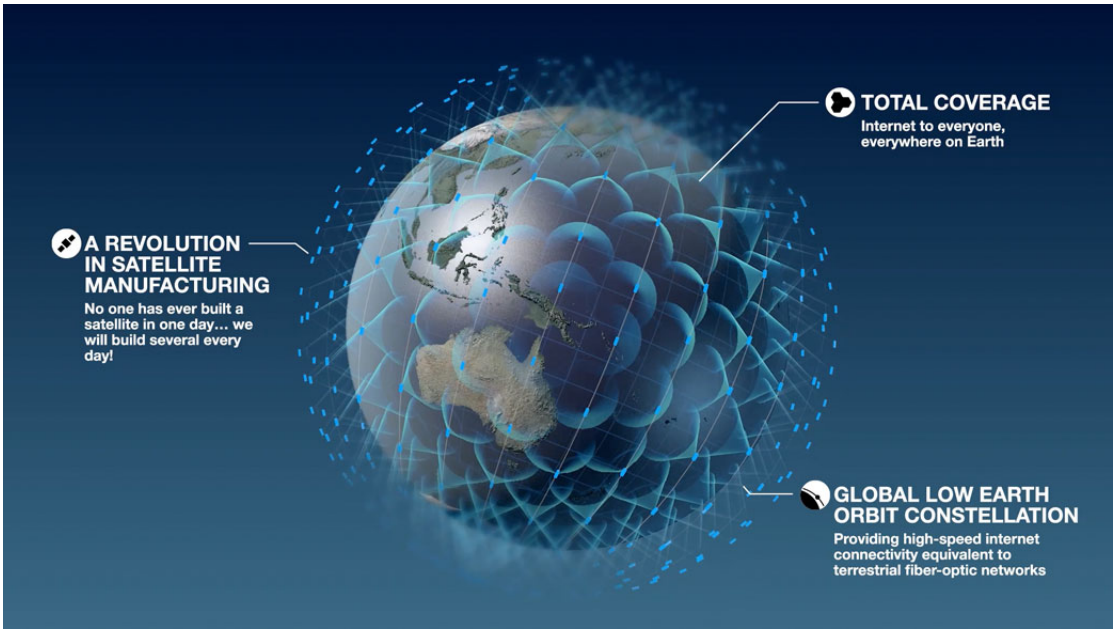

FIGURE 1.5 – La constellation OneWeb, composée de 900 satellites de télécommunications.

production en séries, et non plus sur mesure. En conséquence, les tests doivent également être repensés, ainsi que toutes les autres étapes de la vie du satellite. Ces étapes telles qu'elles existent aujourd'hui seront abordées dans la prochaine section.
La quantité de données à traiter augmente fortement dans le cadre de constellations, et le croisement des données entre les satellites, difficile jusqu'alors, apporte ici une forte valeur ajoutée. Cette propriété a également motivé la création de la thèse.

1.2 Cycle de vie du satellite

Le cycle de vie du satellite peut se résumer en quatre phases principales. Tout d'abord, le satellite est imaginé lors de la phase de conception, où de nombreuses études sont réalisées pour démontrer la faisabilité du projet. Le satellite voit ensuite le jour lors de la phase d'assemblage. Par la suite, le satellite est validé et qualifié lors des tests. Enfin, les opérations se déroulent une fois que le satellite est lancé dans l'Espace et qu'il a atteint son orbite finale. Les phases de la vie du satellite sont représentées Figure 1.6. Il est important dans le cadre de la thèse de détailler ces étapes et de saisir la provenance des données au travers des processus qui les ont engendrées pour mieux comprendre leur utilité et la valeur que nous pouvons leur apporter.

FIGURE 1.6 – Le cycle de vie du satellite.

1.2.1 Conception du satellite

La première étape du cycle de vie du satellite est une phase exploratoire, où l'on réalise dans les grandes lignes les études pour définir ce que sera le futur satellite. L'article de Hodges [42] nous offre une description détaillée de chacune de ces étapes.
Chaque projet commence par un appel d'offres émis par un client pour le besoin d'un nouveau satellite. Le client que nous évoquons peut être par exemple une agence gouvernementale ou un opérateur de télécommunications. Dans l'appel d'offres, il spécifie les particularités qu'il requiert pour son futur satellite. Le but de cette étape pour un constructeur comme Airbus est d'étudier la faisabilité de la mission spécifiée par le client. Ces études ont pour objectif d'estimer le prix du futur satellite, et de proposer un planning pour anticiper les délais nécessaires à la fabrication du satellite. Une réponse à l'appel d'offres est donc proposée au client, qui est alors libre de choisir entre les différents constructeurs.
Une fois le contrat signé, le client donne plus de détails sur ses besoins. Les équipes de recherche et développement vont pouvoir concevoir l'architecture totale du satellite, et simuler son comportement une fois dans l'Espace grâce à des modélisations et simulations thermiques et mécaniques. Cela passe entre autres par la modélisation du satellite en trois dimensions pour le visualiser dans sa globalité, mais aussi pour identifier l'emplacement de chacun de ses équipements. Ensuite, la modélisation mécanique structurelle permet de valider la résistance de la structure vis-à-vis des différentes contraintes qui s'appliquent sur elle. Puis la modélisation cinématique offre une visualisation des mouvements du satellite, comme par exemple le déploiement des panneaux solaires et des antennes après lancement, ou encore les mouvements permettant l'orientation des instruments optiques. Enfin, le modèle thermique valide la résistance de la structure et des équipements aux variations de températures que le satellite connaîtra dans l'Espace, beaucoup plus extrêmes que celles observées sur Terre.
A la fin de cette étape, une architecture préliminaire est proposée. Après quelques itérations avec le client, et une fois la faisabilité agréée par toutes les parties, le satellite est prêt à voir le jour.

1.2.2 Assemblage

Une fois qu'une solution préliminaire est approuvée, le satellite peut être assemblé. Un satellite est conçu pour être le plus compact possible, pour faciliter son lancement. L'as-

semblage doit donc être réalisé en suivant scrupuleusement ce qui a été décidé lors de la conception du satellite. La modélisation du satellite en trois dimensions apporte donc de l'aide précieuse dans cette tâche. De la même manière, l'assemblage des différentes sous-parties doit être réalisé suivant un ordre prédéfini.
Les satellites sont assemblés en salle blanche afin de limiter les pollutions générées par des particules non désirables, et les techniciens qui manipulent les satellites disposent d'équipements de protection appropriés.
Enfin, pour inspecter l'intégration du satellite, on lui fait subir une série de tests pour valider sa conformité vis-à-vis de la spécification préalablement établie. Le principe est de s'assurer et qu'il pourra endurer les dures conditions de vie dans l'Espace et les conditions de son lancement, qui peuvent être très éprouvantes, notamment d'un point de vue mécanique et thermique. Ces tests concernent aussi bien les équipements de la charge utile que de la plateforme, et en général on différencie les tests des équipements radio-fréquence des tests mécaniques. Tout d'abord, les tests des équipements radio-fréquence sont réalisés afin de contrôler les performances des équipements de la charge utile. Ensuite, les tests mécaniques ont pour but de valider la résistance du satellite aux phénomènes rencontrés dans l'Espace : vibrations, fortes variations de températures, adaptation du satellite lorsqu'il est soumis à de fortes interférences.
Certains tests ciblent des équipements particuliers au sein du satellite, et d'autres tests concernent le satellite dans sa globalité.
Les tests peuvent être réalisés à vide, dans différentes conditions de température et de pression. Toutes les fonctionnalités du satellite sont testées pour s'assurer que le satellite réagit selon ce qui était attendu. A titre d'exemple, la figure 1.7 représente une chambre anechoïque à Intespace destinée aux tests effectués sur les antennes. Cette pièce est une base compacte de mesures d'antennes, conçue pour simuler la situation du vide spatial d'un point de vue radio-fréquence. Tous les signaux tels que les ondes téléphoniques et radios sont absorbées par les structures situées sur les murs, au sol et au plafond.
Les tests peuvent durer plusieurs semaines, et génèrent une quantité significative des données produites au cours de la vie du satellite.
Une fois que les tests ont été réalisés et les résultats validés par le client, le satellite est prêt à être acheminé vers son lieu de lancement, par exemple Kourou en Guyane ou Baïkonour au Kazakhstan.

1.2.3 Lancement et opérations

Une fois assemblé et testé, il ne reste plus au satellite qu'à être propulsé dans l'Espace, grâce à un lanceur à l'instar d'Ariane. Une fois lancé, il doit atteindre son orbite. Nous avons vu précédemment que cette phase était suivie de très près.
Le satellite réalise lui-même des manœuvres pour atteindre l'orbite à laquelle il est affecté. Il se déplace, comme nous l'avons décrit précédemment, à l'aide de la propulsion. Il peut ensuite commencer sa mission.

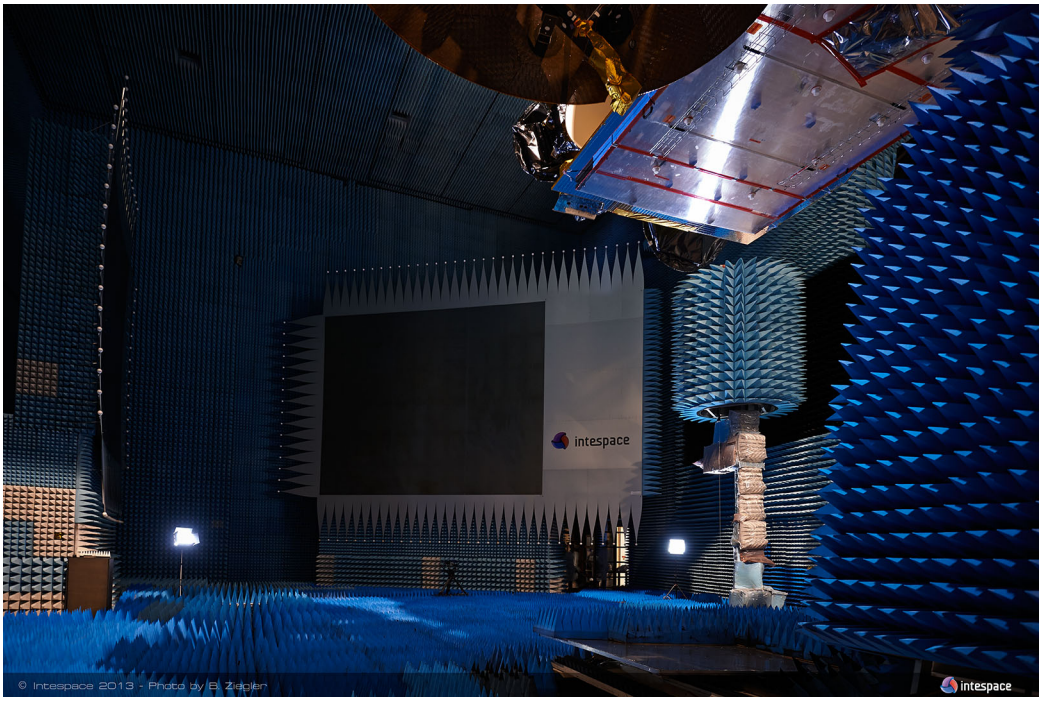

FIGURE 1.7 – Tests d'une antenne dans une chambre anéchoïque. Les structures en relief sur les murs empêchent la réflexion des ondes pour simuler le vide spatial.

La mission d'un satellite dure en général entre 10 et 15 ans. Il arrive qu'elle s'étende au-delà lorsque le satellite est toujours fonctionnel et en bonne santé. Le satellite communique avec la Terre grâce à ses antennes et un segment sol installé sur Terre qui se charge de collecter et traiter les données. Ces données sont des signaux de télécommunications, tels que des signaux télévision ou téléphoniques pour les satellites de télécommunications, des images pour les satellites d'observation. De plus, quelque soit la nature du satellite, il envoie, en plus des signaux relatifs à sa mission, des milliers d'indicateurs pour attester de son bon fonctionnement : ce sont les télémesures.
Quand le satellite arrive en fin de vie, il est désorbité. Les satellites situés sur une orbite basse se désintégreront en s'approchant de l'atmosphère, alors que les satellites géostationnaires, trop éloignés de la Terre, seront envoyés plus loin encore de la surface de notre planète, pour ne pas surcharger l'orbite géostationnaire.

FIGURE 1.8 – Salle de contrôle de l'Agence Spatiale Européenne (ESA).

1.3 Optimiser les phases de la vie du satellite grâce à l'analyse de données

Pendant les phases de la vie du satellite, de nombreuses données sont générées. Nous avons pu voir que ces données peuvent être très hétérogènes. Par exemple, les données de spécifications provenant des appels d'offres ont peu de variables, et peuvent contenir des valeurs numériques aussi bien que des données de statut, tel que la présence ou absence d'un équipement spécifique, par exemple. Les modèles provenant de la phase de conception contiennent un grand nombre de données sur la géométrie et le comportement du satellite, mais doivent être interprétées à partir des modèles.
Enfin, les tests génèrent une grande quantité de données puisque de très nombreuses mesures sont réalisées. Les sorties de ces tests sont généralement des variables numériques et plus particulièrement des courbes, donc des données fonctionnelles. En dernier lieu, les télémesures envoyées par le satellite tout au long de sa vie sont des séries temporelles. Les satellites d'observation envoient régulièrement des images au cours de sa mission qui peuvent faire l'objet de traitements particuliers.
Dans cette section, nous évoquerons quelques applications d'utilisation de ces données pour améliorer les phases de la vie du satellite. Ces applications ont pu bénéficier d'une étude en complément de la thèse, et s'appliquent aux mêmes familles de données : données provenant des tests et des télémesures spatiales.

1.3.1 Eviter certains tests

Comme évoqué précédemment, les tests sont très importants pour valider le bon fonctionnement du satellite au sol. Cependant, certains de ces tests peuvent être considérés comme redondants. Par exemple, certains des équipements ont déjà été testés par les fournisseurs, et sont parfois testés à nouveau individuellement avant d'être intégrés au satellite. Des modèles d'apprentissage pourraient justifier que les résultats obtenus sont les mêmes que ceux réalisés chez le fournisseur, évitant ainsi de réaliser des tests redondants parfois coûteux.
Un autre exemple est celui des tests réalisés sur les guides d'ondes, qui sont eux aussi testés au cas par cas en fonction de leur géométrie. Les guides d'ondes fonctionnent comme des câbles électriques, mais sont en fait des structures creuses et rigides qui vont permettre de diffuser le signal. Certains guides d'ondes peuvent avoir une géométrie très complexe, comprenant plusieurs coudes, par exemple. A l'heure actuelle, les experts décident au cas par cas si les guides d'ondes doivent être soumis à des simulations numériques de stress vibratoire et thermique. Les guides d'ondes ayant des caractéristiques géométriques plus simples, ou redondants d'un satellite à un autre peuvent être dispensés d'être sujets à ces simulations. Des modèles d'apprentissage peuvent apprendre le jugement humain et décider à sa place si un nouveau guide d'onde nécessite une simulation numérique, sur la base de sa géométrie. De plus, il pourrait également dans un second temps prédire la sortie de ces tests. Il est donc possible de réduire considérablement le nombre de tests si la prédiction des sorties est jugée suffisamment fiable pour une partie de ces guides d'ondes.

Une autre application pourrait être d'utiliser les différentes itérations d'une phase de test afin d'identifier les étapes les plus indispensables. En effet, il arrive qu'un même test soit réalisé plusieurs fois dans des contextes différents, de température ou de pression par exemple. Cette situation est schématisée en Figure 1.9. Il serait intéressant d'en déduire quel est l'ensemble minimum de ces itérations nécessaire à observer tous les types de comportements que peut avoir le test, ou encore de développer un modèle prédictif fidèle pour les tests qui ont été abandonnés, à l'image du schéma présenté Figure 1.10. Il existe donc des pistes pour améliorer les conditions des tests. Ces améliorations permettent de réduire la durée des tests et de ne se focaliser que sur les tests les plus informatifs au sujet de l'état du satellite.
Ces exemples d'améliorations ont pu bénéficier d'une étude en complément de la thèse, via l'encadrement des stages de Anaïs Charcosset et Jérémy Pirard à Airbus.

1.3.2 Détecter des comportements atypiques

La détection d'anomalies est une des problématiques les plus importantes autour de l'analyse de données dans le domaine du spatial. C'est dans ce cadre que se situe cette thèse. Lors des phases de test et lors des opérations, une grande quantité de données est générée.

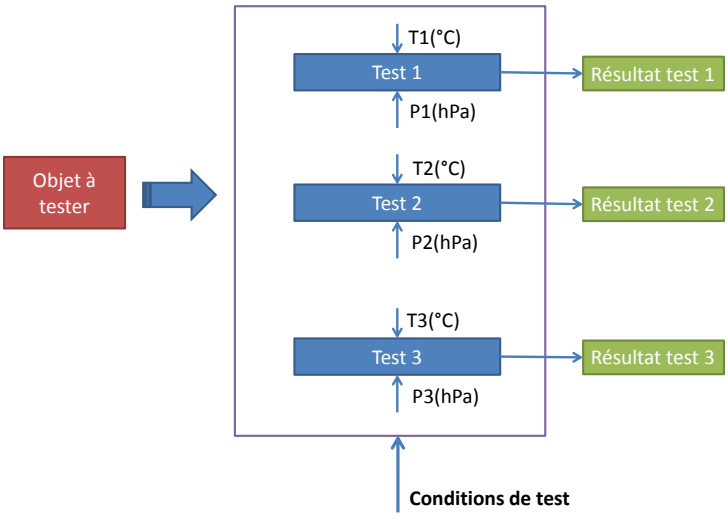

FIGURE 1.9 – Exemple de réalisation d'un test sous l'effet de plusieurs conditions de température et de pression.

Ces données sont des mesures de paramètres physiques. Si elles permettent de se figurer de l'état du satellite à un moment précis, l'historique de toutes les données nous permet de comprendre comment fonctionne un satellite, et d'analyser les situations qui ont pu mener à des pannes. Ce que l'on sait aujourd'hui, c'est que les pannes sont rares, ne se ressemblent pas et ne sont pas clairement répertoriées. La détection et l'anticipation de pannes est donc une tâche très ardue. En conséquence, le contexte d'étude est non supervisé, dans la mesure où toutes les pannes n'ont pas été observées, et donc ne peuvent être apprises. Il est donc nécessaire de développer des méthodes pour donner des indications sur l'atypicité des données et les changements de comportement que l'on peut observer sur le satellite. Cette problématique étant le cœur de la thèse, elle sera détaillée dans la prochaine section.
Les études de fiabilité en général font également partie des cas d'analyse de données applicables au domaine spatial.

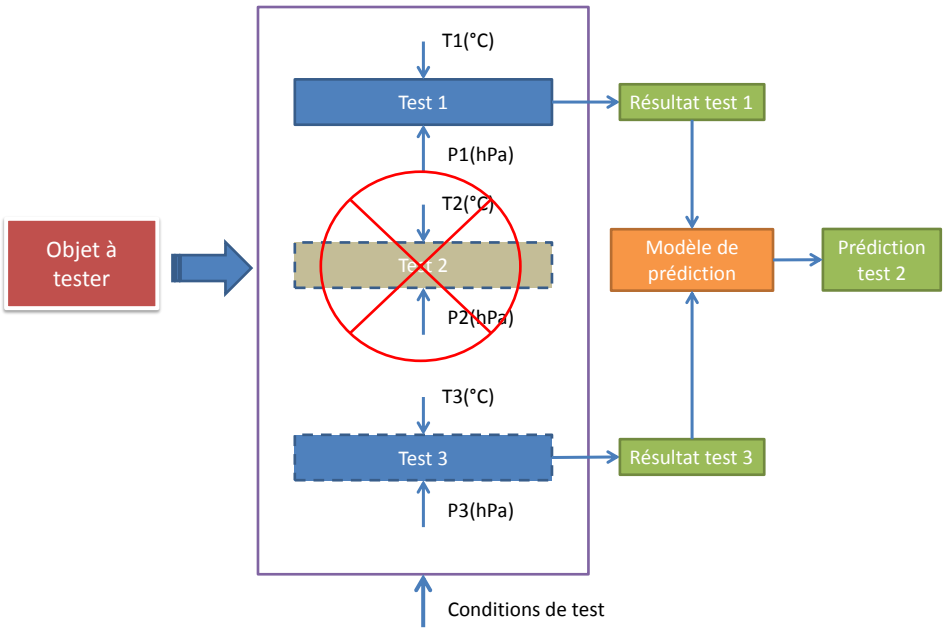

FIGURE 1.10 – Exemple de prédiction des résultats d'un test à partir des résultats des autres itérations de ce test.

1.4 Détection d'anomalies dans les données spatiales

1.4.1 Enjeux industriels

Détecter les divergences de comportement du satellite peut avoir plusieurs buts.
Le premier apport de la détection automatique d'anomalies dans les satellites est un gain de temps. En effet, la revue des tests et le suivi quotidien des satellites sont des activités coûteuses puisqu'une grande partie de ces travaux est réalisée au cas par cas par des experts.
Des méthodes automatiques permettraient aux ingénieurs de se focaliser uniquement sur les données les plus alarmantes. En effet, que ce soit pour les tests ou pour le suivi en vol des satellites, seulement une faible proportion des données requiert une attention particulière à un instant donné. Ordonner les observations en fonction des priorités offre un gain de temps considérable pour les experts chargés du suivi des satellites.
Le contexte industriel confirme ce besoin, puisque les satellites sont de plus en plus complexes, ou nombreux à surveiller, avec notamment l'apparition de constellations de satellites. L'ordonnancement des données en fonction de leur criticité est dans ce contexte une problématique croissante.

L'automatisation de la surveillance des données peut apporter de nouveaux points de vue lors de l'investigation d'anomalies, et rendre ces investigations plus efficaces. Lorsqu'une réelle panne est avérée, il est parfois difficile de retracer le scénario qui a conduit à cette panne. L'élaboration de ces scénarios permet d'éviter qu'ils ne se reproduisent sur d'autres satellites. Avec l'étude de ces données, on peut être capable de capter toutes les divergences, d'identifier les équipements qui ont été impactés et dans quel ordre. Ces investigations peuvent être une source de stress pour les équipes concernées, les clients attendant une explication le plus tôt possible quant à l'origine de la panne. Des méthodes automatiques permettent donc aux ingénieurs d'organiser leur travail et de les orienter vers des scénarios de pannes possibles.
L'analyse croisée des différentes sources de données peut offrir de très nombreuses opportunités. Il s'agit par exemple d'exploiter les données issues des tests et des opérations conjointement. En premier lieu, on peut qualifier l'état de santé global du satellite, en utilisant toutes les sources de données, telles que les données provenant des tests et des opérations. Le croisement des données permet d'offrir une vision de l'état de santé global du satellite, en mettant en lumière les fragilités du satellite au cours du temps, et prendre ainsi en compte son vieillissement. Ainsi, les études de fiabilité peuvent être renforcées grâce à l'identification des équipements les plus fragiles en général, ou encore les équipements se dégradant plus rapidement que leurs homologues, au sein de constellations par exemple.

1.4.2 Définition de la donnée

Dans cette thèse, nous disposons de données provenant de deux sources différentes, les données de tests et les télémesures satellites. Dans cette section, nous tâcherons de donner une description précise des deux sources de données, de leurs modes d'acquisitions ainsi que de leur représentation mathématique.

1.4.2.1 Tests Radio-Fréquence des satellites télécoms

Les données que nous traitons dans le cadre de la thèse sont des données provenant des tests réalisés sur les équipements radio-fréquence des satellites de télécommunications. Les satellites concernés sont donc uniquement des satellites de télécommunications géostationnaires, bien que les mêmes méthodes puissent être généralisées à tous les satellites. Pour les satellites de télécommunications, le principe général de validation est d'envoyer des signaux Radio-Fréquence (RF) de référence à l'entrée du répéteur. Ces signaux (ou porteuses pures) représentent les signaux envoyés par une station sol et captés par le satellite au travers de son antenne de réception. On mesure alors le signal de sortie qui, après amplification, est renvoyé vers la zone de couverture terrestre, au travers de l'antenne de transmission. Les signaux provenant d'une station sol sont appelés Uplink, et les signaux renvoyés au sol sont appelés Downlink.

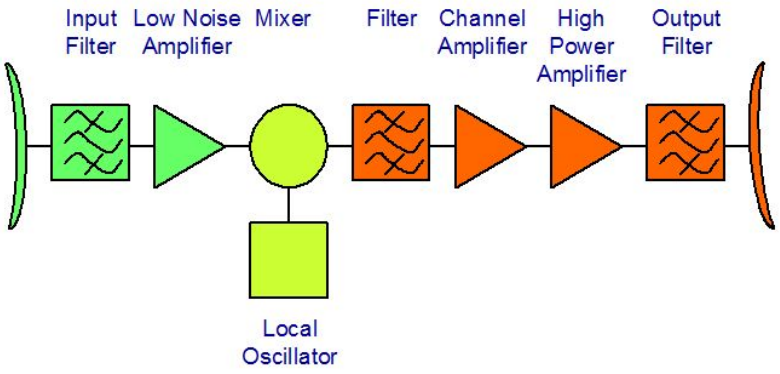

FIGURE 1.11 – Exemple d'une chaine d'amplification.

Parmi les tests réalisés pour valider le satellite, le test le plus complexe et le plus informatif est le test Gain-Fréquence, car il s'applique à toute une chaîne d'amplification, de la réception du signal d'entrée vers le renvoi du signal de sortie. La composition schématique d'une chaîne d'amplification est représentée Figure 1.11.
En pratique, les chaînes d'amplifications sont liées entre elles dans la charge utile pour permettre d'y ajouter par exemple des équipements redondants et pour reconfigurer l'utilisation de la charge utile grâce à l'utilisation de switchs : des interrupteurs qui orientent le signal vers une des directions possibles. Une charge utile peut être schématisée selon la représentation donnée en Figure 1.12. Ainsi, plusieurs chaines d'amplifications peuvent avoir des équipements en commun.

Le test Gain-Fréquence (GF) réalise un balayage en fréquence et permet d'observer le gain entre le signal d'entrée et le signal de sortie sur une bande passante donnée. L'objectif est de vérifier la stabilité du gain dans la bande passante et de détecter les comportements atypiques tels que des non-linéarités, oscillations, pentes. Un exemple de courbe obtenue est représenté en Figure 1.13. Pour des raisons de confidentialité, les vraies échelles ont été effacées. La courbe principale est représentée en rouge et sa dérivée en vert. La courbe marron représente la spécification du client. Visuellement, cela signifie que la courbe principale doit se situer au-dessus de cette courbe pour que le test soit conforme à la spécification. Il existe souvent une spécification sur la dérivée également, représentée en pointillés irréguliers ici. La pente doit alors se situer en-dessous de cette spécification.

Dans la pratique, le gain est un bon indicateur de performance du répéteur. Il s'agit de vérifier l'impact du chemin parcouru par le signal au travers des équipements passifs (guides d'ondes, cables coaxiaux...) ainsi que l'impact des équipements actifs traversés sur cette performance attendue. La réponse idéale du Gain-Fréquence serait une indicatrice sur la performance. L'objectif de cette application est donc de trouver les résultats les plus dégradés parmi les courbes résultant du Gain-Fréquence. En utilisant la connaissance métier, on pourrait réaliser une méthode paramétrique de détection d'anomalies en réalisant

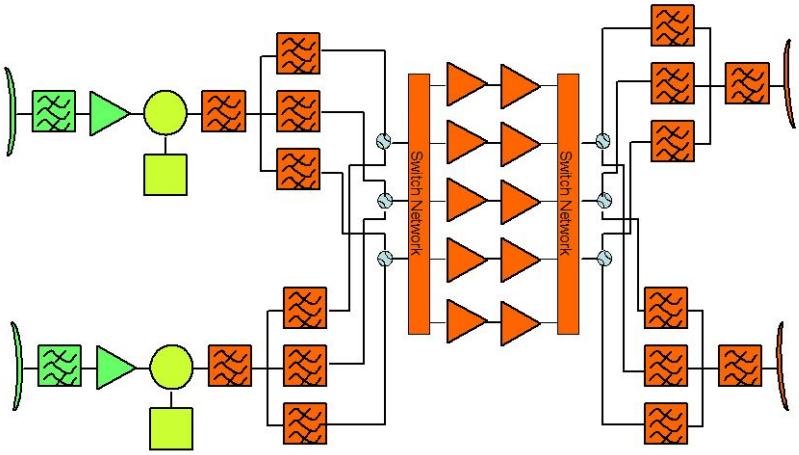

FIGURE 1.12 – Schéma simplifié d'une charge utile.

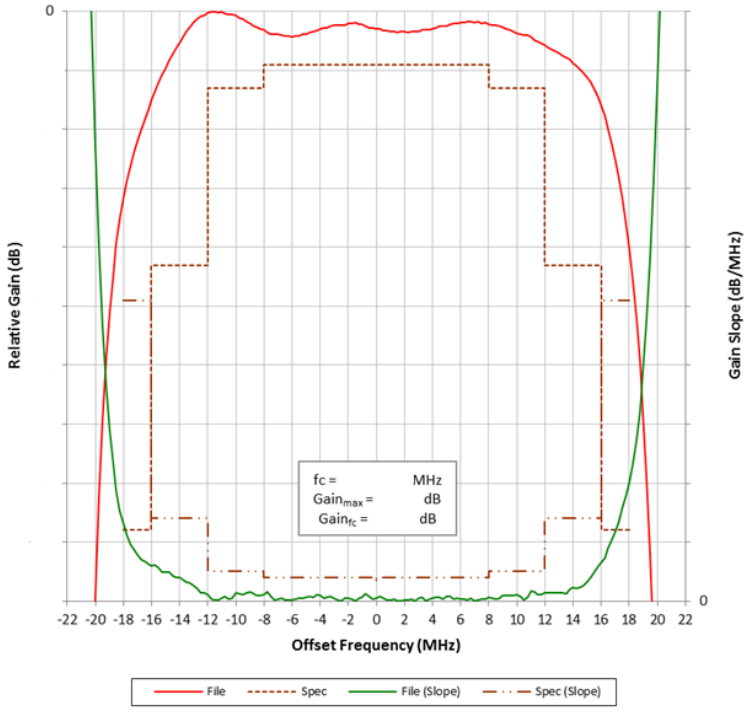

FIGURE 1.13 – Exemple d'une courbe Gain-Fréquence (en rouge), ainsi que sa dérivée (en vert) et la spécification client (en marron).

par exemple un calcul de distance par rapport au signal idéal. Cependant, avec ce genre

d'approche on se prive de rendre générique les méthodes de détection d'anomalies.
Le test Gain-Fréquence est réalisé plusieurs fois dans différentes conditions de température et de pression. L'évolution d'un même test en fonction des facteurs extérieurs est également intéressant pour mesurer l'impact de l'environnement extérieur sur la performance associée à la chaine d'amplification.

1.4.2.2 Télémesures satellites

Une fois le satellite en vol, il renvoie régulièrement sur Terre de très nombreuses données attestant de sa santé, appelées télémesures. Un satellite peut enregistrer plusieurs milliers de télémesures. Le nombre total de télémesures dépend de la taille du satellite et de sa complexité. Ces télémesures sont échantillonnées très régulièrement, en général toutes les vingt ou trente secondes. On peut estimer qu'un satellite de télécommunications produit l'équivalent de 100 Giga-octets de mesures chaque année.
Selon une définition générale, les télémesures sont des valeurs de mesures obtenues à distance au sein de systèmes complexes. Pour les satellites, on peut classer les télémesures en deux catégories selon leur nature.

- Les télémesures numériques sont des mesures de capteurs de paramètres physiques au sein du satellite, comme par exemple les températures des différents équipements, les mesures de courant, des indicateurs relatifs à la position du satellite par rapport à la Terre ou au Soleil. Ce sont donc majoritairement des signaux continus. On dénombre également certains des signaux de comptage, qui prennent par conséquent des valeurs entières. Ces mesures peuvent concerner aussi bien la charge utile que la plateforme. Un exemple de télémesures est présenté Figure 1.14, et illustre bien la périodicité auxquelles sont soumises de nombreuses télémesures. Pour des raisons de confidentialité de données, les échelles ont été simplifiées.

- On dispose également de télémesures de statut, dont le but est de donner des informations sur la situation interne du satellite et de son environnement extérieur. Ce sont donc des données discrètes. Par exemple, on peut savoir quels sont les équipements actifs au cours du temps, si on est en période d'éclipses, ou non. Ces statuts nous permettent de mieux comprendre la situation dans laquelle le satellite se trouve.

Ces mesures sont encodées au sein du système spatial puis envoyées sur Terre sous forme binaire. Une fois reçues, les télémesures sont décodées pour retrouver la valeur physique réellement mesurée.
Le stockage des télémesures est complexe. Stocker en dur toutes les télémesures de tous les satellites était il y a quelques années impossible dans un monde où on ne parlait pas encore de "Big Data". C'est pourquoi les télémesures satellites sont stockées sous forme d'archives, et sont donc difficile d'accès, surtout quand on cherche à accéder à un

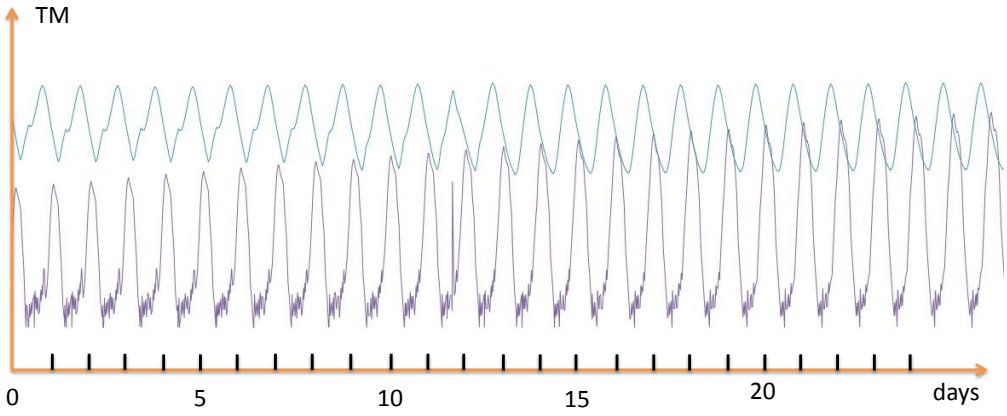

FIGURE 1.14 – Exemple de deux télémesures satellites.

historique de données. Afin de rendre une partie de la donnée immédiatement accessible, les valeurs minimum, maximum et moyenne par heure d'un grand nombre de télémesures sont directement disponibles. En revanche, pour disposer de la totalité des mesures, il faut ouvrir les archives et les "décommuter" par intervalles de temps définis, c'est à dire retrouver la vraie valeur physique à partir d'une compression en binaire. Cette opération peut prendre un temps considérable.

Le logiciel utilisé par Airbus Defence and Space permettant d'accéder aux télémesures et de réaliser leur suivi s'appelle TELMA. Outre l'aspect visualisation des données et ouverture des archives à la demande, il permet de pouvoir détecter automatiquement certaines divergences du satellite. Ces algorithmes, appelés processing, sont définis au cas par cas par les experts. Les événements détectables par ces algorithmes sont en général des dépassements de seuils sous certaines conditions. De nouveaux processing peuvent être intégrés à l'application dès qu'un nouvel événement apparaît, dans le but de pouvoir le détecter si il se reproduit.

Ces processing étant pensés au cas par cas, il est impossible d'en définir pour toutes les télémesures. Par conséquent, de nombreuses télémesures ne sont pas surveillées de manière automatique. Les méthodes que nous développons ont pour but de détecter les divergences que ces processing ne peuvent pas mettre en lumière, mais aussi d'adresser le suivi de l'intégralité des télémesures. C'est dans ce but que la thèse s'inscrit.

1.4.2.3 Modèle d'observation

Dans les deux cas d'applications, on peut considérer que l'on se place dans un cadre d'analyse de données fonctionnelles. En effet, les tests correspondent à des courbes, et les télémesures sont des séries temporelles ayant une forte périodicité. Une fois celles-ci découpées en journées (ou périodes d'orbite pour les satellites situés en orbite basse),

on obtient des données fonctionnelles. Cette approche nous permet de lier les deux cas d'applications et d'appliquer sur ces données des méthodes similaires.
Que ce soit pour les télémesures ou les tests, on considère que dans le cadre univarié, on dispose du modèle d'observation (1.1).

$$X_{i,j} = f_i(t_j) + \varepsilon_{i,j}, \tag{1.1}$$

où :

- $i = 1, \ldots, n$ est un indice permettant d'identifier le numéro du test (pour les données de test), ou bien correspondant au jour ou à la période considérée, (pour les télémesures).
- $t_j, j = 1, \ldots, p$ est le temps auquel la mesure a été acquise. Pour les données de tests, t_j peut être une fréquence. On supposera que $t_j \in [0, 1]$, c'est à dire que $t_j = j/p$, si les mesures sont acquises de manière régulière.
- f_i est une fonction de $\mathbb{L}^2([0, 1])$ observée sur p points par le vecteur $\boldsymbol{X}_i \in \mathbb{R}^p$.
- $\varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$ est un bruit gaussien de variance σ^2 inconnue, et les $\varepsilon_{i,j}$ sont supposés indépendants.

On utilisera ce modèle simplifié pour toutes les applications univariées. De plus, nous avons mentionné notre intérêt concernant l'aspect multivarié des données. Comme nous l'avons précisé, nous serons susceptibles de comparer les tests d'une phase à une autre, ou bien de comparer plusieurs télémesures entre elles. Dans ce cas, nous utiliserons la notation suivante (1.2).

$$X^{(i)}_{j,k} = f^{(i)}_k(t_j) + \varepsilon^{(i)}_{j,k}, \tag{1.2}$$

- $i = 1, \ldots, n$ est toujours l'indice permettant d'identifier le numéro du test, ou le jour (ou autre période considérée). Il est mis en exposant pour simplifier les notations qui en découlent.
- $t_j, j = 1, \ldots, p$ est également le temps (ou la fréquence) pour laquelle la mesure a été acquise.
- $f^{(i)}_k$ est une fonction de $\mathbb{L}^2([0, 1])$.
- $k = 1, ..., m$ est le numéro de la télémesure ou de la phase dans laquelle a été réalisé le test.
- $\varepsilon^{(i)}_{j,k} \sim \mathcal{N}(0, \sigma^2)$ est un bruit gaussien de variance σ^2 inconnue, et les $\varepsilon^{(i)}_{j,k}$ sont supposés indépendants.

Cette notation, plus complète que la précédente, est adaptée au point de vue multivarié. Dans ce cas, nous analyserons l'ensemble des courbes $\{f_k^{(i)}, k = 1, ..., m\}$ conjointement pour chaque journée $i = 1, ..., n$.
Que ce soit de manière univariée ou multivariée, les individus seront qualifiés en fonction de la journée ou du test i, c'est à dire qu'on cherchera à savoir si la journée ou le test a un comportement normal ou atypique.

1.4.3 Objectifs de la thèse

L'objectif de cette thèse est de pouvoir identifier, à partir d'un jeu de données de test ou de l'historique d'une ou de plusieurs télémesures satellites, les données ayant des comportements atypiques ou divergents. Nous ne disposons pas d'historique des anomalies passées. Nous savons, par la connaissance métier, que nous pouvons d'ores-et-déjà les classer en plusieurs familles d'événements atypiques :

- Les anomalies locales ne s'observent que sur une petite partie de la donnée, sur un nombre restreint de mesures. Elles peuvent s'exprimer par des pics, des décrochages dans les courbes. Elles peuvent être liées à des anomalies dans la mesure, et pas nécessairement dans le système. Il est néanmoins intéressant de pouvoir les détecter.

- Les anomalies de motifs s'observent sur toute ou presque toute la courbe, sur une ou plusieurs périodes, concernant les télémesures périodiques. Il s'agira d'anomalies potentiellement plus faciles à repérer.

- Les anomalies de famille s'observent sur plusieurs données en même temps. Par exemple, pour une série de test, il s'agira de détérioration anormale du test au travers des différentes phases de tests subies. Pour une télémesure, il peut s'agir d'événements simultanés sur plusieurs télémesures, pouvant traduire un changement de comportement du satellite.

Définir une méthode générique pouvant détecter tous les types d'anomalies à partir de toutes les différentes sources de données n'est pas trivial. Il est donc nécessaire de pouvoir définir des méthodes de détection d'anomalies qui soient capables de mettre en lumière tous les types de divergences de comportement. Pour cela, les objectifs de la thèse sont multiples.

- Disposant d'une grande quantité d'observations de données fonctionnelles, il est nécessaire de pouvoir réduire la dimension de ces données en ne conservant qu'un nombre restreint de paramètres représentant la donnée. En effet, les vecteurs comportant les observations de la donnée peuvent en pratique être analysés directement, mais la quantité de données est trop importante et redondante pour identifier clairement les anomalies.
  Les paramètres ou coefficients que l'on va sélectionner devront être judicieusement

choisis de manière à pouvoir détenir de l'information sur les anomalies, quelles que soient les aspérités, locales ou plus globales. Ces paramètres que l'on sélectionnera pour la détection d'anomalies seront désignés comme les indicateurs de la fonction.

- Appliquer une méthode de détection d'anomalies adaptée à la structure de nos données sur les indicateurs préalablement définis.

- Définir une méthode de détection d'anomalies dans un contexte multivarié.

Outre la production de méthodes et d'algorithmes pour l'aide à l'identification d'anomalies, il est attendu que les résultats soient utilisables par les ingénieurs chargés du suivi des satellites et des tests. Pour cela, les méthodes doivent avoir des sorties facilement interprétables, tel que des scores ou des probabilités. Les méthodes peuvent être paramétrables par ces ingénieurs, tant que les paramètres ne sont pas trop nombreux et qu'ils restent compréhensibles par tous les corps de métiers.
Dans un contexte se rapprochant du "Big Data", les méthodes développées seront parallélisées grâce à l'utilisation de Spark. Le développement des algorithmes est réalisé en tenant compte de ses contraintes de parallélisation.

1.4.4 État de l'art

Nous disposons de données fonctionnelles et de séries temporelles dans lesquelles nous souhaitons détecter des anomalies. Dans cette section, nous proposons une revue des méthodes et des axes de recherches exploitables dans ce cadre de travail, du traitement préalable de la donnée jusqu'aux interprétations des méthodes.

1.4.4.1 Données fonctionnelles et séries temporelles

Les données dont nous disposons sont donc des observations de données fonctionnelles et de séries temporelles. Il est donc pertinent d'exploiter au préalable les méthodes relatant de l'analyse de données fonctionnelles (FDA) et des séries temporelles.
De manière générale, l'analyse de données fonctionnelles suscite beaucoup d'intérêt depuis quelques décennies. La multiplication des moyens de mesures, l'augmentation de leur précision, de leur fréquence d'échantillonnage et des moyens de stockage a vu augmenter le nombre de données fonctionnelles dans beaucoup de domaines. De très nombreux ouvrages traitent de l'analyse de données fonctionnelles au sens large, tel que l'ouvrage de Ramsay et Silverman [68], ou celui de Ferraty et Vieu [29].
L'analyse des séries temporelles suscite tout autant d'intérêt. Cependant, les méthodes les plus fréquemment appliquées aux séries temporelles sont sensiblement différentes de celles utilisées pour les données fonctionnelles, comme le montrent des ouvrages tels que celui d'Hamilton [39].
Il est intéressant de noter que les séries temporelles et les données fonctionnelles sont

néanmoins deux types de données étroitement liés. En effet, dans les deux cas, ces données proviennent de l'observation d'un paramètre en fonction de la variation d'un autre paramètre, qui est systématiquement le temps pour les séries temporelles. Grâce à cette similitude, on peut remarquer que des méthodes identiques peuvent être utilisées pour ces deux sources de données, tel que le montre la thèse de Allen [2]. Dans notre cas, les séries temporelles sont de plus sujettes à une périodicité forte et sont échantillonnées fréquemment, ce qui renforce l'analogie avec les données fonctionnelles.
Parmi les applications usuelles aux données fonctionnelles et aux séries temporelles, un grande nombre gravite autour du pré-traitement de ces données : méthodes pour lisser les courbes par l'usage de splines ou de méthodes à moindres carrés, analyse fréquentielle, utilisation de la décomposition de Fourier et du périodogramme [12], méthodes pour recaler les courbes notamment grâce au dynamic time warping [87] [60], calcul des composantes périodiques grâce à l'autocorrélation... Ces méthodes sont largement utilisées dans la littérature mais sont en apparence assez éloignées de notre problématique. Elles peuvent néanmoins être exploitées lors de la préparation de la donnée, dans certaines situations où la donnée est difficilement utilisable telle quelle.

1.4.4.2 Classification de courbes grâce à des projections

Il existe dans la littérature une grande variété d'applications relevant de la classification de courbes, plus proche de notre problématique. Certains auteurs tels que Tarpey [81] et Jacques et Préda [46] ont réalisé une revue des méthodes de classification de courbes. Dans ce deuxième ouvrage, Jacques et al. ont proposé une arborescence des méthodes de classifications de données fonctionnelles, que l'on a repris Figure 1.15. Pour ce type d'applications, on remarquera qu'une des manières de procéder, appelée dans le schéma méthode 2-step, est d'exprimer les données fonctionnelles par leur développement dans des bases fonctionnelles, le plus souvent orthonormées. Ensuite, les projections des fonctions dans ces bases sont utilisées comme des représentants de ces fonctions, permettant ainsi de réduire la dimension des données.
Cette méthode permet de résumer au mieux les courbes et de réduire la dimension des données. Ce type d'approche se retrouve dans beaucoup de travaux, tellr que l'étude menée par Auder et al. [6] où de nombreuses bases fonctionnelles sont testées, ou encore les travaux de Gonzàlez et al. [36] où des bases de fonctions construites dans des espaces de Hilbert à noyaux auto-reproduisant (RKHS) sont choisies, ou bien encore chez Antoniadis [4] où les bases de fonctions considérées sont des bases d'ondelettes. Il est également possible de construire les bases de fonctions directement à partir des données, grâce à l'analyse en composantes principales fonctionnelle [68], ou bien l'analyse en composantes principales à noyau [74].
Il existe déjà des études purement basées sur la détection de données atypiques à partir des projections sur des bases fonctionnelles, tels que les travaux de Hoffmann [43], Ren [69], ou même Hubert [44], ce dernier traitant déjà de l'aspect multivarié des données

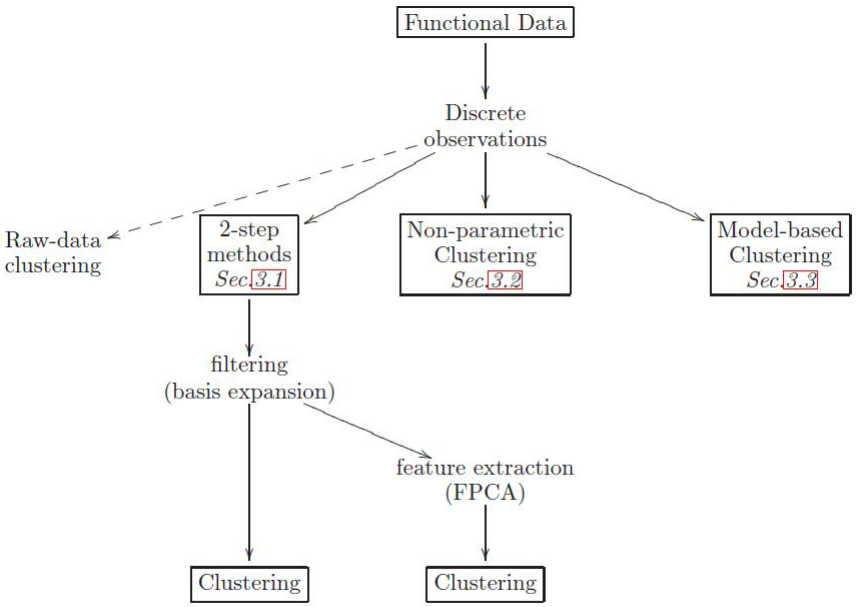

Figure 2: Classification of the different clustering methods for functional data.

FIGURE 1.15 – Méthodes de classifications de données fonctionnelles selon Jacques et al.

fonctionnelles grâce à l'utilisation d'une méthode basée sur le calcul de la profondeur des courbes par rapport à un comportement nominal.

1.4.4.3 Détection de données atypiques

Les méthodes pour détecter des données atypiques peuvent être très nombreuses et varient beaucoup en fonction du point de vue que l'on choisit et du contexte. Il existe des méthodes paramétriques, non paramétriques, supervisées, semi-supervisées, bayésiennes etc...
De nombreux articles proposent une revue exhaustive de toutes les méthodes que l'on peut trouver dans la littérature, tels que [8], [40], [63], [56] ou bien [26].
Concernant notre problématique, les méthodes paramétriques ou supervisées semblent difficiles à appliquer. Parmi les méthodes les plus adaptées à notre problématique, on peut nommer toutes les méthodes basées sur des notions de distance et de densité, parmi lesquelles on peut se référer à [80] ou encore [48]. Parmi ces exemples, les méthodes One-Class SVM [75] et Isolation Forest [52] sont des applications de méthodes d'apprentissage supervisées (SVM et Random Forest) à la détection de données atypiques. D'autres méthodes s'appuyent sur le calcul de score d'atypicité, tels que le Local Outlier Factor [14], le Local Outlier Probability [49], et l'utilisation de la distance de Mahalanobis [38].
On peut également citer la notion de profondeur (depth) qui est étroitement liée à la no-

tion d'anomalie, car plus un individu est facilement séparable des autres données, plus il est atypique. Voir [53] et [18] pour plus de détails. Febrero et al. [28] ont également utilisé la valeur de profondeur pour détecter des anomalies dans des données fonctionnelles.
Certaines études sont particulièrement ciblées sur la détection d'anomalies dans des séries temporelles, telles que [76], où l'auteur a choisi d'appliquer la One-Class SVM sur des portions glissantes de signal. Certains auteurs utilisent des modèles autorégressifs pour détecter automatiquement les anomalies à partir de l'erreur de prédiction.
Une revue des méthodes envisageables pour les séries temporelles est proposée par Gupta et al. [37].

1.4.4.4 Contexte industriel similaire

Nous nous intéressons également aux études réalisées dans le cadre d'applications industrielles similaires. Parmi ces applications, il existe des études qui ont été réalisées spécifiquement sur des données provenant de télémesures spatiales, telles que les études réalisées par l'ESA (l'agence spatiale européenne) [57], le CNES (l'agence spatiale française) [32] et le JAXA (l'agence spatiale japonaise) [33]. Dans ces études, on retrouve la notion d'indicateurs constuits à partir des données, de manière fixe pour les deux premières études, et calculés grâce à des noyaux pour l'étude de l'agence spatiale japonaise. Il existe des initiatives similaires dans le secteur aéronautique, comme par exemple l'étude de Thomas et al. [82] pour Airbus, ou encore la thèse de Rabenoro [65] ainsi que les travaux de Abdel-Sayed et al. [1] pour Safran concernant la maintenance prédictive des moteurs d'avion.
Pour résumer cet éventail de méthodes, nous proposons la figure 1.16 pour illustrer l'arborescence des différentes approches citées et identifiés, ainsi que les liens que l'on peut trouver entre elles. Bien sûr, cette représentation n'est pas complètement exhaustive mais elle permet de se représenter le contexte dans lequel nous nous plaçons.

1.4.4.5 Analyse multivariée

Nous nous intéressons également à l'aspect multivarié des données pour détecter des anomalies. Une méthode basée sur des modèles auto-régressifs multivariés est proposée par Galeano et al. [34] pour la détection d'anomalies dans les séries temporelles multivariées. Comme nous l'avons déjà cité, Hubert et al. [44] ont également exploité l'aspect multivarié pour les données fonctionnelles.
Une autre approche est d'utiliser la covariance des données pour la détection d'anomalies. Fremdt et al. [30], Ilea [45], mais aussi Cai [16] ont proposé des tests statistiques s'appliquant sur la covariance de données fonctionnelles, que l'on peut utiliser pour détecter des anomalies. De manière générale, la détection d'anomalies dans des données fonctionnelles multivariées a été peu exploitée dans la littérature.

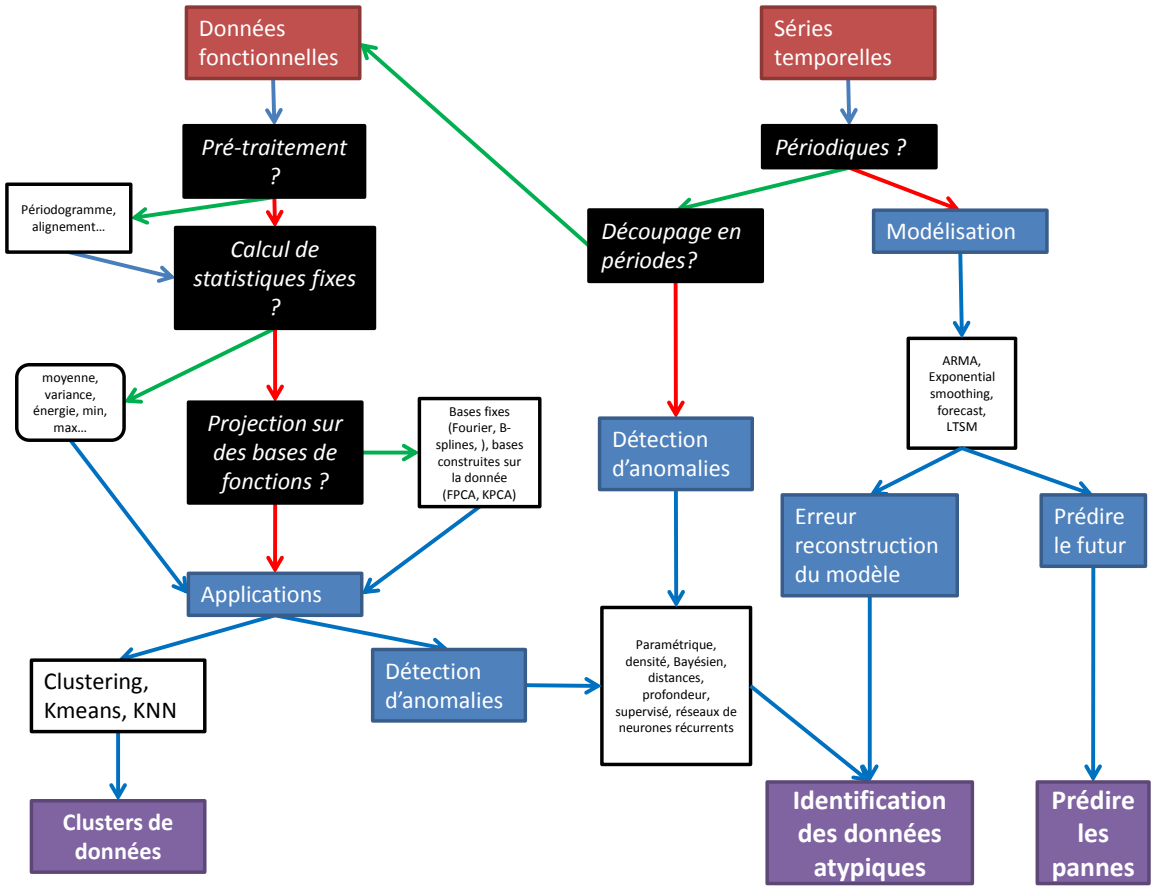

FIGURE 1.16 – Arborescence des méthodes et des différentes approches identifiés pour détecter des anomalies dans des données fonctionnelles et les séries temporelles.

1.4.5 Méthode développée

Dans cette section, nous détaillons différentes approches que nous avons choisi pour répondre à notre problématique.

1.4.5.1 Extraction de coefficients issus de projections

Comme pour beaucoup des applications citées précédemment, une partie du travail consiste à utiliser les coefficients issus de projections dans des bases orthonormées de fonctions ou de vecteurs. Nous choisissons de considérer plusieurs bases afin d'identifier les familles de comportements naturellement isolés par chacune de ces bases. A partir du modèle 1.1 que nous avons défini précédemment Section 1.4.2.3, cela revient, pour chaque courbe $f_i \in \mathbb{L}^2([0,1])$ pouvant correspondre à la période $1 \leq i \leq n$ d'une télémesure choisie, à

utiliser le développement suivant :

$$f_i(t) = \sum_{\lambda \in \Lambda} \theta_{i,\lambda} \phi_\lambda(t),$$

où Λ est l'ensemble des niveaux de coefficients, où $\text{card}(\Lambda) = p$, $(\phi_\lambda)_{\lambda \in \Lambda}$ est une base orthonormée dans $\mathbb{L}^2([0,1])$ et où $\theta_{i,\lambda} = \langle f_i, \phi_\lambda \rangle$ si $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans $\mathbb{L}^2([0,1])$. A partir des observations obéissant au modèle (1.1), les coefficients $\theta_{i,\lambda}$ sont estimés par $\hat{\theta}_{i,\lambda} = \frac{1}{p} \sum_{j=1}^{p} X_{i,j} \phi_\lambda(t_j)$.
Nous considérons également les coefficients obtenus après projection non pas des courbes observées f_i, mais des courbes pré-traitées $\tilde{f}_i$. Par exemple, les $\tilde{f}_i$ peuvent correspondre aux courbes alignées grâce au Dynamic Time Warping, ou alors au périodogramme des fonctions f_i, $i = 1...n$. Concernant les données alignées, le Dynamic Time Warping permet de recaler les observations X_i pour chaque individu $i = 1, ..., n$, mais on peut supposer qu'il existe des fonctions sous-jacentes $\tilde{f}_i$ auxquelles sont sujettes ces observations, une fois alignées. Cela revient à la transformation suivante

$$(X_i, f_i) \xrightarrow{\text{transf.}} \left(\tilde{X}_i, \tilde{f}_i(t) = \sum_{\lambda \in \Lambda} \theta_{i,\lambda} \phi_\lambda(t) \right).$$

Les différentes bases de fonctions fixes que l'on considère sont les suivantes : Fourier, bases à noyaux, ondelettes (Haar et Daubechies).
Concernant les bases construites sur les données, nous utilisons l'analyse en composantes principales et l'analyse en composantes principales à noyau sur les données brutes, mais aussi sur les données recalées et le périodogramme des fonctions. Une fois tous les coefficients calculés, il nous reste à établir une règle pour savoir quels niveaux de coefficients $\lambda \in \Lambda$ on doit conserver afin d'y appliquer les méthodes de détection d'anomalies.
Le sous-ensemble de coefficients à retenir peut être défini de plusieurs manières. Dans la pratique, sélectionner les coefficients qui maximisent la variance est la méthode la plus employée, notamment pour les applications de classification. Pour la détection d'anomalies en revanche, on ne peut pas en conclure que cette manière de sélectionner les niveaux de coefficients est la plus efficace. En effet, les anomalies sont rares, n'ont pas de signature répétée, et donc ne représentent pas nécessairement une forte proportion de la variance.
Nous avons donc développé une méthode dans un contexte semi-supervisé pour ne sélectionner que les coefficients concentrant de l'information sur les données atypiques. Pour cela, il s'agit d'isoler au préalable un sous-ensemble de courbes connues comme ne comportant pas d'anomalies. Ensuite, pour chaque niveau de coefficient $\lambda \in \Lambda$, il s'agit de tester si les coefficients des deux sous-ensembles ont la même distribution : l'ensemble ne comportant pas d'anomalies et le reste des courbes. Le test de Kolmogorov-Smirnov peut convenir dans ce but, tout comme d'autres tests équivalents construits à partir de la distance de Wasserstein.
Les tests qui ont été rejetés après contrôle du taux de faux positifs constituent le sous-ensemble de coefficients, que nous appelerons features, sur lequel nous appliquerons les

méthodes de détection d'anomalies. Nous comparons plusieurs manières de sélectionner les coefficients en analysant la performance des méthodes de détection d'anomalies qui en découlent.

1.4.5.2 Application de méthodes de détection d'anomalies

On peut comparer l'efficacité de chacune de ces bases en appliquant une grande variété de méthodes de détection d'anomalies sur chacun des sous-ensembles.
Nous utilisons des méthodes basées sur des densités et des distances : classification ascendante hiérarchique avec lien simple, Local Outlier Factor et One-Class SVM. Nous avons également développé une nouvelle métrique permettant de prendre en compte l'aspect temporel des données, en pondérant les distances en fonction de l'écart temporel entre les journées de télémesures considérées. Ces méthodes sont appliquées aux sous-ensembles de coefficients préalablement définis lors de l'étape précédente. Nous pourrons donc comparer les différentes bases et les différentes manières de sélectionner les sous-ensembles de coefficients, en plus de comparer les méthodes elles-mêmes.
Certaines de ces méthodes, telles que la classification ascendante hiérarchique et One-Class SVM réalisent une discrimination binaire. C'est-à-dire qu'elles renvoient deux labels différents pour chaque courbe considérée : anomalie ou nominal. D'autres méthodes renvoient un score, dont le seuil est à déterminer pour réaliser la discrimination.

1.4.5.3 Analyse multivariée

Les télémesures et les tests peuvent être analysés de manière multivariée, c'est-à-dire que l'on va chercher à détecter des déviations de comportement dans des groupes de tests ou dans plusieurs télémesures.
Dans le cas où les courbes sont générées par un processus similaire, on peut tester l'égalité des fonctions dans un modèle de régression. Par exemple, il peut s'agir des différentes courbes provenant de plusieurs réalisations d'un même test dans des conditions différentes. Fromont et al [31] ont défini un test statistique pour tester l'égalité des distributions de deux échantillons, adapté aux données fonctionnelles grâce à l'utilisation de fonctions à noyaux. La zone de rejet du test est estimée grâce à des permutations et l'utilisation du Lemme de Romano et Wolf [70]. Nous utilisons ce test pour identifier les groupes de courbes étant les plus hétérogènes.
Dans le cas où les courbes proviennent de sources différentes, la covariance entre les données est un bon indicateur. Certains des signaux étant fortement corrélés, un changement dans la structure de corrélation peut être dû à un changement de comportement interne au satellite. Pour cela, les matrices de covariances entre plusieurs télémesures sont de bons indicateurs de la structure intrinsèque des télémesures entre elles. Plusieurs auteurs tels que Fremdt [30], Ilea [45] ou encore Cai [16], proposent de tester l'égalité de deux matrices de covariances. Les deux premières études se focalisent sur la comparaison de

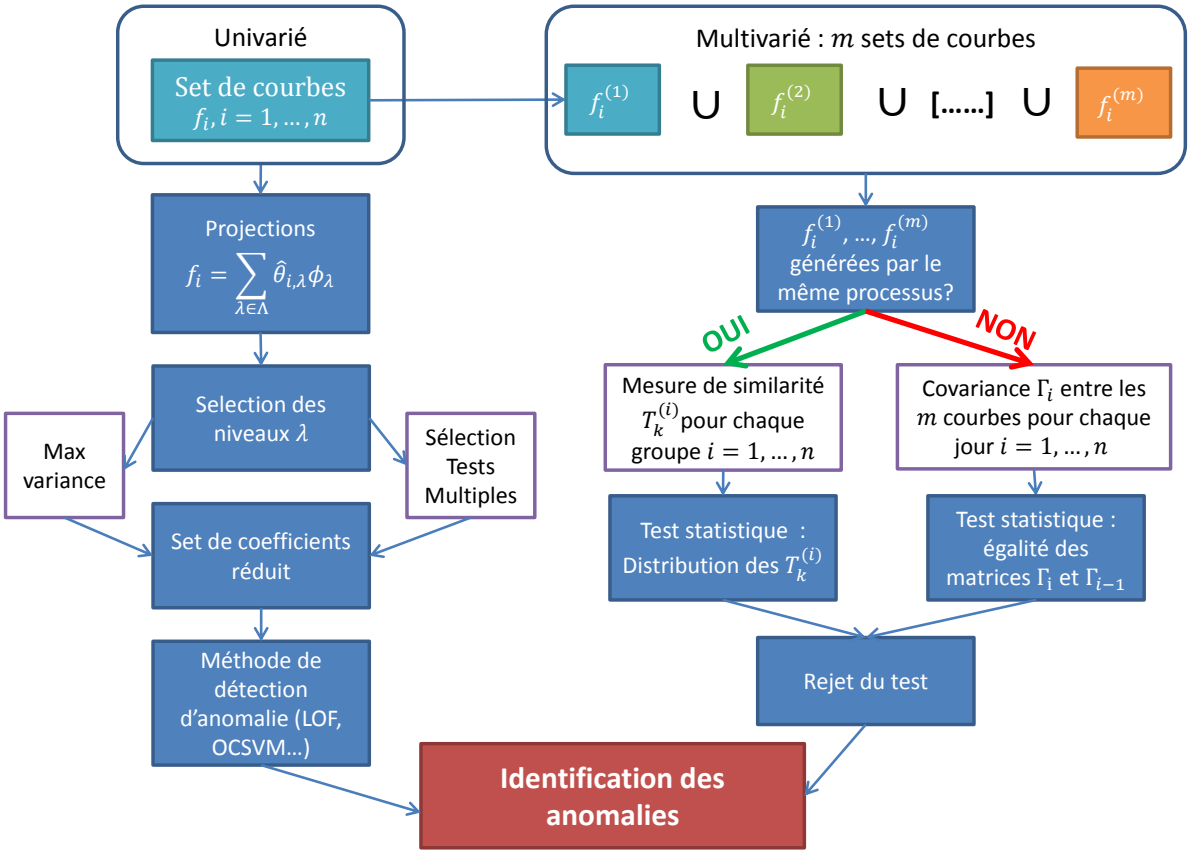

FIGURE 1.17 – Schéma de notre procédure pour détecter des anomalies dans des données fonctionnelles univariées ou multivariées.

leurs décompositions en valeurs propres, alors que Cai [16] propose une autre approche plus parcimonieuse se basant sur la plus grande différence entre les deux matrices. Nous nous inspirons des deux premiers tests pour développer un nouveau test adapté à notre problématique et plus performant, car il permet en particulier de se passer d'inversions matricielles, coûteuses et non applicables dans le cas de matrices mal conditionnées.
Le schéma 1.17 résume la méthodologie utilisée tout au long de la thèse pour la détection d'anomalies dans les données fonctionnelles, univariées ou multivariées.

1.5 Quelques résultats

1.5.1 Comparaison des bases fonctionnelles dans leur aptitude à isoler les anomalies

La première étape de cette thèse consiste à réaliser une revue de la variété des projections que l'on peut appliquer sur les courbes de notre étude afin de mettre en lumière au mieux les anomalies dès les premiers coefficients. Pour cela, nous avons construit une grande quantité d'ensembles à partir de ces différentes projections. Nous avons utilisé des bases de fonctions fixes : Fourier, ondelettes, bases de fonctions à noyaux. Nous avons également utilisé des projections sur des bases de vecteurs orthonormés construits directement à partir de la donnée, telle que l'ACP (Analyse en Composantes Principales) et l'ACP à noyaux. Nous avons également projeté les données pré-traitées, telles que le périodogramme des fonctions et les données recalées grâce aux fonctions SRVF (Square Root Velocity Functions).
Nous avons pu remarquer à l'aide d'un exemple simulé que les anomalies les plus évidentes étaient isolées par toutes les projections, dès les premiers niveaux de coefficients. Cependant, la majorité des projections choisies permettent de détecter seulement les anomalies les plus aberrantes. Les anomalies locales telles qu'une augmentation ponctuelle du bruit, des décrochages, sont donc plus difficiles à mettre en lumière dans les tout premiers coefficients de ces projections.
Les coefficients issus de la projection du périodogramme ainsi que les hauts niveaux d'ondelettes sont les choix les plus judicieux pour détecter des anomalies locales. Comme nous le montrent les figures 1.18 et 1.19, seuls les coefficients issus de la projection du périodogramme des données isolent clairement les courbes comprenant des anomalies locales du reste des courbes, parmi les projections représentées. Les coefficients issus de l'ACP à noyaux permettent également d'isoler les anomalies locales, beaucoup plus légèrement cependant. Ces résultats nous ont inspiré la seconde partie de cette thèse, dont l'idée est de détecter automatiquement les niveaux de coefficients les plus intéressants pour la détection d'anomalies. Nous aborderons les résultats de cette partie au paragraphe 1.5.3.

1.5.2 Détection d'anomalies

Nous avons également établi une comparaison de plusieurs méthodes de détection d'anomalies, parmi lesquelles nous avons pu tester les méthodes basées sur des calculs de densités et de distances.
Dans le cadre de la détection d'anomalies dans un cadre non supervisé, les méthodes que nous avons sélectionnées sont le Local Outlier Factor, que nous dénomerons LOF par la suite, ainsi que la One-Class SVM. Nous choisissons ces méthodes, pour leur applicabilité à notre problématique et leur point de vue différent (densité, distance).
Nous rajoutons à cette liste une nouvelle méthode que nous avons développée appelée

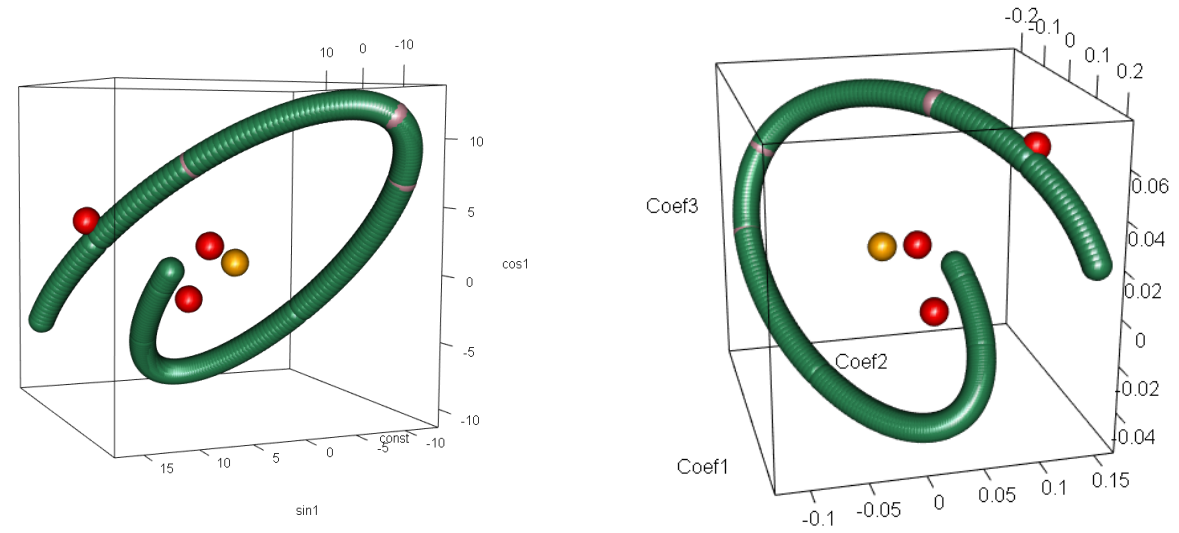

FIGURE 1.18 – Trois premiers niveaux de coefficients issus de projection d'une télémesure simulée sur une base de Fourier (à gauche) et une base à noyau (à droite) et les anomalies insérées : anomalies de motif en rouge, de périodicité en orange et anomalies locales en rose.

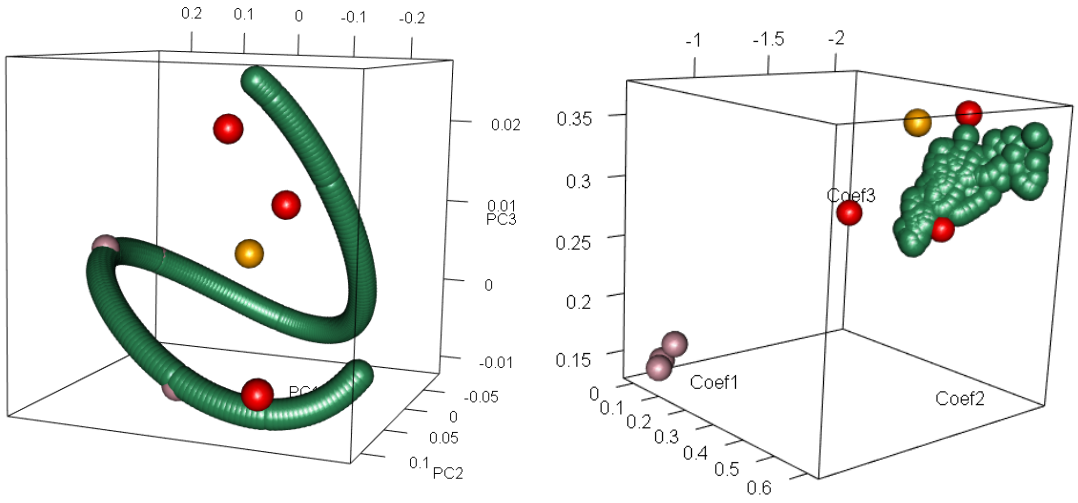

FIGURE 1.19 – Trois premiers niveaux de coefficients issus de la projection d'une télémesure simulée sur une ACP à noyaux (à gauche) et sur une base à noyaux appliquée au périodogramme des données (à droite) et les anomalies insérées : anomalies de motif en rouge, de périodicité en orange et anomalies locales en rose.

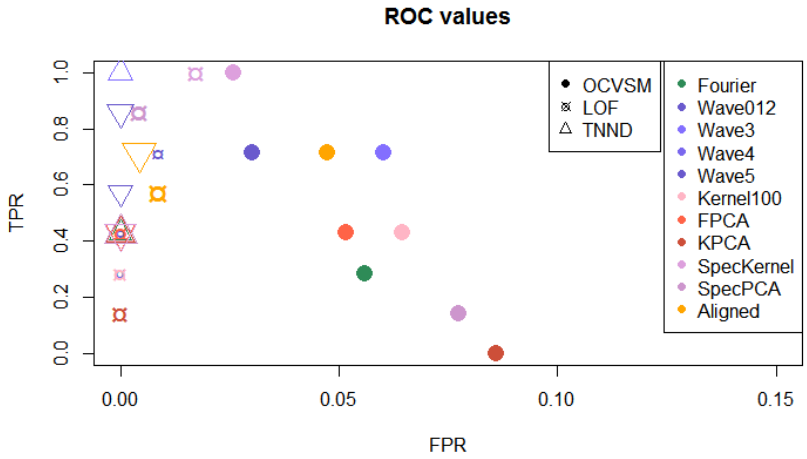

FIGURE 1.20 – Valeurs ROC pour chacune des méthodes et chacune des bases.

TNND pour Temporal Nearest Neighbours Dissimilarity. Nous avons comparé ces méthodes grâce au même exemple simulé que celui présenté dans le paragraphe précédent. Après avoir fixé les paramètres et seuils de chacune de ces méthodes, nous obtenons les résultats présentés en Figure 1.20 au travers d'une courbe ROC illustrant les performances pour chaque couple de méthodes et projections. Cette représentation nous montre que notre nouvelle méthode élaborée en tenant compte des spécificités de nos données est la plus performante, et nous confirme la pertinence des ondelettes et des coefficients construits sur le périodogramme des données pour la détection d'anomalies. Le LOF donne également de très bons résultats. En revanche, la One-Class SVM renvoie trop de faux positifs. Nous suspectons que le caractère évolutif des motifs dans le temps ne soit pas adapté aux méthodes se basant sur le calcul de densité. En effet, comme nous l'avons remarqué sur les figures 1.18 et 1.19, les premiers coefficients ne semblent pas être répartis selon une zone de forte densité, mais plutôt suivant une courbe retraçant la tendance annuelle des télémesures, excepté le cas particulier de l'exemple des coefficients construits à partir du périodogramme. Par ailleurs, cet exemple correspond à l'ensemble pour lequel la One-Class SVM a réalisé la meilleure performance en terme de détection d'anomalies. Les bons résultats du LOF et de la TNND se confirment également sur des données réelles.
Concernant les tests des équipements Radio-Fréquence, la One-Class SVM et le LOF peuvent tous être appliqués pour détecter les comportements atypiques dans les courbes produites. Le LOF semble néanmoins un peu plus adapté car la One-Class SVM a tendance à classer les données "trop parfaites" comme étant atypiques. Un exemple de résultats obtenus grâce à la One-Class SVM est représenté Figure 1.21. On y voit que les courbes

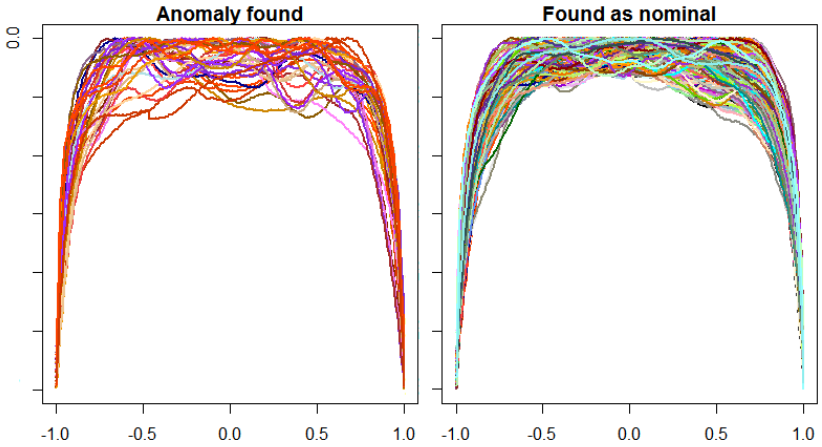

FIGURE 1.21 – Détection d'anomalies dans les courbes Gain-Fréquence grâce à la One-Class SVM appliquée aux coefficients de l'ACP - à gauche les anomalies détectées et à droites le reste des courbes.

comportant les plus grandes oscillations et pentes ont bien été classées comme étant atypiques, notamment celles correspondant à une perte du signal en deçà de la spécification sur la bande passante. Certaines des courbes que l'on peut qualifier de "trop parfaites" ont également été classées comme anormales. En revanche, on peut remarquer que certaine des courbes comprenant des oscillations significatives n'ont pas été classées comme atypiques par la One-Class SVM.

1.5.3 Identifier les coefficients pertinents pour la détection d'anomalies grâce à une procédure de tests multiples

Comme nous l'avons introduit en fin du paragraphe 1.5.1, ne retenir que les premiers coefficients issus d'une projection est usuel dans la littérature pour les applications de classification. En pratique, cette approche n'est pas nécessairement la plus pertinente dans le cas de détection d'anomalies.
En effet, les anomalies, par leur faible occurrence, ne représentent pas une forte proportion de la variance. Pour remédier à ce problème, nous avons créé une procédure de tests multiples dans un cadre semi-supervisé permettant de sélectionner automatiquement les niveaux de coefficients les plus intéressants pour la détection d'anomalies.
Cette procédure est fondée sur la comparaison de la distribution des coefficients dans deux sous échantillons, dont l'un d'eux est connu comme ne comportant pas d'anomalies. Pour cela, nous avons utilisé un test statistique développé à partir de la distance de Wasserstein que nous appliquons aux coefficients issus de la projection sur les ondelettes de

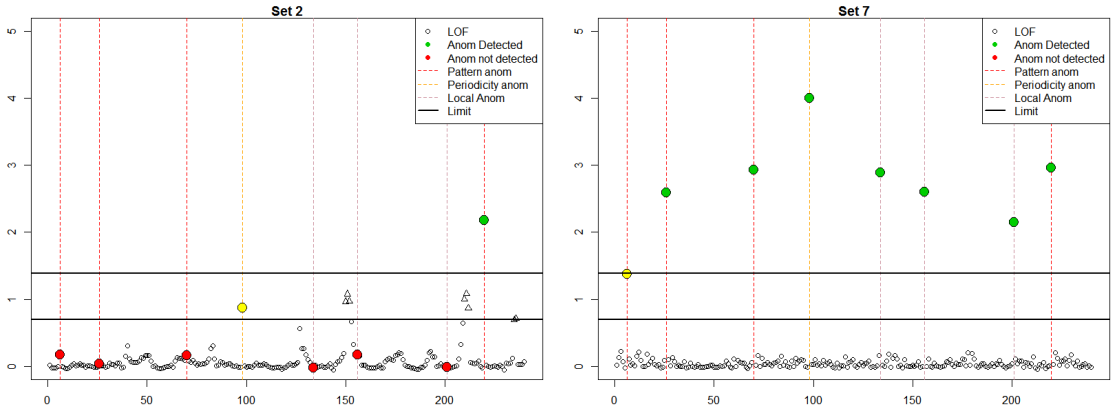

FIGURE 1.22 – Calcul du LOF dans l'échantillon comportant des anomalies. A gauche, sur les coefficients de l'ACP représentant 95% de la variance, à droite sur les coefficients de l'ACP sélectionnés par la procédure de tests multiples.

Haar et de l'ACP. Ces deux bases ont été choisies puisqu'elles sont orthonormées dans $\mathbb{R}^p$, afin de satisfaire la condition d'indépendance requise pour appliquer la procédure de tests multiples de Benjamini-Hochberg.
Pour l'ACP, par exemple, la sélection de coefficients nous donne des résultats très surprenants. Sur une télémesure simulée, nous ne sélectionnons que des composantes principales au delà de la 10ème. A titre d'exemple, les résultats sur le calcul du LOF sur une télémesure simulée sont représentés Figure 1.22, et nous illustrent le gain apporté par cette procédure pour la détection d'anomalies.
Pour une méthode de détection d'anomalies donnée, cette nouvelle approche de sélection des coefficients permet de diminuer les fausses alarmes et d'étendre la marge entre les anomalies et les autres données, par conséquent d'augmenter la puissance de la méthode. De plus, nous avons pu montrer par des simulations que même dans le cas où le sous-ensemble de référence comportait quelques anomalies, la sélection des coefficients permettait tout de même de bien détecter les anomalies, aussi bien celles immiscées dans l'ensemble supposé nominal que dans celui comportant des anomalies. De cette manière, on peut s'abstenir de labelliser une partie des données. Pour plus de détails, le lecteur pourra se référer à l'annexe 3.A.

1.5.4 Analyse multivariée

1.5.4.1 Tests sur les matrices de covariances

Nous nous sommes également intéressés dans cette thèse à l'aspect multivarié des données. En ce qui concerne les télémesures, nous avons cherché à détecter des divergences de comportement majeures au sein d'un groupe de plusieurs télémesures. Pour cela, nous

avons choisi de nous intéresser à l'analyse de la matrice de covariance des télémesures entre elles au cours du temps. Dans ce but, nous avons identifié trois tests statistiques existants pour tester l'égalité de deux matrices de covariances. L'hypothèse à tester est donc la suivante $H_0 : \{\Gamma_1 = \Gamma_2\}$, où Γ_1 et Γ_2 sont les matrices des covariances entre plusieurs télémesures, pour deux journées ou deux périodes orbitales consécutives, et estimées par les matrices de covariances empiriques $\hat{\Gamma}_1$ et $\hat{\Gamma}_2$.

- Le premier test, défini par Fremdt [30], consiste à projeter les deux journées de télémesures sur les vecteurs propres d'une matrice de covariance de référence, par exemple la moyenne ou barycentre des deux matrices de covariances empiriques $\hat{\Gamma}_1$ et $\hat{\Gamma}_2$. Les covariances de ces deux journées de télémesures $\hat{\Delta}_1$ et $\hat{\Delta}_2$ dans cette nouvelle base doivent, sous l'hypothèse nulle, s'approcher de la matrice des vecteurs propres de la matrice de référence, pour les deux périodes considérées. La statistique de test est construite à partir de la différence de ces deux matrices $(\hat{\Delta}_1 - \hat{\Delta}_2)$.

- Un test développé par Ilea [45] consiste à définir une statistique de test à partir de la somme des logarithmes des valeurs propres de la matrice $M = \hat{\Gamma}_2^{-1}\hat{\Gamma}_1$.

- Un test développé par Cai [16] est basé sur la plus grande différence entre les éléments de la matrice $\hat{\Gamma}_1 - \hat{\Gamma}_2$. Ce test est conçu pour être robuste dans des conditions sparses, c'est à dire si la matrice $\hat{\Gamma}_1 - \hat{\Gamma}_2$ contient en majeure partie des valeurs nulles.

- Enfin, nous avons développé, en nous inspirant des deux premiers tests, un nouveau test statistique pour tester l'égalité des matrices de covariances. Ce test est construit à partir du logarithme du rapport entre les termes diagonaux des matrices $\hat{\Delta}_1$ et $\hat{\Delta}_2$, telles qu'elles ont été définies dans le test de Fremdt. Voir le chapitre 4 pour de plus amples informations.

Afin de déterminer le test le plus adapté à notre problématique et à nos données, nous avons comparé ces quatre tests statistiques, tout d'abord sur une simulation construite à partir de vecteurs Gaussiens, puis sur des télémesures simulées. Nous avons pu comprendre que le test défini par Ilea ne s'appliquait pas à nos données car, la plupart du temps, les matrices de covariances sont mal conditionnées puisqu'en pratique, certaines télémesures sont presque identiques, et donc très corrélées. Les tests de Fremdt et de Cai renvoient un nombre non négligeable de fausses alarmes. De plus, le test de Fremdt est très sensible au nombre de vecteurs propres que l'on considère. Notre nouveau test, bien qu'il renvoie des résultats similaires au test de Ilea sur des vecteurs Gaussiens, semble être le plus adapté à notre problématique et à nos données. Sur un exemple de télémesures simulées, nous avons illustré la puissance de ces tests grâce à une courbe ROC, représentée Figure 1.23. Nous pouvons constater que lorsqu'on augmente le nombre de composantes principales considérées d, notre test est capable de détecter de plus en plus d'anomalies, dont anomalies locales plus difficiles à discerner. Les autres tests renvoient de nombreuses

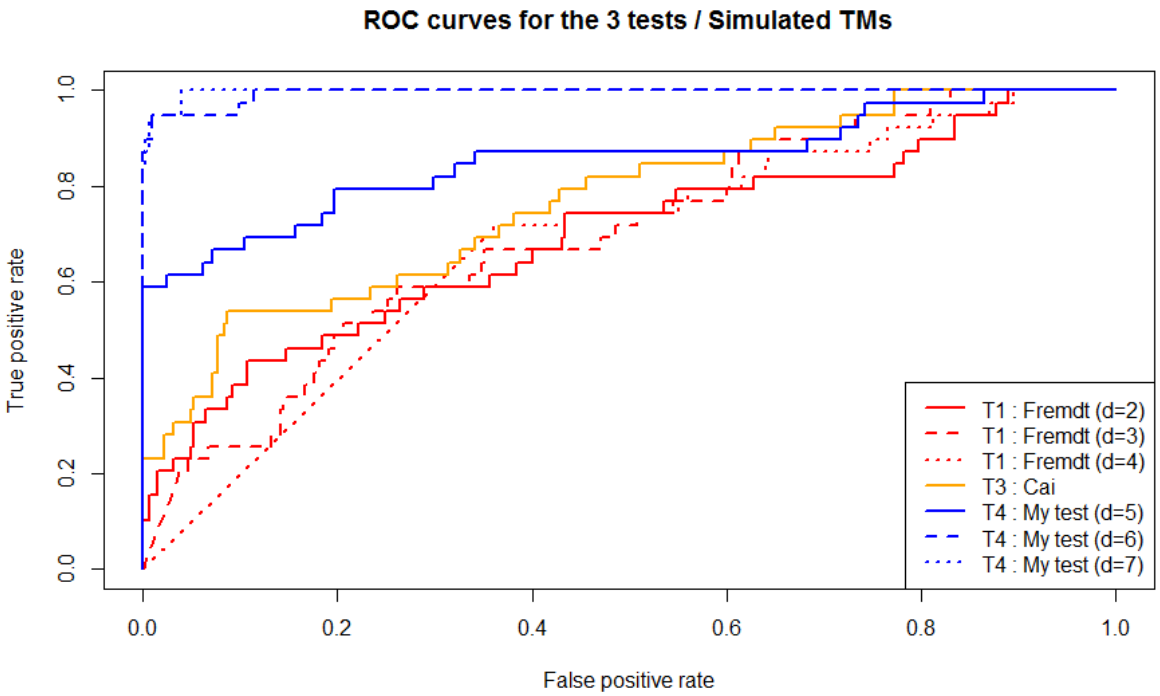

FIGURE 1.23 – Comparaison des trois tests sur des télémesures simulées comportant des anomalies non liées entre elles. Le nouveau test est le plus performant.

fausses alarmes.
Sur de vraies télémesures, l'algorithme réussit à mettre en lumière les changements les plus importants dans la structure de covariance, coïncidant avec des événements connus à bord du satellite.

1.5.4.2 Tests de l'égalité des distributions entre n courbes

Lorsque les courbes que l'on considère sont générées par le même mécanisme, l'approche consistant à comparer les matrices de covariances n'est pas la seule approche possible.
Dans notre cas, on s'intéresse à plusieurs réalisations du test Gain-Fréquence sur une chaine d'amplification donnée, dans plusieurs conditions de températures et de pression. On souhaite montrer que l'environnement extérieur n'impacte pas la performance de la chaîne d'amplification, et donc les courbes considérées. Pour cela, nous utilisons une généralisation d'un test défini par Fromont et. al [31] pour tester l'égalité des distributions des courbes Gain-Fréquence correspondant à la même chaîne d'amplification. La loi de la statistique sous l'hypothèse nulle est estimée grâce à des permutations.
En pratique, l'environnement a toujours une influence sur les courbes, donc le test est

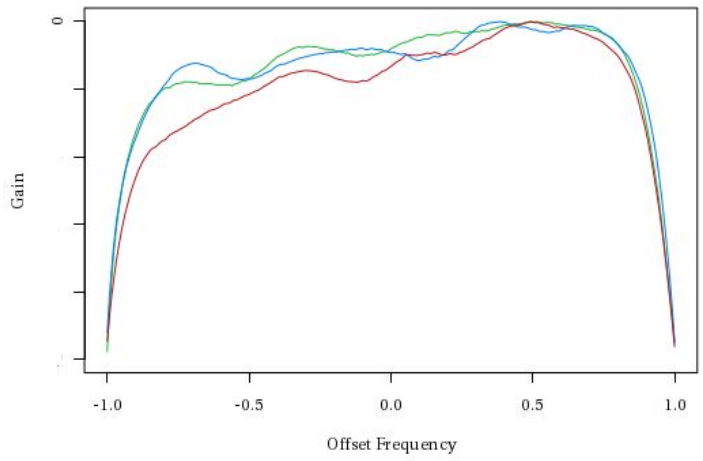

FIGURE 1.24 – Test Gain-Fréquence effectué dans trois environnements différents correspondant à la plus grande statistique de test.

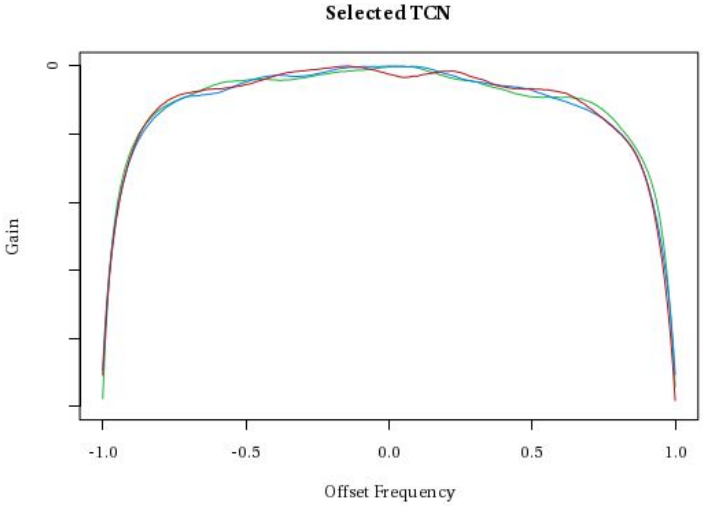

FIGURE 1.25 – Test Gain-Fréquence effectué dans trois environnements différents correspondant à la plus petite statistique de test.

rejeté systématiquement. En revanche, la valeur de la statistique de test définie peut être vue comme une mesure de distance entre les courbes. Ainsi, les chaines d'amplification peuvent être classées de manière à mettre en avant celles qui ont le plus changé à travers les différentes phases. Nous représentons Figures 1.24 et 1.25 les courbes issues des chaines d'amplifications correspondant à la plus grande statistique de test, ainsi qu'à la plus petite.
Cette approche permet d'aider l'ingénieur chargé de la validation du satellite à anticiper d'éventuelles détériorations du signal en fonction de l'environnement dans lequel se trouve

le satellite. Ici, la figure 1.24 nous montre une évolution de comportement inattendue.

1.5.5 Mise en oeuvre opérationnelle des méthodes aux ingénieurs à Airbus

En plus des résultats théoriques et applicatifs que nous présentons dans cette thèse, deux démonstrateurs ont été développés sous forme d'applications web grâce au package *Shiny* de *R*. Ces applications sont disponibles sur tout le site interne Airbus Defence and Space à Toulouse. La première application est purement dédiée à l'analyse des tests Gain-Fréquence, et tend à se généraliser à tous les tests.
Cette application a été développée pour que l'ingénieur chargé de la validation des satellites puisse réaliser la classification des tests obtenus après une ou plusieurs phases de test. Les anomalies sont classées selon leur caractère local ou global.
Quelques paramètres sont laissés libres à l'ingénieur tels que le pourcentage d'anomalies, la base de projection. L'ingénieur peut également revenir sur les sorties de l'algorithme en indiquant les fausses détections. L'outil les classe alors comme nominales. Une fois l'ingénieur satisfait de sa classification, il peut télécharger les résultats sous forme de tableur Excel. Une copie de ces résultats est conservée sur le serveur. Nous conservons cet historique pour se laisser la possibilité à l'avenir d'apprendre cette classification, pour éventuellement réaliser la détection des courbes aberrantes de manière semi-supervisée, voire supervisée. Cela peut nous permettre par exemple d'étendre la sélection des coefficients définie chapitre 3 aux données de test.
La figure 1.26 nous montre l'onglet de l'application dédié à la comparaison de plusieurs phases de test, ici deux phases COLDVAC et HOTVAC. Une visualisation des données en deux dimensions est proposée à l'utilisateur.
La deuxième application est dédiée à l'analyse multivariée et univariée des télémesures. Il s'agit dans un premier temps de charger un jeu de données récupéré dans TELMA, l'outil interne pour visualiser et contrôler les télémesures, et de réaliser l'analyse des matrices de covariance. La figure 1.27 illustre une des fonctionnalités de l'application, la comparaison de deux journées de télémesures grâce aux matrices de covariance. Outre la comparaison de deux journées isolées, l'application effectue la comparaison des covariances entre tous les couples de journées consécutives, afin de mettre en avant les journées où des événements se sont éventuellement produits à bord du satellite.
La détection est également disponible de manière univariée, grâce aux méthodes définies précédemment.
Ces applications sont aujourd'hui utilisées dans le cadre de la validation des tests des satellites de télécommunications, et dans le cadre de l'investigation d'anomalies pour les satellites en vol.
Ce travail est annexe au travail de recherche de cette thèse, néanmoins il reste important pour rendre les travaux disponibles aux ingénieurs Airbus. De plus, la vulgarisation et

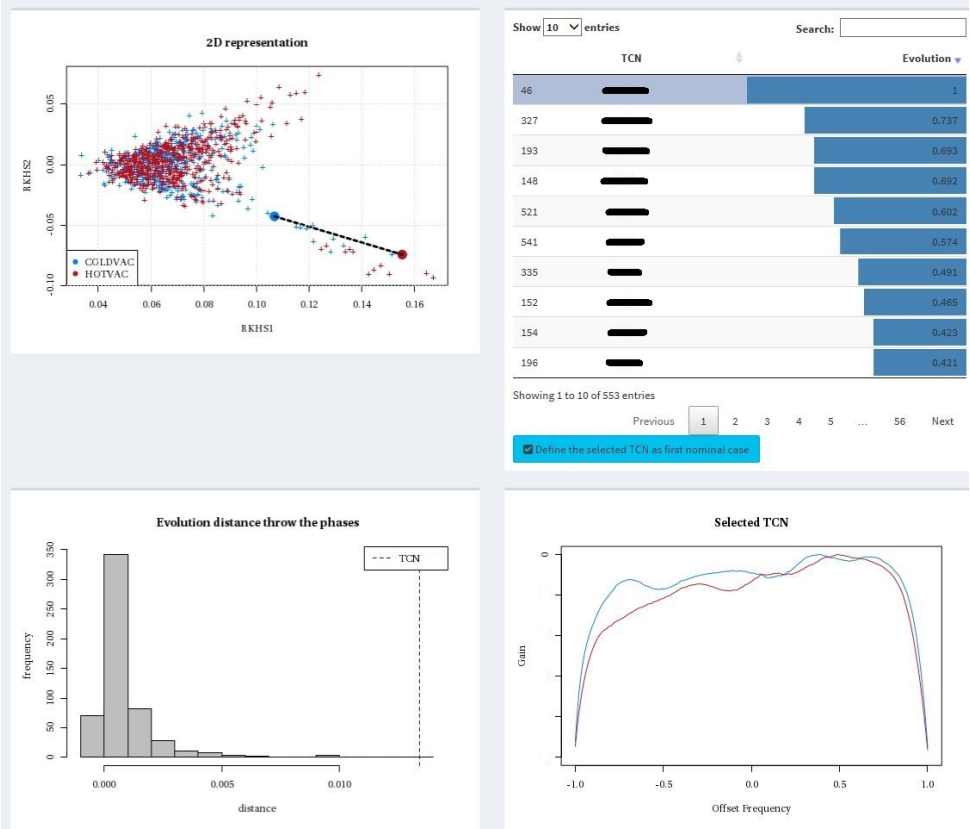

FIGURE 1.26 – Application STAR pour l'analyse des courbes Gain-Fréquence - onglet dédié à l'analyse multivariée.

l'interprétabilité sont des aspects importants pour les chercheurs en mathématiques appliquées, notamment pour rendre leurs travaux utilisables et paramétrables à un public non statisticien.

FIGURE 1.27 – Application STAR pour l'analyse des télémesures - onglet pour visualiser et comparer la matrice de covariance de deux jours de télémesures.

1.6 Organisation du manuscrit

Dans cette introduction, nous avons présenté en détails la problématique de la thèse et plus brièvement la démarche abordée. Dans chacun des prochains chapitres, nous présentons chaque étape de notre étude plus en détails.
Chaque chapitre de cette thèse est rédigé sous forme d'article. Ainsi, les chapitres sont indépendants et peuvent être lus séparément. De la même manière, certaines notations et certaines méthodes sont susceptibles d'être introduits plusieurs fois.
Le prochain chapitre est dédié à la revue des méthodes pour la détection d'anomalies dans les données spatiales, à partir de plusieurs ensembles de coefficients construits grâce à des projections dans des bases orthonormées. Cet article sera prochainement publié dans la revue AIAA - *The American Institute of Aeronautics and Astronautics*, à l'issue de la conférence internationale SpaceOps qui se déroulera à Marseille en mai 2018.
Le deuxième article introduit la procédure de tests multiples destinée à optimiser la sélection des coefficients pour la détection d'anomalies. Cet article a été soumis à la revue *Transactions on Big Data* en janvier 2018. Cet article est d'ores-et-déjà consultable librement via la plateforme HAL et Arxiv.
Le dernier article est dédié aux tests sur les matrices de covariances pour la détection d'anomalies multivariée dans les télémesures. Il a été en partie présenté lors de la conférence *Big Data From Space 2017*, organisée par l'ESA à Toulouse en novembre 2017 et fait l'objet d'un article en préparation, qui devrait être proposé à la publication au cours de l'année 2018.

Chapter 2

Statistical Methods for Outlier detection in Space Telemetries

Comme nous l'avons vu dans l'introduction, la détection d'anomalies dans les données fonctionnelles peut être effectuée de plusieurs manières, selon plusieurs angles d'approche. Les anomalies que nous sommes susceptibles d'observer peuvent également être de plusieurs natures. Il est donc important d'inaugurer ce chapitre en introduisant les anomalies rencontrées, et de réaliser une revue des méthodes que nous pouvons utiliser, afin de définir les meilleures options pour mettre en évidence certaines pathologies.
Les données étant fonctionnelles, nous prenons le parti de construire des indicateurs (ou features) issus de projections dans des bases de fonctions orthonormées. Nous allons donc tester plusieurs bases différentes, telles que des bases fixes ou construites sur les données, qu'elles soient appliquées aux données brutes ou à une version pré-traitée de ces données. Nous appliquons sur ces features également plusieurs méthodes de détection d'anomalies. Nous proposons également une méthode originale, que nous avons définie en tenant compte de la saisonnalité des télémesures pour détecter des anomalies dans des séries temporelles. Les résultats de cette revue nous permettront de choisir les méthodes et les bases de fonctions les plus prometteuses par la suite.
Ce papier sera publié dans la revue AIAA (The American Institute of Aeronautics and Astronautics), qui est une revue de référence dans le domaine du spatial. Ce papier sera en particulier publié dans le cadre de la conférence internationale Space Ops qui se déroulera à Marseille en mai 2018. Une première partie de cette étude dédiée à l'analyse des données de tests a d'ores-et-déjà été présentée dans le cadre des *Journées de la Statistiques* de la Société Française de Statistiques en juin 2016.

Statistical Methods for Outlier detection in Space Telemetries

Clémentine Barreyre[1], Béatrice Laurent[2], Jean-Michel Loubes[3], Bertrand Cabon[1], Loïc Boussouf[1]

Abstract

Satellites monitoring is an important task to prevent the failure of satellites. For this purpose, a large number of time series are analyzed in order to detect anomalies. In this paper, we provide a review of such analysis focusing on methods that rely on features extraction. In particular, we set up features based on fixed functional bases (Fourier, wavelets, kernel bases...) and data-based bases (PCA, KPCA). The outlier detection methods we apply on those features can be distance or density-based. Those algorithms will be tested on real telemetry data.

2.1 Introduction

Analysing functional data and outlier detection have become an increasingly important subject over the years, and the space industry is facing more and more these issues as well. Indeed, unlike other complex systems, the hardware monitoring of satellites is impossible but a single failure in a component or a subsystem may be fatal to it. Consequently, it is highly important to check the behavior of the satellite during all its lifecycle to detect the divergences as soon as possible.
The behavior of the satellites is firstly checked during the tests on the ground, and in flight via the measurements of thousands of health parameters, called telemetries. Most of such signals can be considered as functions depending on time and sampled at high rate, leading to a large number of observed values. In order to automate the monitoring of all these data, existing processing were developed by experts based on physical knownledge. They are run frequently to raise any misbehavior but they rely on a posterior human

[1]Airbus Defence and Space, Z.I. Palays, 31 rue des Cosmonautes, 31400 Toulouse, France
[2]Institut de Mathématiques de Toulouse (UMR 5219), INSA Toulouse, Université de Toulouse, 135 avenue de Rangueil, 31400 Toulouse, France
[3]Institut de Mathématiques de Toulouse (UMR 5219), Université Paul Sabatier, Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse Cedex 9, France

analysis. However, because of the increasing complexity of the concerned systems and the increasing number of data, designing custom-made processing for all data is currently impossible, hence developing automatic data analysis has become essential.
The automatic monitoring of satellites data has already been adressed by the ESA [57] (European Space Agency), the CNES [32] (Centre National d'Etudes Spatiales - French Space agency) and the JAXA (Japanese space agency) to detect real-time anomalies in telemetries. In the aeronautic field, many health monitoring studies have been done, such as the studies performed by Thomas [82], Rabenoro [65] and Abdel-Sayed [1]. All these studies address the anomaly detection with different approaches in various frameworks, firstly to reduce the dimension of the data to concentrate the information, and then to isolate the anomalies.
Given the variety of all the outlier detection methods that have been studied in similar applications, we choose to confront some of them to our data. For this reason, in this paper, we tackle the issue of outlier detection by reviewing statistical methods to analyze typical functional data deriving from satellites that may present some anomaly in unsupervised settings.
Due to the complexity of defining what an outlier is, we aim at extracting main behaviors and identify the observations that differ from it. Hence we want to build features that characterize the normal behavior.
For this purpose, we consider dimension reduction techniques to overcome the high dimensional aspect of the problem. For instance we refer to various clustering applications such as Antoniadis [4], Tarpey [81] and Auder [6], and default detection applications such as Pan [62], where the dimension reduction is done on coefficients arising from projections into functional bases, mostly deriving from wavelet decompositions. Features can be extracted from such reduction techniques, and they are robust when there are some irregularities in the data, such as noise, irregular sampling, missing data.
The features arising from these projections are expected to concentrate the information on a small number of coefficients. As for the previously cited studies, several bases will be tested to build our features.
Then, we apply, on each feature set, several reference outlier detection techniques, for which we refer for example to [19], that we compare with a new method that takes into account the temporal order of the curves, called the Temporal Nearest Neighbours Dissimilarity.
Those methods are applied on simulated data to ease the comparison of all the projection bases and outlier detection methods directly, on a common example. Then, the methods and the results of the common comparison are validated on real telemetries as well as radio-frequency equipments test data. The first part of the paper will be dedicated to the data description, the second part to the dimension reduction and the third part on the outlier detection techniques. The methods will be applied to real telemetry data in the last section of this paper.

2.2 Data

2.2.1 Description

The data that is analyzed is telemetry data. A telemetry is an observation of a continuous signal $t \mapsto X(t)$ sampled at some instants t_j, with $j = 1, ..., T$. Such signals are numerous and each signal characterizes a component/parameter of the satellite. The telemetries are very numerous and a satellite may have as many as 10 000 telemetries or more. The sampling time can be very short, and will vary from one telemetry to another, with most telemetries being sampled every 30 seconds. It means that a full year of telemetry is represented by more than a million instants.
In order to build novelty detection method in an unsupervised setting, we must build confidence areas where the good behavior will be mainly located while anomalies will be away of this area. Hence we need to extract some representative features of these curves that exhibit a pattern that will represent the normal behavior pattern. The telemetries can be very different and vary a lot depending whether the satellite is a telecommunication or an observation satellite.
Telecommunication satellites are mostly located on a geostationary orbit. They always stay above the same location in Earth, completing an orbit each day. Consequently, most of the telemetries coming from telecommunication satellites exhibit patterns that are replicated on a daily basis. These patterns can evolve through the year because of the seasonality. Also, geostationary satellites experience eclipses twice a year during a period of approximatively 15 days. The behavior of some telemetries can be significantly impacted by these events.
Observation satellites are Low Earth Orbit (LEO). Unlike the telecommunication satellites, their location above Earth changes all the time, and it takes a few days for the satellite to scan the whole planet. They often have an inner periodicity which is shorter than a day, with no annual evolution.

In this purpose we present in the following such typical examples and the different kinds of anomalies that may occur more frequently.

2.2.2 Main kinds of signals

The following examples are real satellites data. The methods we develop will be tested later on those telemetries in the section 2.5.

2.2.2.1 Non evolutive telemetry

This first telemetry, represented in the figure 2.1 is not yearly periodical, and exhibits changing behavior. This telemetry contains a large amount of anomalies, hence it is an interesting usecase for applying outlier detection algorithms.

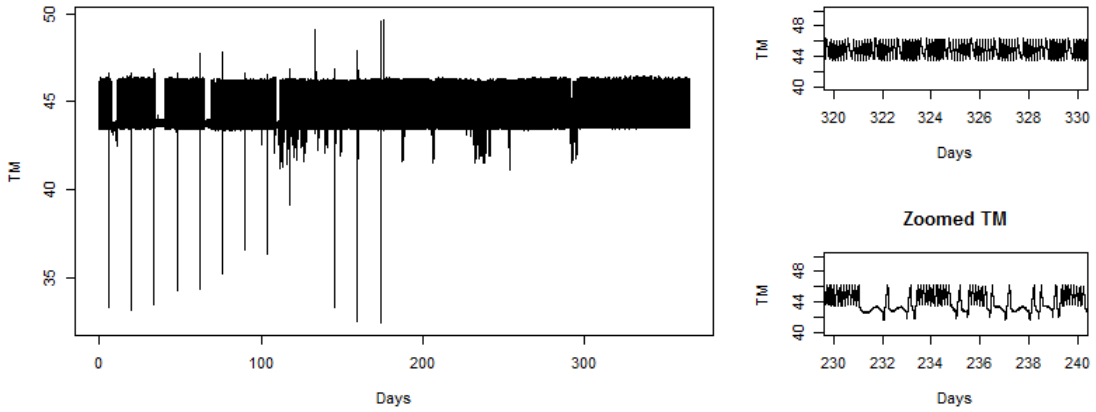

Figure 2.1 – First telemetry.

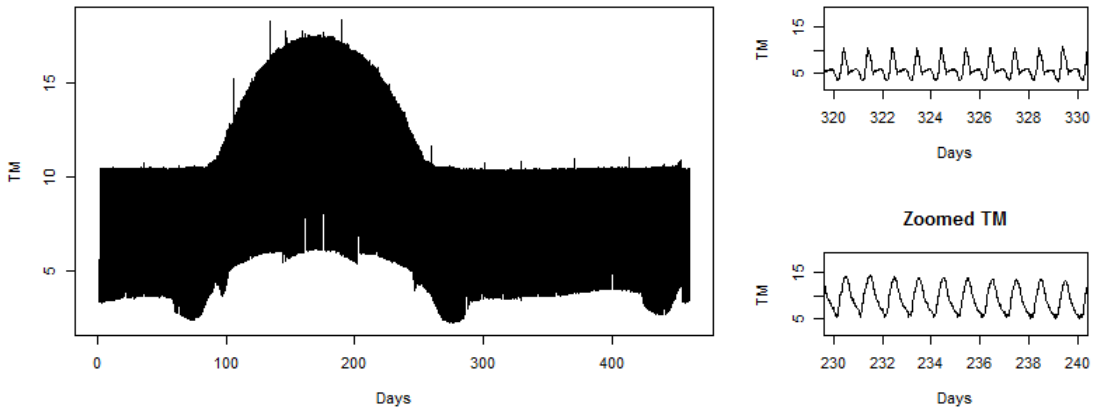

Figure 2.2 – Second telemetry.

2.2.2.2 Evolutive telemetry

This second telemetry, represented in the figure 2.2 has a yearly trend. Unlike the previous example, there is no obvious anomalies in this dataset. The outlier detection here will help us to highlight the changes that can be found in this telemetry.

2.2.2.3 Gain vs Frequency

For this third example, we have proposed the Gain vs Frequency test, which is a test commonly used to characterize the performance of telecommunications channels. The principle

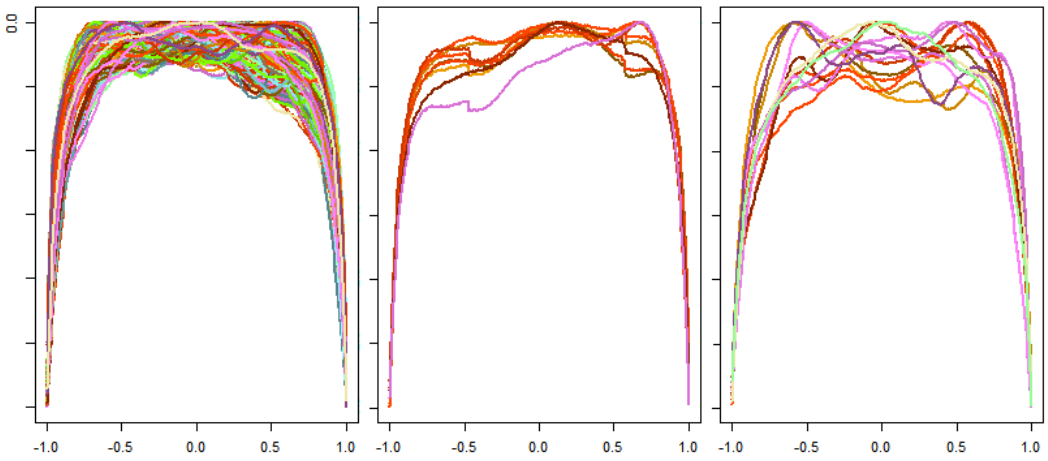

Figure 2.3 – Example of Gain-Frequency tests (left plot), small steps as local anomalies (middle plot) and ripples as global anomalies (right plot).

of the test is to measure the gain between the input and the output at all frequencies within the channel bandwidth, and is an excellent mean to verify the impact of the passive units and the active components on the expected performance. The ideal response of the Gain-Frequency tests is a constant gain function over the frequency bandwidth. The goal is to find the most altered results among the curves resulting from the Gain-Frequency. The signal can be altered locally (steps, spikes...) or globally. The figure 2.3 illustrates these tests and shows how their anomalies look like.

2.2.3 Anomaly description

By definition, an anomaly is an event that differ from the usual behavior of the data. As it is unusual, it is infrequent and therefore difficult to detect. Indeed, defining a normal region that encompasses every possible normal behaviors is very difficult. In addition, the boundary between normal and anomalous behavior is often not precise. Hence our task is very difficult, hence we will focus on specific types of anomalies. As a matter of fact, space telemetries' anomalies can be classified into three categories.

- Local anomalies: those anomalies occur in a limited time durations. They are characterized by discontinuities of the time series at some breaking points.
- Pattern anomalies: one of the pattern of the telemetry does not exhibit the expected shape.
- Periodicity anomaly: the frequency, or the phasis of the physical phenomenon changes.

We can see that these anomalies can have several different origins : anomaly on duration, anomaly on frequency. The anomalies can be due to real phenomenon in flight, but they can also be the consequence of measurement errors. Thus a single method is unlikely to highlight all those types of anomalies with a small rate of false positive. As we do not know the past anomalies on the telemetries, we use a simulated example, close to our data, on which we insert several types of anomalies. It will enables us to evaluate and compare the performances of our methods on a common application. In a second step, these methods are validated on real telemetries.

2.2.4 Simulated example

In order to compare the different methodologies we apply in this paper, we have created a simulated example to ease the validation of each method. This telemetry has a daily and a yearly periodicity, to carry the difficulties we may find in real telemetries.
We used a combination of periodic signals and given patterns to generate this telemetry. We simulated 240 days of telemetry in order to get a significant number of individuals after splitting the signal into days. Several types of anomalies have been introduced. The shape of this telemetry is represented in the figure 2.4. In this figure, the evolution of the repetitive patterns through the days is easily noticeable. We can see that from one day to another, the patterns do not evolve much. We simulated three local anomalies which are represented on the figure 2.5. The first two are local spikes on the signal. The third anomaly corresponds to the increase in the amplitude of the noise during a small time duration. We also added pattern and periodicity anomalies, represented on the figure 2.6. The anomalies can be minor, such as amplitude attenuation, or stronger such as having two patterns within a given day instead of one. Consequently, we have 7 anomalies in total. These anomalies have been chosen in order to investigate if both the obvious ones and the minor ones can be raised by our algorithms. In the next section, we firstly define how to extract the features for highlighting the anomalies.

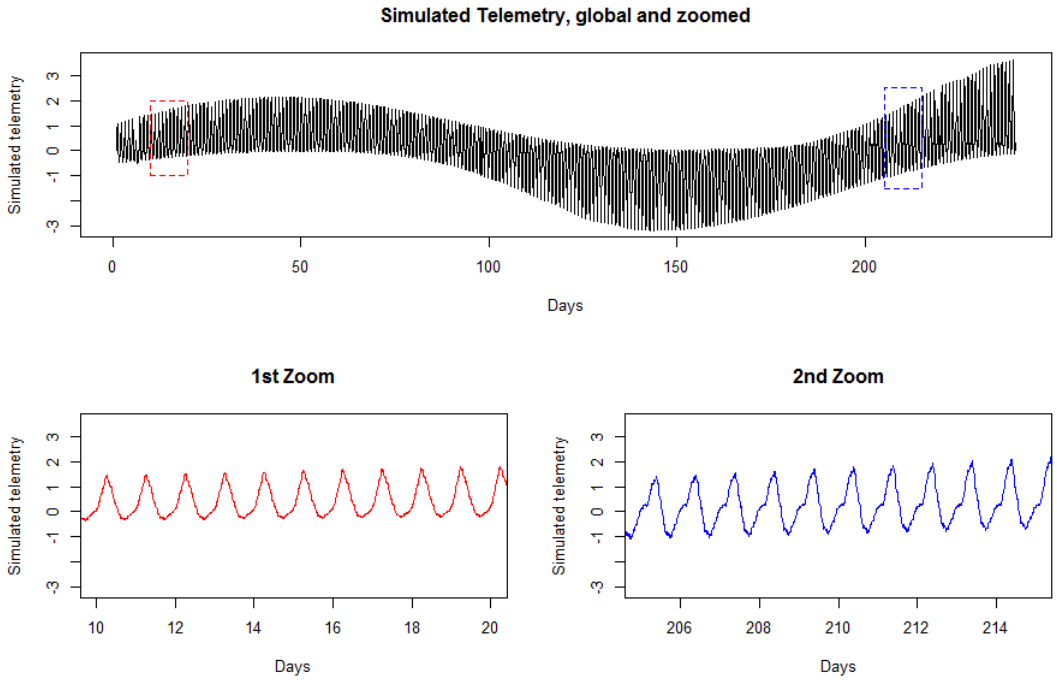

Figure 2.4 – Simulated telemetry and its zoomed version.

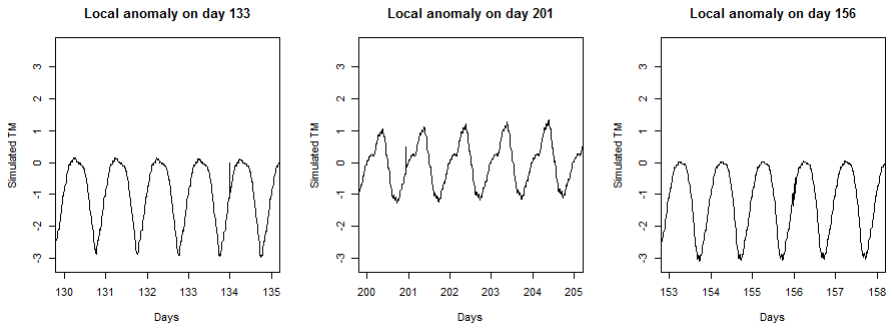

Figure 2.5 – Simulated local anomalies.

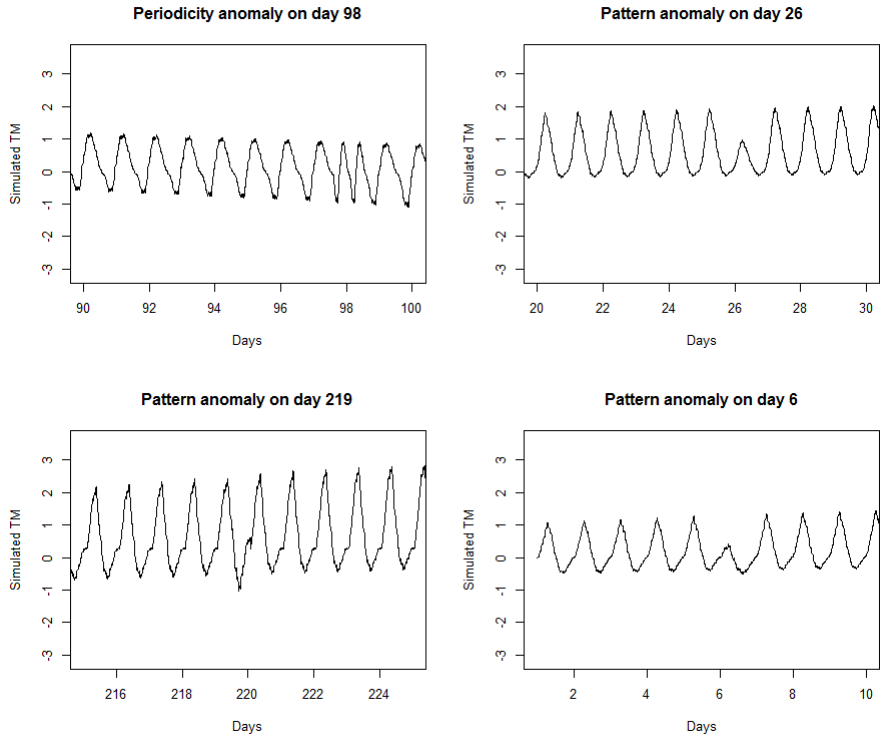

Figure 2.6 – Simulated Pattern anomalies.

2.3 Anomaly detection with feature extraction

2.3.1 Methodology

Our aim is to extract appropriate features that convey some relevant information to characterize anomalies as functions that deviate from these main features. In fact, we recall that the data we monitor are functional data, consequently the features will be built on projection onto functional bases.
Let $x(t) \in \mathbb{R}$, with $t = 1, ..., T$ be a univariate telemetry time series. At first, the time series are cut into n intervals of length p, with $n \times p \leq T$ and $(n+1) \times p > T$. For geostationnary satellites, the telemetries can be cut into days. For observation satellites, several time durations, such as the orbit period, can be tested.
In a first step, have to analyze the matrix $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_n]^T$. The dimension of the matrix $\mathbf{X}$ is $n \times p$. The idea is to reduce the dimension of the vectors $\mathbf{X}_i, i = 1, ..., n$ by resuming the information contained in this vectors by $d < p$ features $(\theta_{i,1}, \ldots, \theta_{i,d})$ chosen properly. It is important to recall that those features are built in order to highlight the best the anomalies. We use the following methodology.
After splitting the data into regular intervals, the vector $\mathbf{X}_i$ to analyze can be seen as observations of functional data. This corresponds to the observation model

$$X_{i,j} = X_i(t_j) + \varepsilon_{i,j}, \ i = 1...n, \ j = 1, ..., p,$$

where $\varepsilon_{i,j}$ is a random noise which variance σ^2 is unknown and we suppose that $X_i \in \mathbb{L}^2([0,1])$. Without loss of generality, we assume that for all $j, t_j \in [0,1]$. Moreover, if the telemetries are regularly sampled, we can assume that $t_j = j/p$.
There exist many ways to represent functional data in a reduced dimension. Some bases are more adapted to some types of functions. We can consider for instance Fourier bases, wavelet bases... Auder and Fischer [6] introduced some well-known functional bases in order to apply functional clustering. Our problem is to identify several types of anomalies. The functional features have to carry relevant information on the anomalies, such that any outlier detection algorithm has a chance to detect them.
We proceed by identifying some features that are able to accentuate anomalies in a reduced dimension. Then, the feature selection is made in order to maximize the variance explained by those features.

2.3.2 Projection onto functional bases

Assume that for all $i = 1...n$, $X_i \in \mathbb{L}^2([0,1])$. Those functions can thus be represented in an orthonormal functional basis in $\mathbb{L}^2([0,1])$. (See Ramsay [68] for more details). Then, if $(\phi_\lambda)_{\lambda \geq 1}$ is an orthonormal basis in $\mathbb{L}^2([0,1])$,

$$X_i(t) = \sum_{\lambda=1}^{\infty} \theta_{i,\lambda} \phi_\lambda(t), \text{ with } \theta_{i,\lambda} = \langle X_i, \phi_\lambda \rangle \tag{2.1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbb{L}^2([0,1])$. In practice, the coefficients $\theta_{i,\lambda}$, $\lambda = 1, ..., d$ are estimated by their empirical counterparts.

$$\hat{\theta}_{i,\lambda} = \frac{1}{p} \sum_{j=1}^{p} X_{i,j} \phi_\lambda(t_j). \tag{2.2}$$

2.3.2.1 Fixed basis

The Fourier basis When the data is periodical, analysing the frequency properties of the signal is widely applied in signal processing. We know that we can decompose the functions X_i according to

$$X_i(t) = \sum_{\lambda=0}^{+\infty} \theta_{i,\lambda} \phi_\lambda(t)$$

where $t \in [0,1]$. In this case, the functional basis $(\phi_\lambda)_{\lambda \geq 0}$ is the usual Fourier basis. We have, for $k \in \mathbb{N}^*$,

$$\begin{cases} \phi_0(t) = 1 \\ \phi_{2k}(t) = \sqrt{2} cos(2\pi k t) \\ \phi_{2k+1}(t) = \sqrt{2} sin(2\pi k t). \end{cases} \tag{2.3}$$

We now have to select d features properly. Let $S_d(X_i) = \sum_{\lambda=0}^{d} \theta_{i,\lambda} \phi_\lambda(t)$ be the partial Fourier series, and let $\boldsymbol{\theta}_{\cdot,\lambda} = (\theta_{0,\lambda}, \ldots, \theta_{n,\lambda})$ be the coefficients of all the curves for the λ^{th} feature. In practice, the vector $\boldsymbol{\theta}_{\cdot,\lambda}$ is unknown and is estimated by $\hat{\boldsymbol{\theta}}_{\cdot,\lambda}$ as defined in (2.2). We know that the more d increases, the better the approximation becomes. Based on this assumption, we choose the smallest d that satisfies the condition

$$\frac{\sum_{\lambda=0}^{d} Var(\hat{\boldsymbol{\theta}}_{\cdot,\lambda})}{\sum_{\lambda=0}^{p} Var(\hat{\boldsymbol{\theta}}_{\cdot,\lambda})} \geq 0.99.$$

It means that the d first features represent more than 99% of the variance of all the features we can compute. At last, we get $\hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{i,0}, ..., \hat{\theta}_{i,d})$ as remaining data to analyze, for $i = 1, ..., n$.

Example 1. *We applied the Fourier decomposition on the simulated example introduced Section 2.2.4. The five first coefficients deriving from this projection represent already more than 99% of the variance. To illustrate the kind of anomalies that could be highlighted by the Fourier basis, we have represented the three first components in the figure 2.7. The nominal days are represented as green dots in green, the local anomalies in pink, the periodicity anomaly in orange and the pattern anomalies are in red. As a first remark, we can to notice that those features do not appear as an agglomerate of points but as a continuous line, resuming the yearly trend of this telemetry. As we can see, the local anomalies seem unlikely to be raised through the Fourier basis whereas the other anomalies*

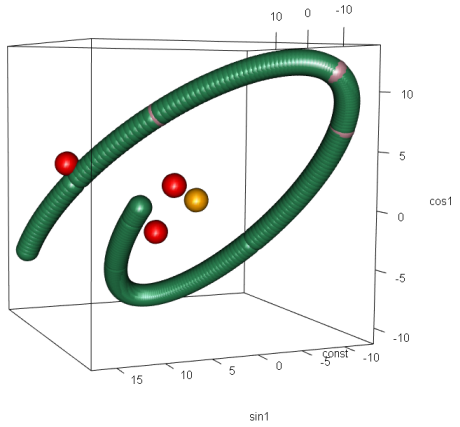

Figure 2.7 – Coefficients in the three first functions of the Fourier basis.

appear as outliers. The local anomalies in pink start slowly to stall the nominal curves in the following features. The anomaly detection will help us to confirm this hypothesis.

Wavelet basis A wavelet basis is defined from a pair of functions (ϕ, ψ) respectively called father and mother wavelet. We consider only compacted supported wavelets. We denote by $supp(\psi)$ the support of ψ. For all $j \geq 0$, $k \in \mathbb{Z}$, let $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ and for all $j \geq 0$, let

$$\Lambda(j) = \{k \in \mathbb{Z}; [-k, 2^j - k] \cap supp(\psi) \neq 0\}.$$

The functions $(\phi, \psi_{j,k}, j \geq 0, k \in \Lambda(j))$ form an orthonormal basis of $\mathbb{L}^2([0,1])$.
The easiest example to present is the Haar basis where $\phi = \mathbb{1}_{[0,1]}$; $\psi = \mathbb{1}_{[0,1/2[} - \mathbb{1}_{[1/2,1[}$.
In this case, $\Lambda(j) = \{0, 1, \ldots, 2^j - 1\}$ and $|\Lambda(j)| = 2^j$, for all $j \geq 0$.
The Haar wavelets are not continuous, and hence sometimes inappropriate for practical purposes. Many other wavelet bases can be considered such as Daubechies symmlet bases (see Daubechies [21] for more details).
A function $X \in \mathbb{L}^2([0,1])$ can be represented by its expansion onto a wavelet basis

$$X(t) = \alpha\phi(t) + \sum_{j \geq 0} \sum_{k \in \Lambda(j)} \beta_{j,k}\psi_{j,k}(t).$$

We extract, from the estimated wavelet coefficients $(\hat{\alpha}_i, \hat{\beta}_{i,jk}, j \geq 0, k \in \Lambda(j))$ of our functions X_i, a small number of suitably chosen coefficients. In the wavelet case, the coefficients are indexed by the level j and the position k. Thus, it does not have a sense to take the d first coefficients since they are not ordered in one dimension. The variance analysis is likely to encourage us to keep only the coefficients corresponding to the first levels.

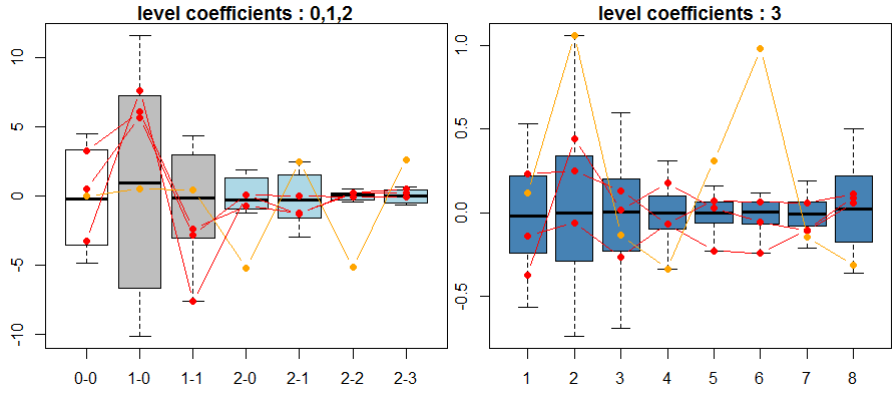

Figure 2.8 – Wavelet coefficients with $j \leq 4$ and details for pattern anomalies (in red) and periodicity anomaly (in orange).

However, we know that the first levels of wavelet coefficients are not designed to catch local events. Consequently, those events are unlikely to be highlighted in these levels. Thus, we have to consider larger levels as well. Coifman et al. [20] proposed to pick only a finite number of wavelets by keeping d pairs of coefficients $(j_1, k_1), ..., (j_d, k_d)$ such that $supp(\cup_{m=1}^{d} \psi_{j_m,k_m}) = [0, 1]$. Auder et al. [6] noticed that it is possible to keep all the position coefficients from a same level. In order to get the anomaly types that can be detected by each wavelet level, we will analyze them separately. As the first levels contain a small number of coefficients, we group the levels for which $j \in \{0, 1, 2\}$ together. The features of the following levels will be taken entirely. Finally, the features extraction we get can be

$$\hat{\boldsymbol{\theta}}_i^{012} = (\hat{\alpha}, \hat{\beta}_{0,0}, \hat{\beta}_{1,0}, \hat{\beta}_{1,1}, \hat{\beta}_{2,0}, ..., \hat{\beta}_{2,3});$$

or, for some $j > 2$, $\hat{\boldsymbol{\theta}}_i^{j} = (\hat{\beta}_{j,0}, ..., \hat{\beta}_{j,2^j-1})$. Several values for $j > 2$ will be tested.

Example 2. *We would like to choose the levels that contain informations on the anomalies. To have an idea of the pathologies raised up by each level, we choose to represent all the coefficients of some wavelet level in order to get relevant informations on each anomaly type. For pattern anomalies, the small levels are expected to explain better the global changes. For this, we represent for a full level, a boxplot per feature. We add the corresponding values for each anomaly within a continuous line. For pattern anomalies, we can see in Figure 2.8 that in the small levels only one pattern appears clearly as an outlier (refer to the lines in red). For the two other pattern anomalies it is impossible to conclude based on this figure. We can also see that the periodicity anomaly starts to appear as abnormal at the level $j = 2$, and it stall even more than the other data in the coefficients for which $j = 3$. For the local anomalies, we consider the levels 4 and 5, that we represent in Figure 2.9. We can see that the fourth level is the first level where local anomalies start to appear clearly as outliers. It is even more evident in the fifth level.*

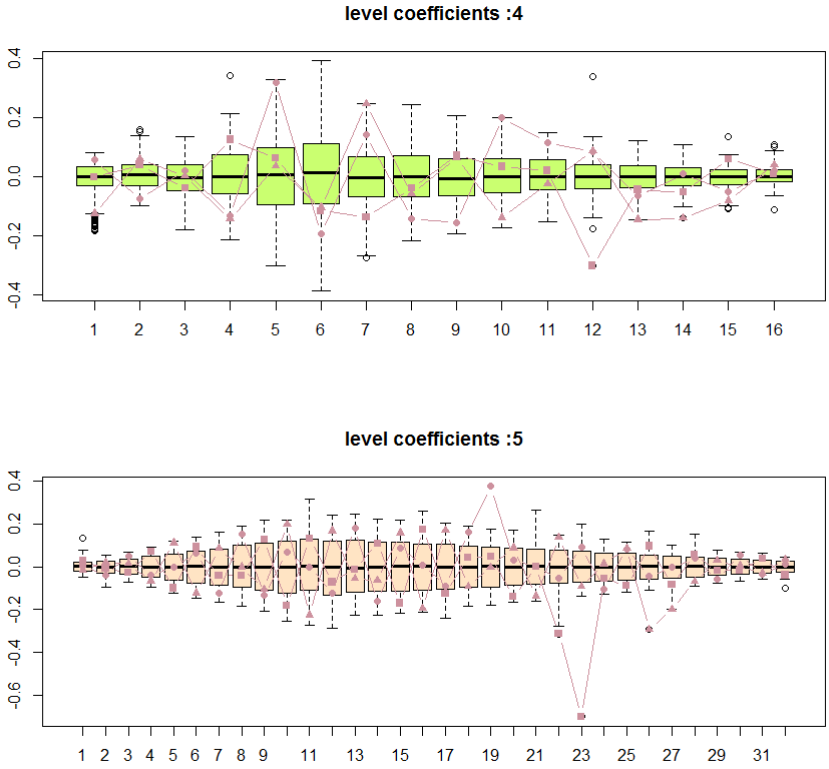

Figure 2.9 – Wavelet coefficients with $j = 4, 5$ and details for the local anomalies. Different markers to identify the local anomalies.

We can see that the different levels enable to raise several types of anomalies. Larger levels than $j = 5$ returns false positives, and the number of features becomes too large.

Kernel basis We consider in this paragraph a basis obtained from the Gaussian kernel eigenfunctions. For a given $\gamma > 0$, the Gaussian kernel is defined as

$$K_\gamma(s,t) = exp(-\gamma(s-t)^2).$$

The Gaussian kernel is universal in the sense that its eigenfunctions can represent any regular function (see Steinwart [78]). Thanks to this property, it is possible to reduce the dimension of the observations by keeping only the d first eigenfunctions, with $1 < d < p$. When the telemetry exhibits a pattern reproducing itself regularly, this basis is relevant. As the basis is fixed, if a unique pattern has a different shape, the features are likely to be highly impacted.
The Gaussian kernel K_γ is a Mercer kernel (a continuous symmetric, definite positive

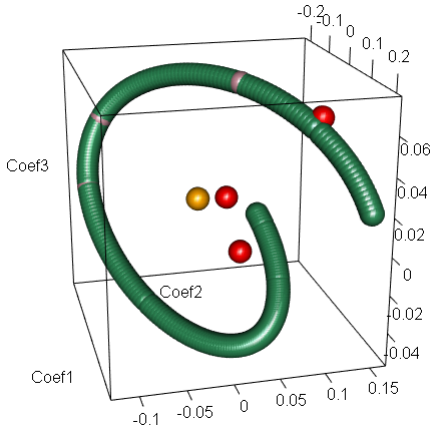

Figure 2.10 – Coefficients in the three first functions of the Kernel basis for $\gamma = 100$.

function). Thanks to Mercer theorem [59], we know that K_γ admits the following representation

$$K_\gamma(s,t) = \sum_{\lambda=1}^{\infty} \alpha_k \phi_\lambda(s)\phi_\lambda(t).$$

The functions $(\phi_\lambda)_{\lambda\geq 1}$ are the eigenfunctions of the kernel K_γ, and $(\alpha_\lambda)_{\lambda\geq 1}$ form a decreasing positive series, such that $\sum_{\lambda=1}^{\infty}\alpha_\lambda < +\infty$. Thanks to Steinwart, [78], we know that the eigenvalues $(\alpha_\lambda)_{\lambda\geq 1}$ decrease exponentially. This property justifies the fact that we can retain only a finite number d of eigenfunctions. As for the Fourier basis, we retain the d first coefficients that encompass more than 99% of the variance of the coefficients. The function X_i, $i = 1, ..., n$, can be approximated by $X_{i,d}$ corresponding to its projection onto the d first eigenfunctions (see Schölkopf and Smola [73] for more details):

$$X_{i,d}(t) = \sum_{\lambda=1}^{d} \theta_{i,\lambda}\phi_\lambda(t), \text{ where } \theta_{i,\lambda} = \int_0^1 X_i(t)\phi_\lambda(t)dt. \tag{2.4}$$

In practice, for all $i = 1 \dots n$, we estimate the coefficients $(\theta_{i,\lambda})_{\lambda=1,...,d}$ by the empirical coefficients $(\hat{\theta}_{i,\lambda})_{\lambda=1,...,d}$ as defined in the equation (2.2). The remaining vector to analyze is $\hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{i,1}, ..., \hat{\theta}_{i,d})$.

Example 3. *For the simulated example, we test several Gaussian kernels with* $\gamma = 10, 100, 10000$.
In this case, the four first functions we get from this kernel basis represent already more than 99% of the variance. To illustrate the anomalies that can be highlighted by this basis, we plot in Figure 2.10 the three first components corresponding to the kernel eigenfunctions with $\gamma = 100$*, the other options give similar results. The colors that we use are the same as previously. Once again, the local anomalies (in pink) do not appear as outliers, whereas the others do.*

2.3.2.2 Data-dependent bases

The principal component analysis (PCA) is a very powerful way to reduce the dimension of the data by retreiving the vectors that recover the best the variance of the data. If all the portions of the signal are really similar, we will see that, by taking a small number of features, we are likely to resume the data in a good way.
The kernel principal component analysis (KPCA) was introduced by Schölkopf et al. [74]. The idea is to build a non-linear form of PCA.
Let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a continuous fonction. Denote $\bar{\psi} = \frac{1}{n}\sum_{i=1}^{n} \psi(\mathbf{X}_i)$. The KPCA finds the eigenvectors of the matrix

$$\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n} \Big(\psi(\mathbf{X}_i) - \bar{\psi}\Big)\Big(\psi(\mathbf{X}_i) - \bar{\psi}\Big)^T.$$

As we can see, if $\psi = I_d$, then $\mathbf{C}_n$ is equivalent to the empirical covariance matrix, therefore the principal component analysis is a particular case of the kernel principal component analysis. Let $\mathbf{K}$ be the Gram matrix defined as

$$\mathbf{K}_{il} = \langle \psi(\mathbf{X}_i) - \bar{\psi}, \psi(\mathbf{X}_l) - \bar{\psi} \rangle = (\psi(\mathbf{X}_i) - \bar{\psi})^T(\psi(\mathbf{X}_l) - \bar{\psi}).$$

Finding the eigenvalues α_λ and the eigenvectors $\mathbf{V}_\lambda = \sum_{i=1}^{n} \theta_{i,\lambda}(\psi(\mathbf{X}_i) - \bar{\psi})$ of the matrix $\mathbf{C}_n$ can be done by solving the eigenvalue problem

$$n\alpha_\lambda \boldsymbol{\theta}_\lambda = \mathbf{K}\boldsymbol{\theta}_\lambda. \tag{2.5}$$

The projection of a test point $\psi(\mathbf{X}^*)$ onto the eigenvector $\mathbf{V}_\lambda$ is equal to $\gamma_\lambda \boldsymbol{V_\lambda}$, where

$$\gamma_\lambda = \sum_{i=1}^{n} \theta_{i,\lambda} \langle (\psi(\mathbf{X}_i) - \bar{\psi}), (\psi(\mathbf{X}^*) - \bar{\psi}) \rangle. \tag{2.6}$$

Example 4. *Let $(\phi_\lambda, \lambda \in \Lambda)$ be an orthonormal family of functions of $\mathbb{L}^2([0,1])$, and*

$$\psi(X_i) = \left(\frac{1}{p}\sum_{j=1}^{p} X_{ij}\phi_\lambda(t_j)\right)_{\lambda\in\Lambda} = (\hat{\theta}_{i,\lambda})_{\lambda\in\Lambda}$$

where $\psi : \mathbb{R}^p \to \mathbb{R}^d$, and $d = |\Lambda|$. Then, $\psi(X_i)$ returns the projection coefficients of X_i in an orthonormal basis, as defined earlier. Then we get

$$\psi(X_i) - \bar{\psi} = \begin{pmatrix} \hat{\theta}_{i,1} \\ \vdots \\ \hat{\theta}_{i,d} \end{pmatrix} - \begin{pmatrix} \bar{\hat{\theta}}_1 \\ \vdots \\ \bar{\hat{\theta}}_d \end{pmatrix} = \hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}}.$$

The covariance function becomes $\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})^T$. In this case, we can see that the KPCA applies a PCA on the features corresponding to projections in an

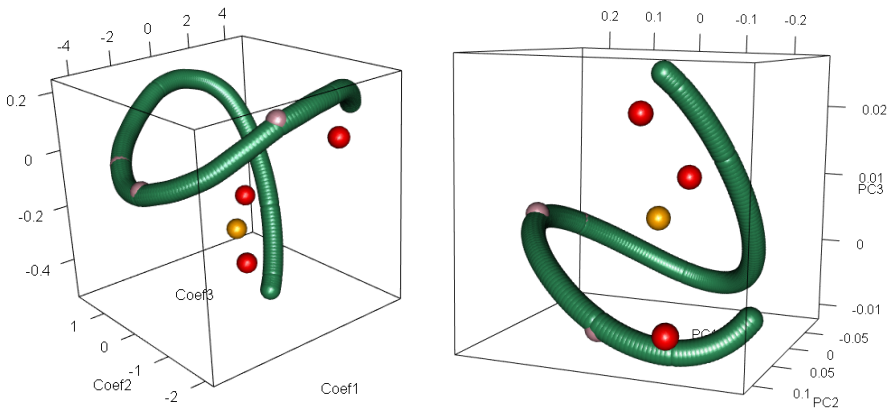

Figure 2.11 – Coefficients in the three first functions of the PCA basis (left) and the KPCA Component basis (right).

orthonormal basis. We can take for example a Gaussian kernel to recover the results of the previous section. We can also use projections on other fixed bases such as Fourier, wavelet bases.

Example 5. *We apply both the standard PCA and the KPCA with a Gaussian kernel* $K(x, y) = exp(-\gamma(x - y)^2)$*, with* $\gamma = 100$ *as in the Kernel basis section. For the PCA case, the three first components we get from this basis represent already more than 99% of the variance, whereas for the KPCA we need to retain the 4 first principal components. The three first components for both cases are represented in Figure 2.11. We can see that the pattern anomalies and the periodicity anomaly are clearly isolated. However, we also notice that the local anomalies seem to appear as more isolated than what we observed with the Fourier and the Gaussian kernel basis.*

2.3.3 Frequency spectrum

In the analysis of telemetries, changes in the period can be interesting to catch. By analysing the periodogram, also denoted as the frequency spectrum of the signal, some anomalies of non-periodicity can be detected. Instead of computing our features on the original signal, it is possible to build them on the frequency spectrum. The frequency spectrum has been detailed by Schlindwein et al. [72]. Let X be an arbitrary function of length p. Denote by $X(t)$ the t^{th} sample of the sequence.
This function can be decomposed in a Fourier basis using the discrete Fourier transform

$$d(\nu) = \frac{1}{p} \sum_{t=0}^{p-1} X(t) e^{2\pi i \nu t}.$$

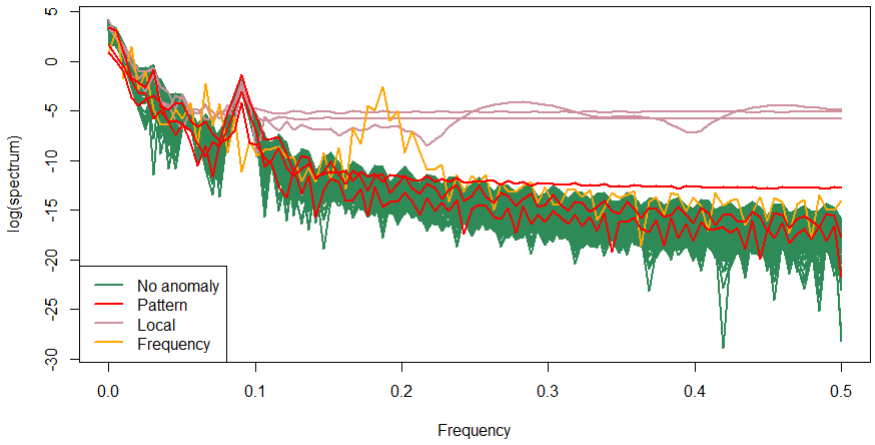

Figure 2.12 – Periodogram on simulated data.

The periodogram of the function is highlighting the frequencies corresponding to the function's period. It is equal to $I(\nu) = p|d(\nu)|^2$. If the daily signal is periodical with an unknown period, then it is more relevant to apply the functional decomposition on the periodogram than on the initial values $\mathbf{X}_i$. The periodogram is able to catch changes in the period without being disturbed by changes in the phasis. We can then look for pattern anomalies in the frequency spectrum.

Example 6. *We often represent the logarithm of the spectrum rather than the spectrum itself. In our case, the periodogram is represented in Figure 2.12. As we can see, the periodogram enables to catch local anomalies as well as periodicity anomalies. Then, we can extract the features (for example on kernel bases or the principal components). The local anomalies appear clearly on Figure 2.13 as outliers. However, as we have computed our signal with periodical functions, we must validate this hypothesis on real data.*

2.3.4 Curve registration

When extracting the information conveyed by curves, an additional difficulty may come from the fact that the curves may not be aligned. Indeed when finding a meaningful representative function that characterizes the common behavior of the sample, capturing its inner characteristics, as trends, local extrema and inflection points can be really relevant. In fact, the curves can be subject to both amplitude (variation on the y-axis) and phase (variation on the x-axis) variations with respect to the common pattern, as pointed out in [68], or [84] for instance. Hence, in the two last decades, there has been a growing interest

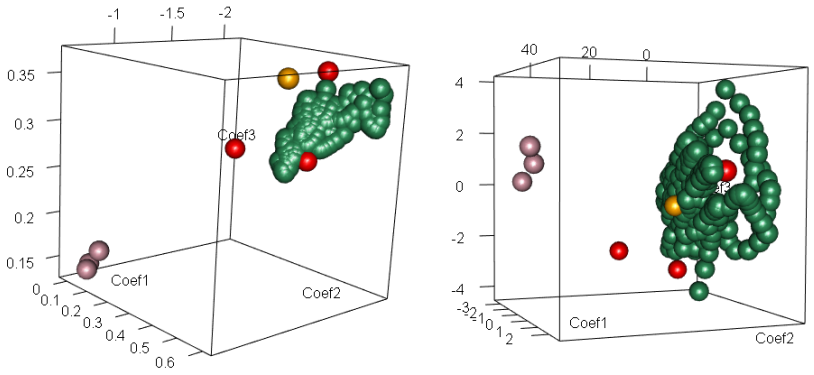

Figure 2.13 – Coefficients in the three first functions of the Kernel basis and on the Functional Principal Component basis computed on the periodogram.

for statistical methodologies and algorithms to align the curves prior to any analysis in order to remove this variability. Among all the methods we refer for instance to [87], [35], [25] and references therein. Hereafter, we focus on warping fonctions γ defined as follows. Denote $\mathcal{G}$ the set of all diffeomorphism γ subject to :

$$\mathcal{G} = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0, \gamma(1) = 1\}$$

γ is a function that warps the time. $X_i \circ \gamma$ denotes the time-warping of X_i by γ. When having two functions f_1 and f_2, the classical dynamic time warping algorithms finds the warping function that minimizes $\inf_{\gamma \in \mathcal{G}} ||f_1 - f_2 \circ \gamma||$, where $||\cdot||$ denotes the standard $\mathbb{L}^2([0,1])$ norm. This method is well adapted when we want to align two curves. To align more than two curves within the same algorithm, a new method was implemented.
Denote $q_i = sgn(X_i'(t))\sqrt{|X_i'(t)|}$ the squared-root slope function (SRSF) of X_i, can be shown that the SRSF of $X_i \circ \gamma$ that we denote (q_i, γ) is equal to $q_i(\gamma(t))\sqrt{|\gamma'(t)|}$. Then, we can prove that $||q_1 - q_2|| = ||(q_1, \gamma) - (q_2, \gamma)||$. Thanks to this property, we can define a distance considering the time warping. We call this distance the y-distance D_y.

$$D_y(f_1, f_2) = \inf_{\gamma \in \Gamma} ||q_1 - (q_2 \circ \gamma)\sqrt{\gamma'}||$$

The Karsher-means μ_f and μ_q are the functions that minimise

$$\mu_f = \underset{f \in L^2}{\operatorname{argmin}} \sum_{i=1}^{n} D_y(f, X_i)^2$$

$$\mu_q = \underset{q \in L^2}{\operatorname{argmin}} \sum_{i=1}^{n} \inf_{\gamma_i \in \Gamma} ||q - (q_i \circ \gamma_i)||^2$$

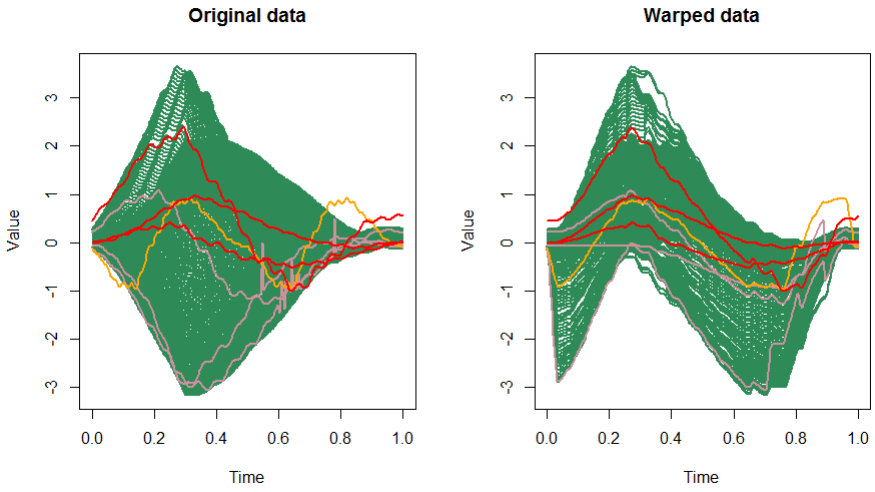

Figure 2.14 – Original curves and their aligned versions.

The idea of the algorithm is to start by considering the empirical mean of all functions, and then to warp all the curves according to this mean. At each iteration, compute the new Karsher means and update the warping functions. The algorithm stops after a given number of iterations or if the increment $||\frac{1}{n}\sum_{i=1}^{n}(q_i \circ \gamma_i)\sqrt{\gamma_i'}||$ is small enough. The aligned functions we get from the output of the algorithm are the functions $\tilde{X}_i = X_i \circ \gamma_i^*$. Those methods can be really useful when the patterns have a small phase or scale gap that can be ignored.

Example 7. *We align the curves we got from our simulated data thanks to the function align_fPCA which is part of the R package fdasrvf. This function takes a set of curves, align them thanks to shape functions as described earlier, and then we compute the functional principal component analysis on the aligned curves. On the following figure are represented the original data and their warped version after 12 iterations of the algorithm. After computing the FPCA on the aligned data, we keep the 7 first coefficients in order to keep 99% of the variance. We choose to represent the three first FPCA components before and after the warping algorithm.*

As we can see on Figure 2.14, aligning the data generates some noise in the components we get. Some of the curves can be so much warped that they no longer look like the original ones. Those extreme deformations can generate outliers that are not true outliers before the time warping procedure. However, on the figure 2.15 we notice that the local anomalies stall much from the nominal data. The outlier detection methods will help us in determining which case is the best for the outlier detection purpose.

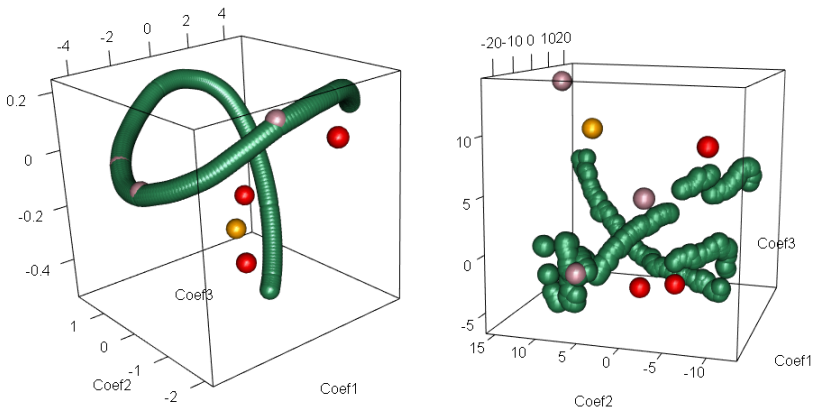

Figure 2.15 – Three first coefficients of the PCA applied on the original and aligned data.

2.4 Anomaly detection

In this section, we detail several methods to detect outliers using the features that have been previously selected. First it is important to define what is an outlier. Several definitions are used in the literature, and the most famous is probably the definition of Hawkins [40]. "An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism." Based on this assumption, we try to build a mathematical framework in order to detect them automatically. For this purpose, several ways can be considered.

- An outlier is an observation that lies at an abnormal distance to the other observations. Distance-based methods may be adapted to this point of view.
- An outlier is an observation situated in a low-density area. Then the density-based methods and minimum volume set estimation can be applied.

We develop outlier detection methods based on those two point of views. We will start with distance-based methods, and then density-based methods, inspired by Chandola's anomaly detection classification [19]. For all the outlier detection methods, we apply them on the feature sets we have defined in our previous examples, that we summarize as follows.

1. The first Fourier coefficients representing at least 99% of the variance.
2. The coefficients deriving from the Daubechies wavelet levels for which $j \in \{0, 1, 2\}$.
3. The coefficients deriving from the Daubechies wavelet levels for which $j = 3$.
4. The coefficients deriving from the Daubechies wavelet levels for which $j = 4$.
5. The coefficients deriving from the Daubechies wavelet levels for which $j = 5$.

6. The coefficients deriving from the projections onto a Gaussian kernel basis for which $\gamma = 100$, representing at least 99% of the variance.
7. The PCA coefficients representing at least 99% of the variance.
8. The KPCA coefficients where the kernel is a Gaussian kernel with $\gamma = 100$ representing at least 99% of the variance.
9. The coefficients deriving from the projections onto a kernel basis of the periodogram (we use the same kernel as the set 6).
10. The PCA coefficients applied on the periodogram.
11. The PCA coefficients applied on the aligned data.

For each application, we also present the results when we aggregate the novelties detected on every feature set.

2.4.1 Anomaly detection using distance methods

Some of the methods to detect outliers in functional data are based on distances between the curves. Usually, the usual distance refers to the Euclidian distance on the raw-data. As we have reduced the dimension of the data to get d coefficients for each curve, the distance we use is the Euclidian distance built on the features :

$$d(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j) = \sqrt{\sum_{\lambda=1}^{d} (\hat{\theta}_{i,\lambda} - \hat{\theta}_{j,\lambda})^2} = ||\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_j||.$$

As we have defined the feature-distance, we can now apply distance-based methods for outlier detection using this distance.

2.4.1.1 Hierarchical ascending clustering

The hierarchical ascending clustering is a widely used distance-based clustering method. Let $\boldsymbol{D}$ be the distance matrix of all our features, where $D_{ij} = d(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j)$, with $i, j = 1, ...n$. The usual clustering, as it is often applied, is unlikely to isolate outlier clusters. However, we know that if the intergroup similarity is the "single linkage", the outliers are likely to be situated in very small clusters.
All we have to do is to set the number of clusters in order to isolate the outliers. Tibshirani et al. [83] introduced the gap statistics, which is a criteria to set up automatically the number of clusters.
For each feature set defined earlier, we apply this automatic hierarchical clustering on the simulated data. We label as anomalous the individuals which are not in the biggest

Features	Number of clusters	Anomalies Pattern-Period-Local 3 - 1 - 3	False alarms
1 - Fourier	3	1 - 1 - 0	0
2 - Wavelet (0,1,2)	3	1 - 1 - 0	0
5 - Wavelet (5)	13	2 - 1 - 3	6
6 - Kernel ($\gamma = 100$)	3	1 - 1 - 0	0
7 - PCA	4	2 - 1 - 0	0
8 - KPCA	4	2 - 1 - 0	0
9 - Periodo + ker	6	1 - 1 - 3	0
10 - Periodo + PCA	5	1 - 1 - 3	0
11 - Shape + PCA	2	0 - 1 - 0	0
Aggregating	-	All anomalies detected (3 - 1 - 3)	6

Table 2.1 – Hierarchical clustering on simulated data.

cluster. The results we get with the simulated data are reported in the table 2.1. For almost all features types, the anomalies are in singleton clusters. The fifth wavelet levels feature set generates some misdetections, but almost all anomalies can be found there. We firstly notice that for all the other feature sets, we have detected nothing but real anomalies. We can see that the feature sets computed on the periodogram seem to be the best features for detecting outliers.
The automatic criteria for selecting the best number of clusters is difficult to set up. In fact, several numbers of clusters are possible, and as we do not know in advance the number of anomalies, it is difficult to fix it properly.
However, aggregating all the results of the clutering provides good results : all anomalies are found at leat once.

2.4.1.2 Temporal nearest neighbours dissimilarity

For the telemetry application, we have chosen to consider the signal as repetitions of curves. However, there is a logical order in these curves that can be taken into account. Thus, we have developed a novel distance-based method to handle this property.
In this section, we use a distance that we denote the kernel-distance d_K, which depends on a kernel K. This kernel has to be chosen in order to satisfy the following property.

For all $x, y, h \in \mathbb{R}$, $K(x, y) = K(x + h, y + h)$. Kernels such as Laplacian kernels and Gaussian kernels can be considered. The kernel-distance we choose is $d_K(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j) = K(i, j) \times d(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j)$. Then we define the temporal nearest neighbours dissimilarity (TNND) Δ as

$$\Delta_i = \sum_{j \leq i} d_K(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\theta}}_j).$$

This dissimilarity is motivated by the fact that some of the telemetries are repetitions of patterns that slowly evolve through time. Thus, a day of telemetry must be compared to the other days by giving more importance to the most recent previous days.

Proposition 1. *Suppose that* $\hat{\boldsymbol{\theta}}_i = \hat{\boldsymbol{\theta}}_{i-1} + \boldsymbol{\varepsilon_i}$, *with* $\boldsymbol{\varepsilon_i} \in \mathbb{R}^d$. *As the telemetry evolves slowly, we suppose that, if there is no anomaly, for all* $i = 1, ..., n$, $||\boldsymbol{\varepsilon}_i|| < \eta$ *for some* $\eta > 0$.
Then, if K *is a Laplacian Kernel,* $K(x, y) = exp(-\rho|x - y|)$,*where* $\rho > 0$, *we can show that*

$$\Delta_i \leq \eta \frac{e^{-\rho}}{(1 - e^{-\rho})^2}. \tag{2.7}$$

Proof. We know that for all $j = 1, ..., i - 1$, we have

$$d(\hat{\boldsymbol{\theta}}_{i-j}, \hat{\boldsymbol{\theta}}_i) = ||\sum_{l=1}^{j} \varepsilon_{i-l}|| \leq \sum_{l=1}^{j} ||\varepsilon_{i-l}|| = j \times \eta$$

Thus, we have

$$\Delta_i = \sum_{j=1}^{i-1} K(i - j, i) \times d(\hat{\boldsymbol{\theta}}_{i-j}, \hat{\boldsymbol{\theta}}_i) \tag{2.8}$$

$$\leq \eta \times \sum_{j=1}^{i-1} j \times e^{-\rho j} \leq \eta \times \frac{e^{-\rho}}{(1 - e^{-\rho})^2} \tag{2.9}$$

□

In the case where no anomaly occur, we can estimate η by the maximum $||\varepsilon||$ observed. It is also possible to consider that at most a proportion α of the observations are outliers. In this case, it is possible to estimate η by taking the $1-\alpha$ quantile of the $||\varepsilon_i||$, $j = 1, ..., n$ estimation, and to detect outliers when Δ_i exceeds this value.
This method can only be applied to telemetry data. However, we notice that, in some cases, it can also be generalized to test data, because several curves can come from the same test, where an input parameter evolves at each iteration. Another method will be provided in the appendix 2.A.

Features	Anomaly Pattern - Period - Local 3 - 1 - 3	False alarm
1 - Fourier	2 - 1 - 0	0
2 - Wavelet (lev 0,1,2)	3 - 1 - 0	0
3 - Wavelet (lev 3)	3 - 1 - 3	0
4 - Wavelet (lev 4)	2 - 1 - 3	0
6 - Kernel ($\gamma = 100$)	2 - 1 - 0	0
7 - PCA	2 - 1 - 0	0
8 - KPCA	2 - 1 - 0	0
9 - Periodo + kernel	0 - 0 - 3	0
10 - Periodo + PCA	0 - 0 - 3	0
11 - Shape + PCA	3 - 1 - 1	1
Aggregating	All anomalies detected	1

Table 2.2 – TNND results on simulated data.

Example 8. *Let us suppose that at most 2% of the observations are anomalous. Then we fix* $\eta = q_{0.98}(||\varepsilon||)$*. The results for each basis are stored in the table 2.2. We can see that the third level of wavelet is able to raise all the anomalies with no false alarm by setting the threshold thanks to the 98% quantile. The figure 2.16 shows the evolution of* Δ *for this feature set. But by analysing the evolutions of the TNND for the other features spaces, we can see that we could have detected all the anomalies, or almost, by setting the threshold to a lower value. The figure 2.16 also shows the results for the PCA built on the periodogram, for example. We also notice that only the feature set built on the aligned data generates false alarms.*

2.4.2 Local Outlier Factor

The Local Outlier Factor is a score introduced by Breuning et al. [14] to detect outlier data. In addition of detecting outliers, it returns a score to illustrate how the data is anomalous. This approach is also related to density-based clustering. This method inspired the ESA in the Novelty software [57]. It is local since this degree depends on how the object is isolated with respect to the surrounding neighbourhood. Suppose we have n objects $x_1, ..., x_n$ to cluster. Let k be the number of neighbours to consider. To simplify the notations, we

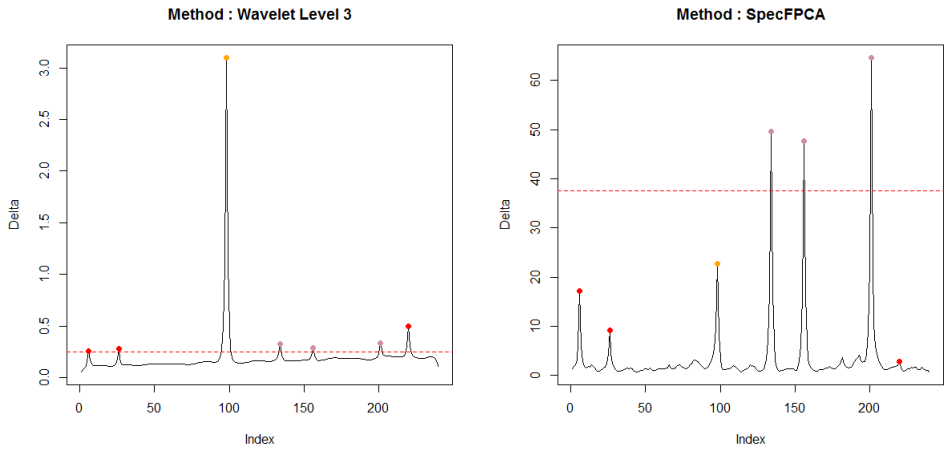

Figure 2.16 – TNND and anomaly detected with the threshold set as 98% quantile for the 3rd level of wavelet (left) and the Periodogram + PCA (right).

suppose that for x_1, x_2, x_3 all different, $d(x_1, x_2) \neq d(x_1, x_3)$.
- Let $d^k(p)$ be the k-distance of an object p, such that for k objects, the distance to p is closer than $d^k(p)$, and the other $n - p$ points are situated further. Let $N_k(p)$ be the k nearest neighbours of the object p.
- The reachability distance of an object p with respect to an object o is then defined as $r_k(p, o) = \max\{d_k(o), d(p, o)\}$. If p and o are sufficiently close, the distance between them is replaced by the k-distance of o.
- Then the local reachability density of p is defined as

$$lr_k(p) = \frac{k}{\sum_{o \in N_k(p)} r_k(p, o)}.$$

- The local outlier factor is then defined as

$$LOF_k(p) = \frac{1}{k} \sum_{o \in N_k(p)} \frac{lr_k(o)}{lr_k(p)}.$$

In other words, the LOF compares the nearest neighbours density distance of a given object with the nearest neighbours density distance of its nearest neighbours. When this score is close to 1, it means that the object is distributed in the same way as its neighbours. When having a large LOF, the corresponding object is likely to be an outlier.

Example 9. *In the table 2.3, we have indicated the results obtained if we consider as anomalous all the observations which lead to a* $LOF > 1.5$*, for* $k = 10$ *neighbours. We can see that, once again all the anomalies were found at least once. The results seem comparable and a little less satisfying than the results we get from the TNND. The choice*

Feature	Anomaly Pattern - Period - Local 3 - 1 - 3	False alarm
1 - Fourier	1 - 1 - 0	0
2 - Wavelet (lev 0,1,2)	2 - 1 - 0	0
4 - Wavelet (lev 4)	1 - 1 - 0	0
5 - Wavelet (lev 5)	1 - 1 - 3	2
6 - Kernel ($\gamma = 100$)	1 - 1 - 0	0
7 - PCA	2 - 1 - 0	0
8 - KPCA	0 - 1 - 0	0
9 - Periodo + kernel	3 - 1 - 3	4
10 - Periodo + PCA	2 - 1 - 3	1
11 - Shape + PCA	2 - 1 - 1	2
Aggregating	All anomalies detected	8

Table 2.3 – LOF results on simulated data.

of the threshold is important, and we can see that the detection varies a lot, from a feature set to another. On the figure 2.17, we have represented the two cases where the LOF exhibits the greatest and the smallest values. With the kernel features built on the periodogram set, we can detect all anomalies, while having some false alarms. On the other hand, with the same threshold, the LOF computed on the KPCA features detects only one anomaly. To explore several choices of threshold, we plot for each feature set the ROC curve. Those curves can be found in the figure 2.18. As we can see, the features based on the kernel basis on the spectrum are the best choices in this example.

2.4.3 Density approach

Anomalies are rare events, and we would like to find them in our coefficients $\hat{\boldsymbol{\theta}}_i \in \mathbb{R}^d$. The idea is to find a measurable subset of $\mathbb{R}^d$ that we call G, of probability greater than γ. Thomas et al. [82] gave a definition of this property. For this we set G such that its volume is as small as possible. In other word, we would like to solve the following problem

$$\min_{G \in \mathcal{B}(\mathbb{R}^d)} \{\mu(G) \mid \mathbb{P}(\hat{\boldsymbol{\theta}} \in G) > \gamma\},$$

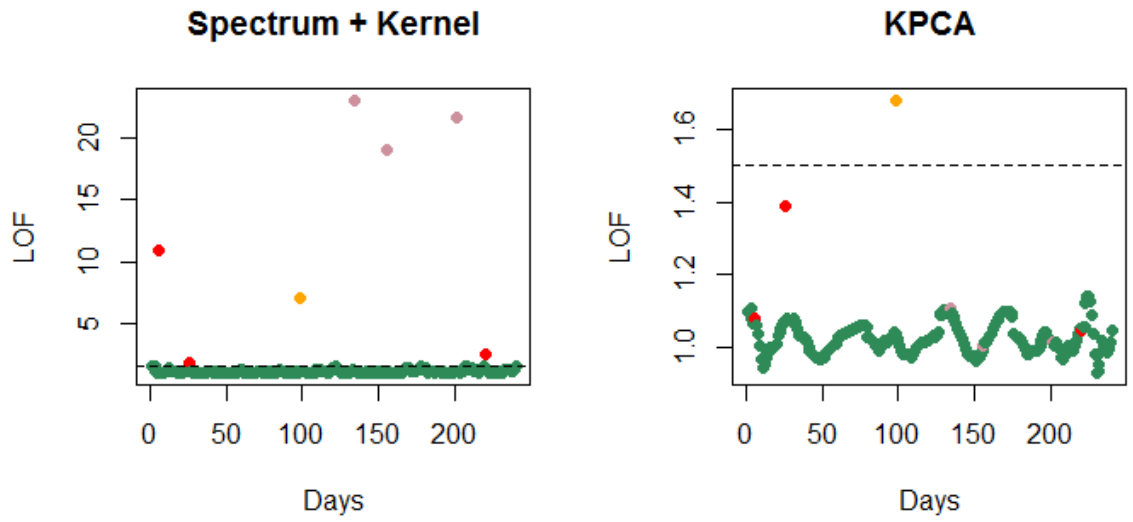

Figure 2.17 – LOF and anomaly detected - threshold at 1.5.

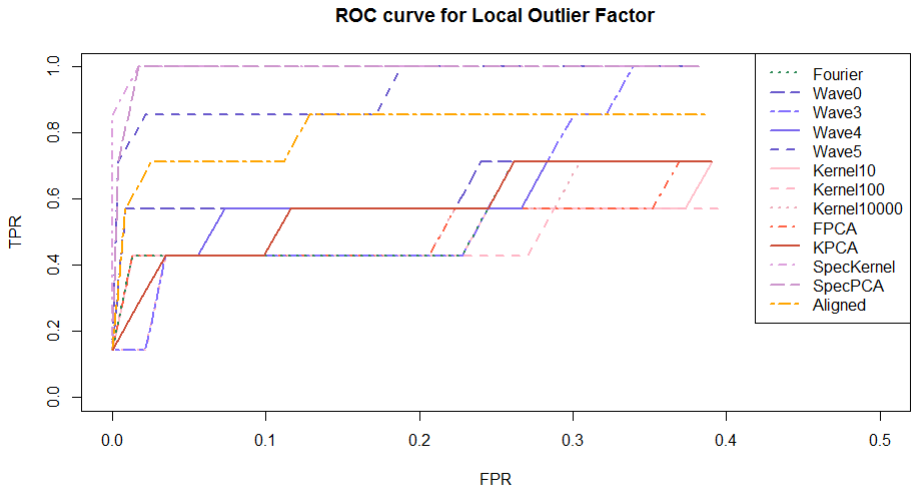

Figure 2.18 – ROC curves for the LOF computed on several feature sets.

where $\mu(G)$ is the Lebesgue measure of G. Here $\mathcal{B}(\mathbb{R}^d)$ denotes the set of all measurable subsets of $\mathbb{R}^d$ and $\mathbb{P}$ is the distribution of the coefficients $\hat{\boldsymbol{\theta}}$. In fact, G is a density level set, see Cadre et al.[15] for further details. By troncating the density to a level γ, we are able to isolate the rare events. The One-Class SVM is one of the most famous algorithm that applies outlier detection by defining the minimum volume set.
The One-class SVM was introduced by Schölkopf et al. [75], and this method was already

applied by Shawe-Taylor et al. [76] to time series data. Let $(\mathbf{x}_1, ..., \mathbf{x}_n) \in \mathbb{R}^{d \times n}$ be a sample data. As in the SVM classification, the idea of the One-Class SVM is to transform the data with a non-linear function ψ.
After transformation, the problem is to find the separating hyperplan orthogonal to the vector $\mathbf{w}$ that defines a minimum volume containing a given proportion of the data. It isolates the individuals situated in low-density regions with a margin ρ. Those points are considered as outliers. The values of $\mathbf{w}$ and ρ are the solutions of the optimization problem

$$\begin{aligned} &\min_{\mathbf{w},\boldsymbol{\xi},\rho} \frac{1}{2}||\mathbf{w}||^2 - \rho + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i \\ &s.t \ \ \langle \mathbf{w}, \psi(\mathbf{x}_i)\rangle \geq \rho - \xi_i, \ \forall i = 1..n \\ &\xi_i \geq 0, \forall i = 1..n. \end{aligned} \tag{2.10}$$

The parameter $\nu \in]0, 1]$ is fixed by the user. It represents an upper bound for the proportion of anomalies highlighted by the One-Class SVM. The decision function of this problem is $h(\mathbf{x}) = sgn(\langle \mathbf{w}, \psi(\mathbf{x})\rangle - \rho)$. It returns 1 if the individual is in a high-density region. Let $K(x, y) = \langle \psi(\mathbf{x}), \psi(\mathbf{y})\rangle$ be a kernel. The problem (2.10) is convex and can be solved by its dual.

$$\begin{aligned} &\min_{\alpha \in \mathbb{R}^n} \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ &s.t. \ \ 0 \leq \alpha_i \leq \frac{1}{\nu n}, \ \sum_{i=1}^{n}\alpha_i = 1. \end{aligned} \tag{2.11}$$

The points x_i such that $\alpha_i \neq 0$ are called the support vectors. We can remark that ν is also a lower bound for the fraction of support vectors. These support vectors define the limit between outliers and nominal data.
The decision function becomes $h(\mathbf{x}) = sgn(\sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}) - \rho)$. The advantages of the One-Class SVM is that the nominal set is defined only by the support vectors. Thus, the computation time for new values is much faster than for many other algorithms. Usually, we choose to use a Gaussian kernel K_γ as it was defined in Section 2.2.1.
This method is widely used for anomaly detection. NOSTRADAMUS is based on this method [32]. Thomas et al. [82] perform a new algorithm to tune the parameters γ and ν in order to find a minimum volume set of a desired level λ. Other methods exist to define the minimum volume set, such as isolation forests, introduced by Liu et al. [52].

Example 10. *On the table 2.4, we can see that the One-Class SVM generates systematically false positive detections. It can be explained as we know that it returns at least 5% of outliers whereas we truely have less than 3% of anomalies. Even if we set ν to a very small level, for example 1%, we would always have a large number of false positives. It can be explained by the fact that, in most feature sets, the features are not distributed*

Feature	Anomaly Pattern - Period - Local 3 - 1 - 3	False alarm
1 - Fourier	1 - 0 - 1	13
3 - Wavelet (lev 3)	1 - 1 - 3	14
5 - Wavelet (lev 5)	1 - 1 - 3	7
6 - Kernel ($\gamma = 100$)	1 - 1 - 1	15
7 - PCA	1 - 0 - 2	12
8 - KPCA	0 - 0 - 0	20
9 - Periodo + kernel	3 - 1 - 3	6
10 - Periodo + PCA	1 - 0 - 0	18
11 - Shape + FPCA	3 - 1 - 1	11
Aggregating	All anomalies detected	78

Table 2.4 – Table : OCSVM results on simulated data, with $\nu = 0.05$ and $\gamma = 1/d$.

according to an agglomerate of points. Hence it is more difficult to set the minimum volume set. Moreover, we detect less anomalies than with the other methods in general. The periodogram seems once again to be the best feature set since all the anomalies were found. These results can be justified by the fact that for most feature sets, the features are ordered according to their temporal occurrence. In fact, it seems that there is no clear high density areas, and density methods are expected to be less efficient than the TNND.

2.4.4 Conclusion on the methods

Given a simulated telemetry, we are able to conclude on several points to the efficiency of the different methods.

- The distance methods can be applied when a telemetry is evolutive through the year.
- The One-Class SVM is less efficient since the trending effect does not always generate high density area.
- The Temporal nearest neighbours distance is the best method to use in the given example, but the parameter η has to be chosen in order to be coherent with the

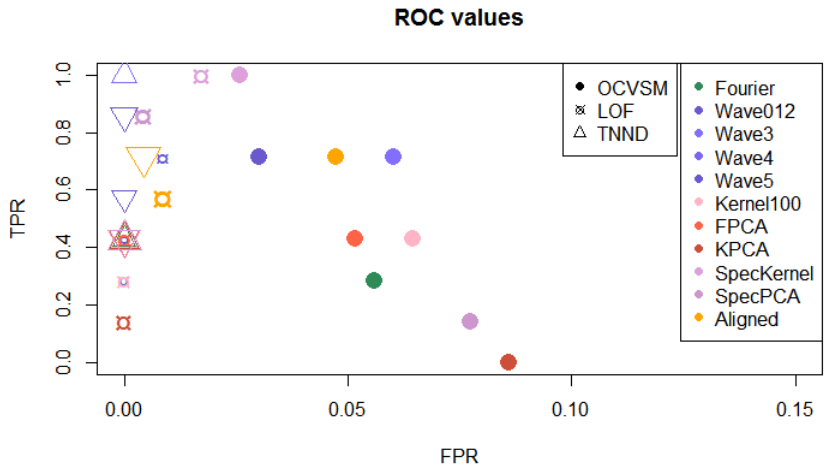

Figure 2.19 – True positive rate over False positive rate for all methods.

percentage of anomalies.

For each outlier detection method, at least one feature set was able to raise all anomalies with only a few false alarms. The TNND coupled with the features built on the periodogram seems to be the best compromize in this case.
All the results from the simulated data are shown in the figure 2.19. All those methods have to be implemented on real data to confirm our results, or to refine them.

2.5 Validation using satellite data

We have validated our methods with simulated telemetry. However, as we noticed in Section 2.2.2, the telemetries can have various behaviors, and consequently we would also like to test them on real data. For each data set, we will comment the results and illustrate the ones for which the methods that seem to return the most interesting results for the given usecase.

2.5.1 Telemetry

2.5.1.1 First application

We are dealing with the first telemetry we described earlier. As we have seen in Section 2.2.2, we know that some outliers are expected to be detected, such as spikes, drops in the amplitude of the signal, and pattern anomalies. On some portions, the signal can be altered. We have labeled by hand each portion of signal and computed all the features we

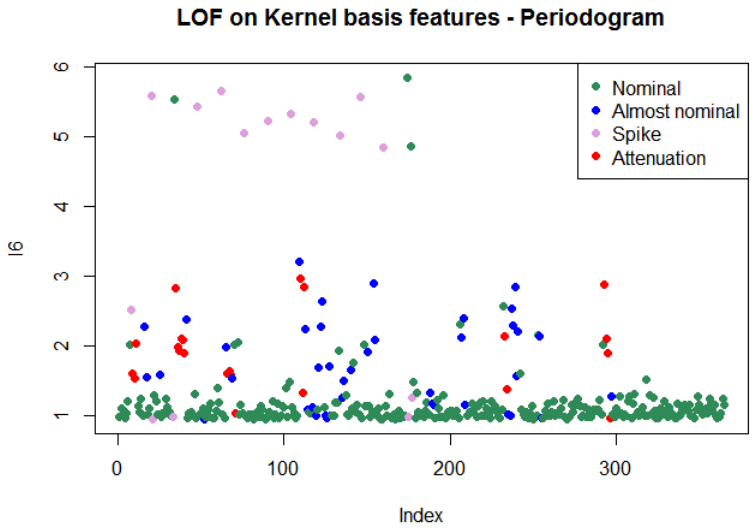

Figure 2.20 – Local Outlier Factor of the kernel features built on the periodogram, first telemetry application.

have developed in Section 2.3. However, this telemetry does not have a daily periodicity, which has a strong importance since the repetitive patterns are not all scaled in the same way (see Figure 2.1). Thus, we focus on the features built on the periodogram and the aligned data.
As expected, the hierarchical clustering returns a very small number of outliers because of the impossibility to fix properly the number of clusters. The One-Class SVM returns many false alarms. As some anomalies occur several times, it seems also that the OCSVM considers those anomalies as "nominal". The best results seem to be obtained with the Local Outlier Factor computed on the kernel basis, on the periodogram. We have labeled some of the days, that are represented in Figure 2.20 in order to visualize the efficiency of this algorithm.

2.5.1.2 Second application

The previous example was a critical one, and usually a telemetry dataset contains only a few outliers. For instance, the second application does not contain known anomalies, and the algorithms will help us to detect the changes in the behavior of this telemetry. The telemetry varies over the year and the daily portions of this telemetry are repetitive. Thus, any functional decomposition can be informative in this case.
As the patterns are really regular, it seems unnecessary to use the periodogram or the aligned data. The Principal Component Analysis seems to provide good results such as wavelets, kernel features, etc...

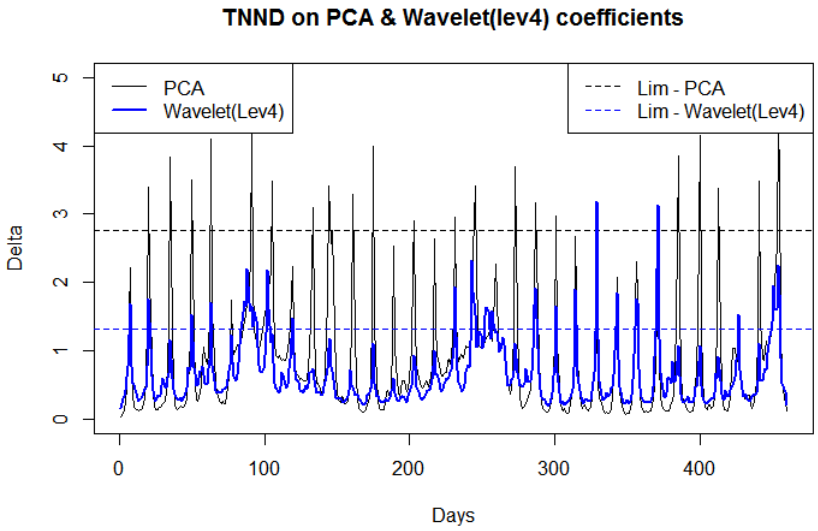

Figure 2.21 – TNND on two feature sets.

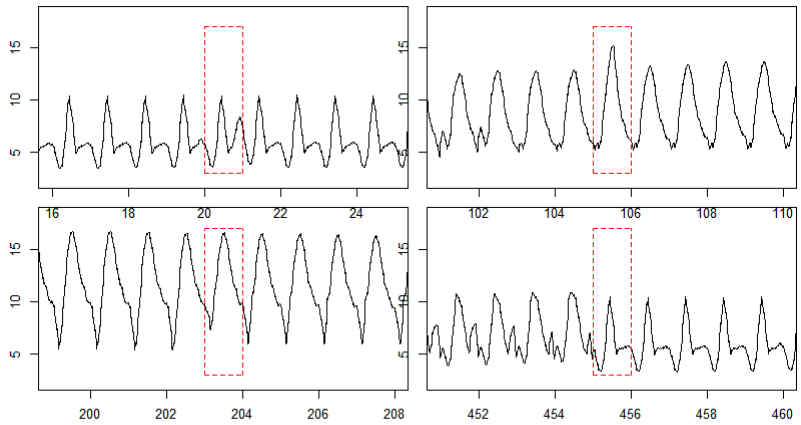

Figure 2.22 – Anomalies found with TNND, with both PCA and Wavelet features.

As the telemetry varies, the temporal nearest neighbour dissimilarity is the most powerful method. We represent the values of the statistics Δ computed on the PCA coefficients and the 4th level of wavelets. These results appear in Figure 2.21. We can see that both features sets enable to highlight more or less the same outliers, but not with the same importance. The figure 2.22 shows some examples of anomalies that can be found thanks to this method. Local changes as well as changes in the behavior of the telemetry can also be detected.

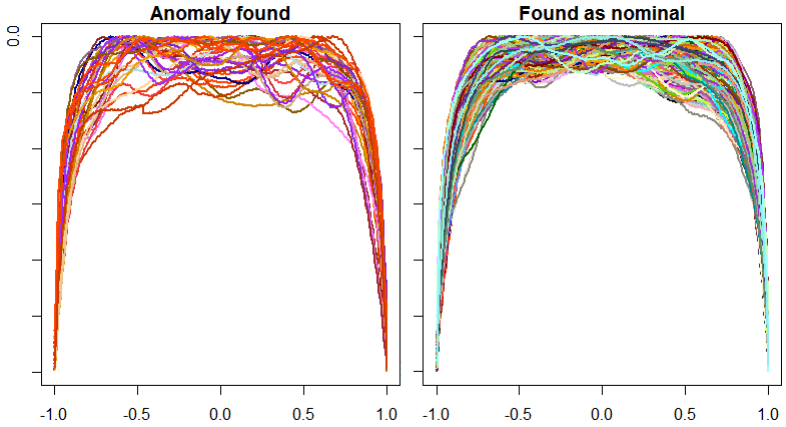

Figure 2.23 – One-Class SVM ($\nu = 0.05$) on the Principal Component features.

2.5.2 Gain Frequency tests

In this section, we apply the algorithms on test data instead of telemetry. As we have extracted functional features, it is important to validate this work directly with functional data. The Temporal Nearest Neighbour Dissimilarity cannot be used in this case, because the curves are not temporaly ordered. We could have used it if, for example, a test would have been tested under several conditions, for example if we get a curve each time we increase the temperature. This will be done in a further study. Refer to the appendix 2.A. We compute the One-Class SVM and the LOF on all the feature sets we have described in Section 2.3. Once again, two different methods are necessary to detect the different types of outliers. As it was already done on a previous study (see [9]), we can see that the best feature basis is the Principal Component Analysis. Both methods (LOF and One-Class SVM) give satisfying results, but the OCSVM generates again many false positives, as we can see on Figure 2.23. In this case, this can be explained by the fact that the alterations can only decrease the values of the curves. Thus, a "perfect" GF curve can be seen as an outlier since it is an upper boundary for all the GF curves. The Local Outlier Factor has similar results, with less false alarms. Therefore, this method seems more adapted to our problem. The fifth level of wavelets highlights the expected local anomalies we have described in Figure 2.3.

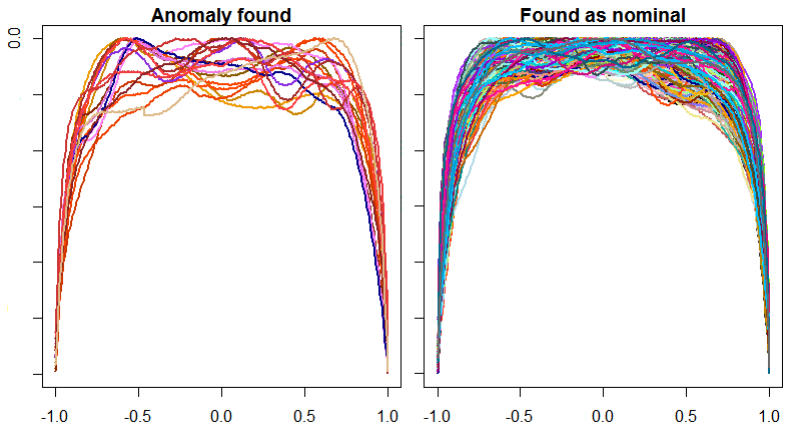

Figure 2.24 – LOF ($k = 20$ neighbours) on the Principal Component features. Threshold at 1.5.

2.6 Conclusion

We have explored several methods for unsupervised outlier detection in functional data. We have seen that one single method is unlikely to bring out all anomalies. But by aggregating the feature sets, we were able to detect all anomalies in all the sources of data. We have seen that the aggregation of the methods generates sometimes many false alarms, as for the OCSVM. Even for these extreme cases, the outlier detection methods still improve the manual review since they enable the expert to focus only on a reduced number of observations.
In addition, it seems that these methods can be more sensitive to some given properties of the data, in particular to the data periodicity. Here is what we can conclude.

- The One-Class SVM seems to generate more false alarms. However, if we can learn from the results of the classification for a while, it would be possible to switch to a semi-supervised approach and then to a supervised model. This method has the best efficiency when we know the percentage of anomalies, when the data are distributed according to an agglomerate of points.
- The Local Outlier Factor is very efficient when the data is not evolutive. It is often the best method to detect local anomalies, as for the Gain-Frequency data. It can be used with any features.
- The TNND is the best method to characterize the evolutive properties of telemetry. However, the percentage of outliers has to remain as small as possible.

We can also conclude on the features to be used in each case.

- The kernel basis, PCA, KPCA and the small level wavelet features can be used as soon as the curves are similar, such as the test data. In this case, the PCA and KPCA provide satisfying results to raise pattern anomalies.
- The periodogram features are very efficient when the data has a changing period, such as the first telemetry example. When the telemetry is very regular, it is unnecessary to use them.
- The high level wavelets are useful to detect local anomalies, as well as the periodogram features.
- Aligning the data can be useful when some patterns appear randomly in a constant telemetry for example. If the scale of the repetitive patterns changes randomly, it is better to use the periodogram.

If we combine the types of anomalies we want to detect and the data type, we can find the best method for the outlier detection. The feedback of our experts on this work will help us improving the method. A further step that could be foreseen would be to use the correlation between several telemetries as a feature, in order to detect abnormal correlations on several parameters.

2.A Appendix : Multivariate outlier detection in test data

In this section, we would like to detect outliers in multivariate functional data that are supposed to similar. For this purpose, we would like to test the equality of the distribution of m curves.
This is motivated by the application to the test data, where each amplification channel is tested several times on several conditions of temperature and vacuum.
For this purpose, we apply a test developed by Fromont et al. [31] that was primary designed to test the equality of the distribution of two samples. This test was designed as follows.

2.A.1 Two-sample test

Let $X^{(1)}$ and $X^{(2)}$ be the observations of two regression functions on p observation points.

$$\begin{cases} X_j^{(1)} &= f^{(1)}(t_j) + \epsilon_j^{(1)} \\ X_j^{(2)} &= f^{(2)}(t_j) + \epsilon_j^{(2)} \end{cases}$$

where $f^{(1)}, f^{(2)} \in \mathbb{L}^2([0,1])$, and $(\epsilon_j^{(1)}, \epsilon_j^{(2)})_{1\leq j\leq p}$ are i.i.d Gaussian noises which variance is unknown. In this framework, the hypothesis to test is $H_0 : \{f^{(1)} = f^{(2)}\}$. In this purpose, we build the test statistics T_K. Given K a kernel function, the test statistics T_K is defined as

$$T_K = \frac{1}{p(p-1)} \sum_{i\neq j=1}^{p} K(t_i, t_j)(X_i^{(1)} - X_i^{(2)})(X_j^{(1)} - X_j^{(2)}).$$

The hypothesis H_0 should then be rejected at the level α if T_K exceeds the $1-\alpha$ quantile of its distribution. In practice, the distribution of T_K is unknown. However, its quantiles can be estimated thanks to bootstrap permutations.
Thanks to Romano and Wolf's lemma [70], we know than if the $f^{(1)}$ and $f^{(2)}$ are independent, then T_K and T_K^b have the same distribution, where T_K^b is computed on bootstrap permutations of both vectors $X^{(1)}$ and $X^{(2)}$.

$$T_K^b = \frac{1}{p(p-1)} \sum_{i\neq j=1}^{p} K(t_i, t_j)\varepsilon_i^b\varepsilon_j^b(X_i^{(1)} - X_i^{(2)})(X_j^{(1)} - X_j^{(2)})$$

where $\varepsilon_j^b \in \{-1, 1\}$ for all $j = 1, ..., p$ are Rademacher random variable of probability $1/2$, and $b = 1, ..., B$ are the bootstrap realisation of the statistics T_K^b. Denote $\tilde{q}$ as the estimated quantile of T_K.

$$\tilde{q} = \frac{1}{B+1}\left(1 + \sum_{b=1}^{B} \mathbb{1}_{T_k^b \geq T_K}\right)$$

Then, we know thanks to the Romano and Wolf's Lemma that under H_0, for all $\alpha \in [0,1]$

$$\mathbb{P}(\tilde{q} \leq \alpha) \leq \alpha.$$

Hence, we reject the hypothesis H_0 at the level α if $\tilde{q} \leq \alpha$.

2.A.2 Generalization to m curves

We generalize this test in the case where we consider $m > 2$ curves. The hypothesis to test is then $H_0 : \{f^{(1)} = \ldots = f^{(m)}\}$, and the test statistics becomes

$$T_K = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} K(t_i, t_j) \sum_{l=1}^{m} (X_i^{(l)} - \bar{X}_i)(X_j^{(l)} - \bar{X}_j)$$

where $\bar{X}_i = \frac{1}{m} \sum_{l=1}^{m} X_i^{(l)}$, for all $i = 1, ..., p$. If $K(t_i, t_j) = \sum_{\lambda \in \Lambda} \varphi_\lambda(t_i)\varphi_\lambda(t_j)$ where $(\varphi_\lambda)_{\lambda \in \Lambda}$ is an orthonormal basis in $\mathbb{L}^2([0,1])$ and $\Lambda \subset \mathbb{N}$, then we get

$$T_K = \sum_{i=1}^{p} \sum_{l=1}^{m} \sum_{\lambda \in \Lambda} \left(\frac{1}{p} \varphi_\lambda(t_i)(X_i^{(l)} - \bar{X}_i) \right)^2.$$

We can estimate the rejection quantile as for the previous application, by using uniform permutations for all the instants t_i between the m curves.

2.A.3 Use the test statistics as a similarity measure

In addition to test the equality of the distributions, the T_K statistics provides a similarity measure within a set of m curves. In our application, we may consider n independent sets of m curves among which we would like to extract the sets where the curves are the most heterogeneous. In practice, this is how we use the T_K statistics, since we are not exactly in the same framework as the one required in the initial test. The sampled curves we have do not obey exactly to the same distribution, hence we obtain only rejections in the bootstrap approach.

2.A.4 Application to Gain-Frequency tests

We choose to compute the test statistics on a set of $n = 329$ amplification channels. Each amplification channel was tested m =4 times in different conditions of vacuum and temperature, and we would like to highlight the channels that changed the most through the phases. We compute the T_K statistics for each of those n sets of 4 curves, where K is a Gaussian kernel, $K(x, y) = exp(-\frac{|x-y|^2}{2\sigma^2})$, with $\sigma^2 = 0.5$.
We chose to represent the two highest T_K values in Figure 2.25 as well as the two smallest T_K values in Figure 2.26. Note that the smallest T_K value is worth around 2% of the

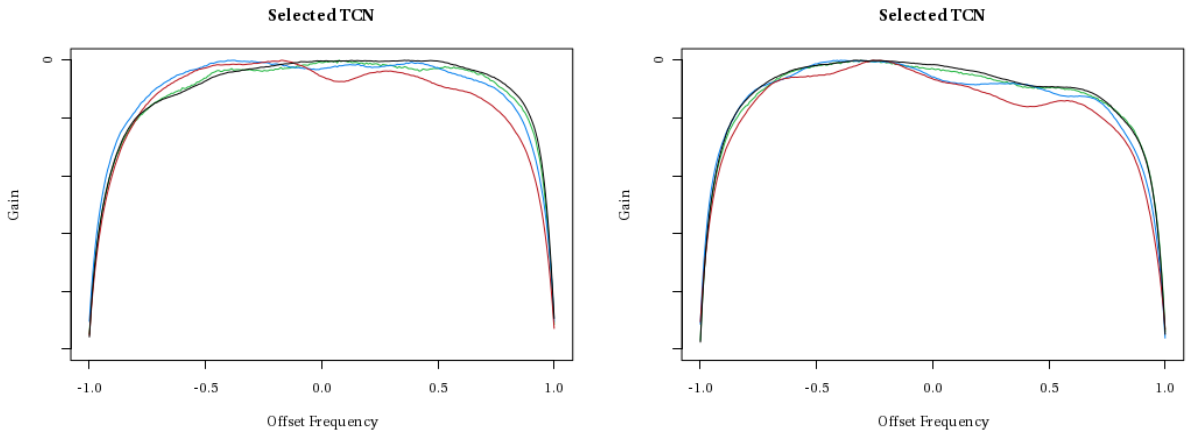

Figure 2.25 – The two amplification channels corresponding to the two highest T_K values.

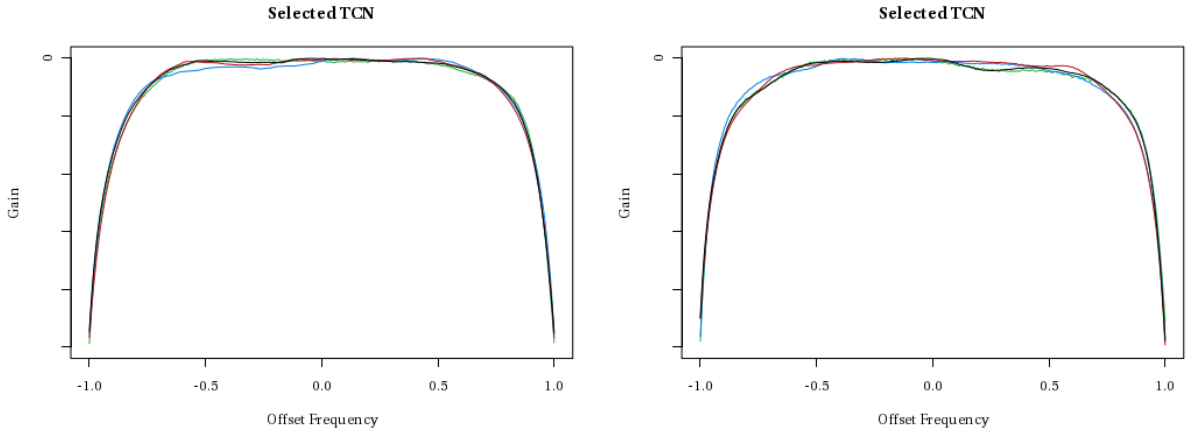

Figure 2.26 – The two amplification channels corresponding to the two smallest T_K values.

highest T_K value. As we can see, for both amplification channels returning the two highest values, the red curve exhibits ripples whereas the other curves do not exhibit such behaviors. Regarding the smallest T_K values, we can see that the curves look all really similar together.
This test statistics is really helpful to identify the channels that are the most sensitive to the external conditions. In practice, the satellites are subject to extreme environmental conditions once in space, that is why it is so important to identify such channels that are the most sensitive to the external parameters.

2.A.5 Conclusion

This approach is really interesting since it enables to target immediatly the channels that evolve the most through the test phases. The test as it was defined by Fromont et al. [31] is not really adapted to our framework. In practice, the test is rejected for all the data we have. However, it would be really interesting as a future work to adapt it to our situation, as it was shown that the order of the sorted T_K is already really relevant to target the biggest changes. Moreover, it is more adapted to our application than the Euclidian distance since the kernel method enables to consider the time warping of the curves in the several occurrences.

Chapter 3

Multiple Testing for Outlier Detection in Space Telemetries

Dans le chapitre précédent, nous avons mis en lumière des projections permettant de réduire la dimension des données. Nous avons vu que certaines de ces projections étaient plus pertinentes que d'autres pour mettre en lumière certaines familles de pathologies.
Dans ce chapitre, nous nous intéressons à la sélection automatique des niveaux de coefficients les plus intéressants pour la détection d'anomalies. Dans ce but, nous testerons seulement deux projections différentes : l'ACP et la décomposition en ondelettes de Haar. Nous réalisons cette sélection grâce à une procédure de tests multiples élaborée dans un contexte semi-supervisé.
Il s'agit d'isoler en premier lieu un sous-ensemble de courbes connues comme ne comportant pas d'anomalies, et de tester, pour chaque niveau de coefficients, l'égalité de leur distributions grâce à un test défini à partir de la distance de Wasserstein.
Les niveaux intéressants sont ceux pour lesquels l'hypothèse a été rejetée, après avoir contrôlé le taux de faux positifs grâce à la procédure de Benjamini-Hochberg.
Ce travail a été présenté à Helsinki en juillet 2017 dans le cadre de la conférence *European Meeting of Statisticians*. Par ailleurs, un article est actuellement en cours de soumission pour un numéro spécial de la revue *Transactions on Big Data* de l'éditeur IEEE consacrée aux applications au données spatiales "Big Data From Space". Une version de cet article est d'ores-et-déjà disponible via les plateformes HAL et arxiv.

Multiple Testing for Outlier Detection in Space Telemetries

Clémentine Barreyre[1], Béatrice Laurent[2], Jean-Michel Loubes[3], Bertrand Cabon[1], Loïc Boussouf[1]

Abstract

We propose a novel procedure for outlier detection in functional data deriving from satellites, in a semi-supervised framework. As the data is functional, we consider the coefficients obtained after projecting the observations onto orthonormal bases (wavelet, PCA). A multiple testing procedure based on the two-sample test is defined in order to highlight the levels of the coefficients on which the outliers appear as significantly different to the normal data. The selected coefficients are then called features for the outlier detection, on which we compute the Local Outlier Factor to highlight the outliers. This procedure to select the features is applied on simulated data that mimic the behavior of space telemetries, and compared with existing dimension reduction techniques.

3.1 Introduction

In this paper, we propose a novel procedure for outlier detection in the telemetries deriving from the satellites in a semi-supervized framework. As described by Chandola [19], an outlier is basically an individual that is significantly different from the normal behavior. In addition, several anomalies do not necessarily exhibit similar characteristics. Hence, detecting anomalies must be done by defining the normal behavior in the first place. Then, the deviation measured between an individual and the normal behavior gives good indications of the irregularities that occur in these novelties.
In the framework of this paper, the normal behavior can be partially learned thanks to a semi-supervised approach. It means that we can isolate a subset of data that do not contain any anomaly and that will be referred as the nominal set. Then, the anomalies can

[1]Airbus Defence and Space, Z.I. Palays, 31 rue des Cosmonautes, 31400 Toulouse, France
[2]Institut de Mathématiques de Toulouse (UMR 5219), INSA Toulouse, Université de Toulouse, 135 avenue de Rangueil, 31400 Toulouse, France
[3]Institut de Mathématiques de Toulouse (UMR 5219), Université Paul Sabatier, Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse Cedex 9, France

be defined as observations that differ from this observed nominal behavior. We therefore have two sets of data, a nominal one and another one containing a small proportion of anomalies, that we want to detect. Similar semi-supervised outlier detection were already treated by [85] and [77] for instance.
In this paper, we are interested in the application to space telemetries, that are frequent measurements of thousands of parameters in the satellite on-board all through its life. Those telemetries represent terabytes of historical data and exhibit various characteristics. Hence the detection of abnormal events must be done thanks to automated processing that must be self-adaptive to the data and to the anomaly types that may occur on these telemetries. We suppose that we have n days of a telemetry that can be considered as functions regularly sampled on p times each day. We assume that our observations are corrupted by independent and identically distributed (i.i.d.) Gaussian noise. This corresponds to the following model:

$$X_{i,j} = f_i(t_j) + \varepsilon_{i,j},\ i = 1...n,\ j = 1, ..., p, \tag{3.1}$$

where f_i is originally defined on a compact set, that can be modeled, without loss of generality, $f_i : [0, 1] \mapsto \mathbb{R}$. The variables $(\varepsilon_{i,j})_{1\leq i\leq n, 1\leq j\leq p}$ are i.i.d. centered Gaussian variables with variance σ^2 which is unknown. Since the telemetries are regularly sampled, we assume that for all $j = 1, ..., p$, $t_j = j/p$.
Outlier detection applied to streams of data and functional data has been treated for example in [3] where the authors used a sliding-window model, by Hoffmann [43], where the Kernel Principal Component Analysis (KPCA) decomposition is used for detecting novelties in hands digits. Ordoñez [61] applied functional data analysis methods on GPS measurements, and the outlier detection is done thanks to the depth metrics. Ren [69] proposed to use projections coupled with high-breakdown mean function estimator to detect outliers.
Our approach relies on three main steps. Firstly, we project the data onto orthonormal bases and collect the coefficients resulting from these projections. As our aim is to identify among the large number of coefficients the levels on which the anomalies are well sepatated to the nominal data, we must apply a multiple testing procedure. In the requirements of this framework, the estimated coefficients must be independent. For this reason, we choose to project our data onto the Haar wavelet basis and the principal component basis that satisfy this condition.
In a second step, we learn with a multiple testing procedure the projection coefficients that enable to separate the normal data from the anomalies. The selected coefficients are called features for anomaly detection.
Once the features are selected, we compute in a third step the Local Outlier Factor (LOF) [14] on these features to detect the abnormal behaviors and we compare our results with some common procedures to select features from observations of functional data. We choose this method because it is well adapted to the data we have. Other methods could have been chosen, as the ones described by Barnett [8], such as One-Class SVM [75],

One-Class Random Forest [27], or depth methods [28], for instance.
We then apply our procedure to benchmark data that are simulated thanks to the internal knowledge on space telemetries, and finally on a real telemetry.
The paper is organized as follows. In the first part, we detail the two orthonormal bases that we use in this paper : the Haar wavelet basis and the principal component basis. In the next section, we apply two-sample tests on each level of coefficients, and then keep the ones that are rejected after controlling the false discovery rate thanks to the Benjamini-Hochberg procedure [10]. We consider the Kolmogorov-Smirnov test, as well as two other tests built on Wasserstein metrics, defined for instance in [67].
The last section is dedicated to the application on a simulated and a real telemetry. We show in this section the good performances of our procedure by comparing it to more classical dimension reduction methods. We also apply the outlier detection on the raw-data, as it was done in [76], to show that this naive method is less powerful in our framework.

3.2 Projection onto orthonormal bases

The observations of the space telemetry data are high-dimensional, hence it is important to identify the levels of coefficients where the anomalies appear as significantly different to the nominal data.
We recall that our observations obey to Model (3.1) where the variables $(\varepsilon_{i,j})_{1\leq i\leq n, 1\leq j\leq p}$ are i.i.d. centered Gaussian variables with variance σ^2 and for all j, $t_j = j/p \in [0,1]$. We suppose that the nominal days are the n_0 first ones, where $n_0 < n$. We consider two types of features, the first ones are obtained from a projection onto an orthonormal basis of $\mathbb{L}^2([0,1])$, the second one correspond to a Principal Components Analysis.
These two bases were chosen because of their orthogonality in $\mathbb{R}^p$, which is needed to satisfy the independence condition to apply the multiple testing procedure. We provide deeper information on Section 3.3.3.

3.2.1 Projection onto the Haar basis

We assume that for all $i = 1...n$, $f_i \in \mathbb{L}^2([0,1])$, and we denote by $\langle ., . \rangle$ the usual scalar product in $\mathbb{L}^2([0,1])$. The functions f_i can be represented in an orthonormal functional basis of $\mathbb{L}^2([0,1])$. See [68] for more details. If $(\phi_\lambda)_{\lambda\in\mathbb{N}^*}$ is an orthonormal basis in $\mathbb{L}^2([0,1])$,

$$f_i(t) = \sum_{\lambda=1}^{\infty} \theta_{i,\lambda}\phi_\lambda(t), \text{ with } \theta_{i,\lambda} = \langle f_i, \phi_\lambda \rangle. \tag{3.2}$$

From the observations obeying to Model (3.1), the coefficients $\theta_{i,\lambda}$ are estimated by their empirical counterparts

$$\hat{\theta}_{i,\lambda} = \frac{1}{p}\sum_{j=1}^{p} X_{i,j}\phi_\lambda(t_j). \tag{3.3}$$

Due to independence constraints, we focus here on a wavelet basis, namely the Haar basis of $\mathbb{L}^2([0,1])$. For the sake of simplicity, we assume that p is a power of 2, namely $p = 2^{J+1}$. We set $\phi_0 = \mathbb{1}_{[0,1]}$ and $\psi = \mathbb{1}_{[0,1/2[} - \mathbb{1}_{[1/2,1[}$. For all $l \geq 0$, $k \in \Lambda(l) = \{0, 1, \ldots, 2^l - 1\}$, the Haar wavelet functions are defined as $\phi_{l,k}(x) = 2^{l/2}\psi(2^l x - k)$. Since we only have p regularly spaced observations per curve, we only keep the p first wavelet coefficients, which corresponds to all the coefficients up to the level $l = J$. We define

$$\Lambda = \{0\} \cup \{\lambda = (l,k),\ 0 \leq l \leq J,\ k \in \Lambda(l)\}, \tag{3.4}$$

and

$$\{\phi_\lambda, \lambda \in \Lambda\} = \{\phi_0, \phi_{l,k},\ 0 \leq l \leq J,\ k \in \Lambda(l)\}.$$

Hence, our initial set of data from Model 3.1 is represented by the same number of coefficients $(\hat{\theta}_{i,\lambda})_{\lambda\in\Lambda}$, $1 \leq i \leq n$. We can of course recover exactly the initial data $(X_{i,j})_{1\leq i\leq n, 1\leq j\leq p}$ from the coefficients.
For the next steps, we need the independence of the random variables $(\hat{\theta}_{i,\lambda}, \lambda \in \Lambda, 1 \leq i \leq n)$ with respect to λ and to i. In fact, we have such independence thanks to the three following properties :

- The Haar basis is orthonormal with respect to the discrete scalar product :

$$\forall \lambda, \lambda' \in \Lambda, \frac{1}{p}\sum_{j=1}^{p} \phi_\lambda(t_j)\phi_{\lambda'}(t_j) = \delta_{\lambda,\lambda'},$$

  where $\delta_{\lambda,\lambda'} = 0$ if $\lambda \neq \lambda'$ and 1 otherwise.

- For all i, the vector $\mathbf{X}_i = (X_{i,1}, ..., X_{i,p})'$ is a Gaussian vector.

- The vectors $(\mathbf{X}_i, 1 \leq i \leq n)$ are independent.

In this case, the feature selection will consist in finding the indices $\lambda \in \Lambda$ of interest in the decomposition. Let us now introduce the other set of features that we consider.

3.2.2 Principal component basis

The principal component analysis (PCA) is a very powerful way to reduce the dimension of the data by finding the vectors that recover the best the variance of the data. As this basis depends on the data, we apply it on a subset I_0 of the data known as nominal. We choose to set up I_0 as the set of even indices $\{2i, 1 \leq i \leq n_0/2\}$. Indeed, in order to

get independent features as previously, we compute the principal components only on a subset of $n_0/2$ nominal days and project the remaining data on this orthonormal basis. Let $\bar{\boldsymbol{X}}_{I_0} = \frac{1}{n_0/2}\sum_{i\in I_0} \boldsymbol{X}_i$ be the mean of the observations. The PCA finds the eigenvectors of the matrix

$$\boldsymbol{\Gamma}_{n_0/2} = \frac{1}{n_0/2}\sum_{i\in I_0}(\mathbf{X}_i - \bar{\boldsymbol{X}}_{I_0})(\mathbf{X}_i - \bar{\boldsymbol{X}}_{I_0})'.$$

Let $(\boldsymbol{\Phi}_\lambda)_{\lambda=1\dots p}$ be an orthonormal family of vectors of $\mathbb{R}^p$ built with the eigen vectors of $\boldsymbol{\Gamma}_{n_0/2}$, ordered by decreasing eigenvalues. This family is orthonormal for the scalar product $\langle\cdot,\cdot\rangle_p$ in $\mathbb{R}^p$ defined by :

$$\forall \mathbf{u},\mathbf{v}\in\mathbb{R}^p,\ \langle \mathbf{u},\mathbf{v}\rangle_p = \frac{1}{p}\sum_{j=1}^{p} u_j v_j.$$

Finally, we project the vectors $\mathbf{X}_i$ on the subspace generated by the vectors $(\boldsymbol{\Phi}_\lambda)_{\lambda=1\dots p}$, in order to get

$$\mathbf{X}_i = \sum_{\lambda\in\Lambda} \hat{\theta}_{i,\lambda}\boldsymbol{\Phi}_\lambda$$

where $\hat{\theta}_{i,\lambda} = \langle \mathbf{X}_i, \boldsymbol{\Phi}_\lambda\rangle_p = \frac{1}{p}\sum_{j=1}^{p} X_{i,j}\boldsymbol{\Phi}_{\lambda,j}$ and $\Lambda = \{1,...,p\}$. We still have the independence of the random variables $(\hat{\theta}_{i,\lambda}, \lambda\in\Lambda, i\notin I_0)$ with respect to λ and i. This is why we need to split the nominal set in two parts in this case.
The feature selection will consist in finding the levels λ of interest in this decomposition. Once we have presented the features that we consider, we will now see how to reduce the dimension by selecting a small set of features, in order to have good performances for the outlier detection procedure.

3.3 From approximation coefficients to features selection

We have represented our initial dataset by the approximation coefficients. Our aim is now to learn the levels on which the anomalies appear as significantly different to the nominal data. We therefore propose a new procedure, based on multiple testing that is well adapted to the problem at hand.

3.3.1 Univariate testing at each level

We remind that we have represented for all $1\le i\le n$, our vector of data $\mathbf{X}_i$ by the features $(\hat{\theta}_{i,\lambda})_{\lambda\in\Lambda}$ with $\Lambda = \{1,\dots,p\}$ for the PCA decomposition and Λ is defined by (3.4) for the Haar decomposition.
For each level $\lambda\in\Lambda$ we would like to know if the vector $\hat{\boldsymbol{\theta}}_{\cdot\lambda} = (\hat{\theta}_{1,\lambda},\dots,\hat{\theta}_{n,\lambda})$ contains

relevant information on the outliers. The vector $\hat{\boldsymbol{\theta}}_{.\lambda}$ is divided into two parts :

$$\hat{\boldsymbol{\theta}}_{0\lambda} = (\hat{\theta}_{1,\lambda}, \dots, \hat{\theta}_{n_0,\lambda})$$

corresponding to the set that is known to be nominal and

$$\hat{\boldsymbol{\theta}}_{1\lambda} = (\hat{\theta}_{n_0+1,\lambda}, \dots, \hat{\theta}_{n,\lambda})$$

for the other values. For this, we apply the two sample test to compare the distribution of the coefficients from both subsets.
Denote n_0' the number of individuals we keep in the nominal set. If the basis is the Haar wavelet basis, we have $n_0' = n_0$: all the nominal individuals can be used since the basis is fixed, therefore the features from both subsets are independent. For the PCA basis, we take $n_0' = n_0/2$ since we do not use the features arising from the data $\boldsymbol{X}_{2i}, i = 1, ..., n_0/2$ that were already used to compute the principal components.
Denote $\tilde{\boldsymbol{\theta}}_{0\lambda} = \hat{\boldsymbol{\theta}}_{0\lambda}$ in the case of the Haar basis, and $\tilde{\boldsymbol{\theta}}_{0\lambda}$ is composed by the odd indices of $\hat{\boldsymbol{\theta}}_{0\lambda}$ for the PCA.
We assume that the components of the vector $\tilde{\boldsymbol{\theta}}_{0\lambda}$ are independent and identically distributed (i.i.d.) with common distribution function $F_\lambda^{(0)}$ and that the components of the vector $\hat{\boldsymbol{\theta}}_{1\lambda}$ are i.i.d. with common distribution function $F_\lambda^{(1)}$. Both sets are independent. In the next section, we propose several testing procedures to test the null hypothesis $\{F_\lambda^{(0)} = F_\lambda^{(1)}\}$.
At first we introduce these tests for a single level λ. We will handle the problem of multiple testing in the next section.

3.3.1.1 Two sample tests

Let us suppose we have two independent vectors $\boldsymbol{X} = (X_1, ..., X_{n_0})$ i.i.d. with common continuous cumulative distribution function F and probability distribution P, and let $\boldsymbol{Y} = (Y_1, ..., Y_{n_1})$ i.i.d. with common continuous cumulative distribution function G probability distribution Q. The generalized inverse functions F^{-1} and G^{-1} are the also called the quantile functions. F and G are estimated by the empirical distribution functions F_{n_0} and G_{n_1}, where $\forall t \in \mathbb{R}$,

$$F_{n_0}(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbb{1}_{X_i \leq t},$$

$$G_{n_1}(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{1}_{Y_i \leq t}.$$

Let $X_{(1)} \leq \ldots \leq X_{(n_0)}$ be the ordered vector X. The quantile function F^{-1} is estimated by $F_{n_0}^{-1}$ defined as :

$$F_{n_0}^{-1}(p) = \begin{cases} X_{(1)} & \text{if } p < 1/n_0 \\ X_{(i)} & \text{if } p \in \left[\frac{i-1}{n_0}, \frac{i}{n_0}\right[\text{ and } 2 \leq i \leq n_0 \\ X_{(n_0)} & \text{if } p = 1 \end{cases}$$

and $G_{n_1}^{-1}$ is computed in the same way. We would like to test the following hypothesis :

$$\begin{cases} H_0 : F = G \\ H_1 : F \neq G \end{cases} \tag{3.5}$$

Many papers deal with the two-sample problem when no prior knowledge is assumed for the shape of the distributions. In this case non parametric tests are used to asses the veracity of the null assumption. The Kolmogorov-Smirnov test is a reference for this problem but other more recent tests can also be implemented using a distance that preserve the shape of the distributions. Here we use tests based on the Wasserstein metrics, reviewed for instance in [67].

3.3.1.2 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test relies on the fact that under the null hypothesis H_0, the distribution of the statistics

$$D_{n_0,n_1} = \sqrt{\frac{n_0 n_1}{n_0 + n_1}} \sup_{x \in \mathbb{R}} |F_{n_0}(x) - G_{n_1}(x)|$$

does not depend on F. The test is rejected when $D_{n_0,n_1} > c_{1-\alpha}$, where $c_{1-\alpha}$ is defined as the $1-\alpha$ quantile of D_{n_0,n_1} under H_0, in order to obtain a level α for the test. See [88] for further explanation.

3.3.1.3 Tests based on Wasserstein distances

When testing the equality of distributions, the choice of the distance used to evaluate the statistical gap between the two samples is important. In the following we introduce a test based on Wasserstein distance. First, for $d \geq 1$, consider the set $\mathcal{W}_q\left(\mathbb{R}^d\right)$ of probabilities with finite r-th moment. For μ and ν in $\mathcal{W}_q\left(\mathbb{R}^d\right)$, we denote by $\Pi(\mu,\nu)$ the set of all probability measures π over the product set $\mathbb{R}^d \times \mathbb{R}^d$ with first (resp. second) marginal μ (resp. ν). The L_q transportation cost between these two measures is defined as

$$W_q^q(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int \|x-y\|^q \, d\pi(x,y).$$

This transportation cost allows to endow the set $\mathcal{W}_q\left(\mathbb{R}^d\right)$ with the metric $W_r(\mu,\nu)$. More details on Wasserstein distances and their links with optimal transport problems can be found in [66] or [86] for instance.
The Wasserstein distance $W_q(P,Q)$ between two probability measures P and Q on $\mathbb{R}$ with $q \geq 1$ finite moments can be easily written as

$$W_q^q(P,Q) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^q dt$$

where F^{-1} and G^{-1} are the quantile functions of P and Q respectively. In our framework, we observe two n samples of i.i.d random variables with distribution P and Q. Let P_n and Q_n be the empirical distributions, hence the Wasserstein distance between these two empirical distributions is given by

$$W_q^q(P_n,Q_n) = \frac{1}{n}\sum_{i=1}^{n} |X_{(i)} - Y_{(i)}|^q.$$

Testing the equality of the two distributions is equivalent to test that the Wasserstein distance $W_q^q(P,Q)$ is equal to zero. As a matter of fact, under the assumption that $X_1,\ldots,X_n$ are i.i.d. with distribution P, $Y_1,\ldots,Y_n$ are i.i.d. with distribution Q and P and Q have finite q-th moment it is easy to conclude that $W_q^q(P_n,Q_n) \to W_q^q(P,Q)$ almost surely. However, designing a test requires to know the asymptotic distribution of a rescaled version of $W_q^q(P_n,Q_n)$ both under H_0 to estimate the level of the test and under H_1 to evaluate its power. Del Barrio et al. [23] and references therein give some insights while the case $P \neq Q$ is tackled in [24].

Yet the asymptotic distribution depends on the distribution P which is unknown. Hence we will use the test proposed in [67] which relies on the following property. If we consider the image by the distribution function G of the distribution P we obtain a distribution $P(G^{-1})$ with cumulative distribution function $G \circ F^{-1}$. Under the null assumption H_0, then $F = G$ and this distribution is the uniform distribution on $[0,1]$. Hence rather than testing the goodness of fit $P = Q$, we can use this non linear transformation to alternatively test the goodness of fit between the uniform distribution and $P(G^{-1})$. The main advantage of this setting is that the asymptotic distribution under the null assumption does not depend on the distribution P.
Assume that f and g are the density functions related to F and G. Let us suppose that there exists $C \in \mathbb{R}$ such that

$$\forall t \in \mathbb{R},\ \frac{g(F^{-1}(t))}{f(F^{-1}(t))} \leq C.$$

Let $\gamma = \frac{n_0 n_1}{n_0+n_1}$. According to Ramdas et al. [67], we know that, under the null hypothesis,

$$\gamma \int_0^1 (G_{n_1}(F_{n_0}^{-1}(t)) - t)^2 dt \to \int_0^1 (\mathbb{B}(t))^2 dt$$

and

$$\sqrt{\gamma} \sup_{t\in[0,1]} |G_{n_1}(F_{n_0}^{-1}(t)) - t| \to \sup_{t\in[0,1]} |\mathbb{B}(t)|$$

where $\mathbb{B}(t)$ is a Brownian bridge on $[0,1]$.
Consequently, it is possible to build a statistical test, to check the equalities of two distributions thanks to the Wasserstein distance, by using the asymptotic distribution of the test statistics under the null hypothesis to calibrate the quantiles. The null hypothesis is rejected if

$$T_2 = \gamma \int_0^1 (G_{n_1}(F_{n_0}^{-1}(t)) - t)^2 dt > c_{2,1-\alpha}$$

for the 2-Wasserstein test, or if

$$T_\infty = \sqrt{\gamma} \sup_{t\in[0,1]} |G_{n_1}(F_{n_0}^{-1}(t)) - t| > c_{\infty,1-\alpha}$$

for the ∞-Wasserstein test, where $c_{2,1-\alpha}$ is the $1-\alpha$ quantile of the distribution of $\int_0^1 (\mathbb{B}(t))^2 dt$, and $c_{\infty,1-\alpha}$ is the $1-\alpha$ quantile of the distribution of $\sup_{t\in[0,1]} |\mathbb{B}(t)|$.

Since we use the asymptotic quantiles of the test statistics under the null hypothesis, we have carried out some simulations to estimate the level of the test from a non asymptotic point of view.

3.3.2 Simulation

3.3.2.1 Level of the test

In this example, to evaluate the non asymptotic level of the tests, we take $n_0 = n_1 = n/2$ and we simulate both samples with standard Gaussian distributions. We simulate m i.i.d. samples. For $k = 1...m$, $X^k \sim \mathcal{N}_{n/2}(0, I_{n/2})$ and $Y^k \sim \mathcal{N}_{n/2}(0, I_{n/2})$, independent of X^k. At each iteration k, we test the equality of the distributions of X^k and Y^k, from which we get a p-value $p_{k,n}$. As usual, we estimate the level of our tests by the empirical estimator, namely the proportion of tests rejected at a level α among the m Wasserstein tests :

$$\hat{\alpha}(n) = \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}_{p_{k,n}<\alpha}.$$

We repeat it for both Wasserstein tests. When n is large, we expect $\hat{\alpha}(n)$ to be close to α. We choose $\alpha = 0.05$, $m = 5000$ iterations and $n/2$ varies from 50 to 10000. The results are summarized in Table 3.1, showing that the level of the test is close to 5% even for small sample sizes. This shows that the asymptotic test reaches the desired level very quickly, mostly with the 2-Wasserstein test.

We make the same work with X^k and $Y^k \sim \mathcal{E}_{n/2}(1)$, and we record the level corresponding to both Wasserstein tests. We present the results in Table 3.2.

	2-Wasserstein test		∞-Wasserstein test	
n	$\hat{\alpha}$	$\sqrt{Var(\hat{\alpha})}$	$\hat{\alpha}$	$\sqrt{Var(\hat{\alpha})}$
50	0.052	3.1×10^{-3}	0.0394	2.8×10^{-3}
100	0.046	2.9×10^{-3}	0.0360	2.6×10^{-3}
500	0.048	3.0×10^{-3}	0.0470	3.0×10^{-3}
1000	0.045	2.9×10^{-3}	0.0514	3.1×10^{-3}
10000	0.048	3.0×10^{-3}	0.0550	3.2×10^{-3}

Table 3.1 – Estimated level of the 2-Wasserstein and the ∞ - Wasserstein test on Gaussian distributions.

	2-Wasserstein test		∞-Wasserstein test	
n	$\hat{\alpha}$	$\sqrt{Var(\hat{\alpha})}$	$\hat{\alpha}$	$\sqrt{Var(\hat{\alpha})}$
50	0.050	3.1×10^{-3}	0.0406	2.8×10^{-3}
100	0.0462	3.0×10^{-3}	0.0400	2.8×10^{-3}
500	0.0462	3.0×10^{-3}	0.0482	3.0×10^{-3}
1000	0.0470	3.0×10^{-3}	0.0536	3.2×10^{-3}
10000	0.0456	3.0×10^{-3}	0.0518	3.1×10^{-3}

Table 3.2 – Estimated level of the 2-Wasserstein and the ∞ - Wasserstein test on exponential distributions.

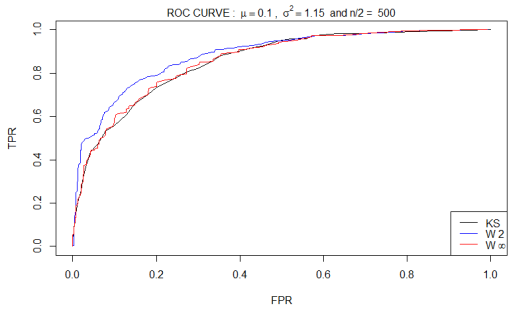

Figure 3.1 – ROC curves with $\mu = 0.1, \sigma^2 = 1.15, n = 1000$.

The level of the test is equivalent with exponential distributions, even for small values of n. We have carried out some simulations to compare the performances of the two tests based on T_2 and T_∞ with the Kolmogorov-Smirnov test, where the results are presented in the next section.

3.3.2.2 Simulation study of the power of the tests

We simulate independent samples arising from two different distributions. For $k = 1, \dots, m$, we simulate

$$X^k = (X_1^k, \dots, X_{n/2}^k) \sim \mathcal{N}_{n/2}(0, I_{n/2})$$
$$Y^k = (Y_1^k, \dots, Y_{n/2}^k) \sim \mathcal{N}_{n/2}(0, I_{n/2})$$
$$Z^k = (Z_1^k, \dots, Z_{n/2}^k) \sim \mathcal{N}_{n/2}(\mu, \sigma^2 \times I_{n/2}),$$

where $(\mu, \sigma^2) \neq (0, 1)$. We denote by F_X, F_Y and F_Z the cumulative distribution functions of X, Y and Z. We know that $F_X = F_Y$ whereas $F_X \neq F_Z$. We denote by $F_{n,X}^k, F_{n,Y}^k$ and $F_{n,Z}^k$ their empirical distribution function at the iteration k.
At each iteration k, we generate X^k, Y^k and Z^k and, based on these observations, we test the equality of the distribution F_X and F_Y, and then the equality of F_X and F_Z. The first test should be accepted whereas the second should be rejected.
Let $p_k^{(0)}, k = 1, ..., m$ be the p-values corresponding to the test $\{F_X = F_Y\}$, and $p_k^{(1)}, k = 1, ..., m$ be the p-values corresponding to the test $\{F_X = F_Z\}$ at the iteration k. From these p-values we can compute the true positive rate (TPR) and false positive rate (FPR) for each level α of the test, where $\text{TPR}(\alpha) = \frac{1}{m}\sum_{k=1}^m \mathbb{1}_{p_k^{(0)} > \alpha}$ and $\text{FPR}(\alpha) = \frac{1}{m}\sum_{k=1}^m \mathbb{1}_{p_k^{(1)} > \alpha}$.
The values reported draw a ROC curve from those simulations.
We test different values for n, μ, σ^2 in order to understand the behaviors of each test. We simulate $m = 2000$ samples, thus we have 4000 results of tests in total.

As we can see on the figures 3.1, 3.2 and 3.3, in all the simulations, the 2-Wasserstein test has the best power.

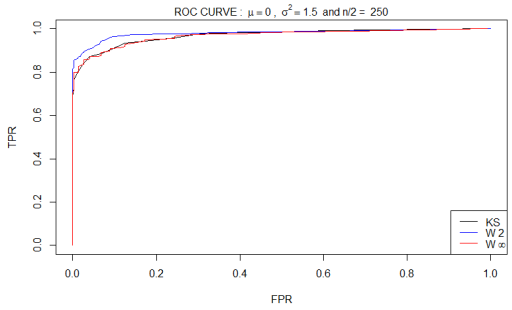

Figure 3.2 – ROC curves, with $\mu = 0, \sigma^2 = 1.5, n = 500$.

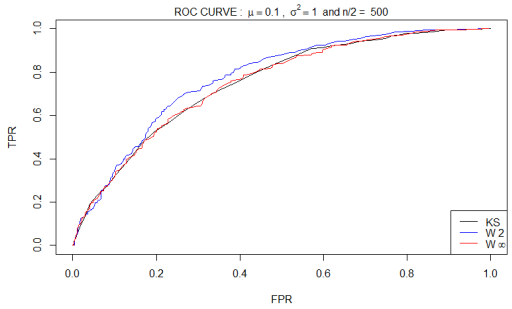

Figure 3.3 – ROC curves, with $\mu = 0.1, \sigma^2 = 1, n = 1000$.

3.3.3 Features selection while controlling the false discovery rate

3.3.3.1 A multiple testing framework

We remind that in both cases described in Sections 3.2.1 and 3.2.2, we have defined features $(\hat{\boldsymbol{\theta}}_{\cdot\lambda})_{\lambda\in\Lambda}$ to represent our data. Each feature vector $\hat{\boldsymbol{\theta}}_{\cdot\lambda}$ is divided in two parts, $\hat{\boldsymbol{\theta}}_{0\lambda} = (\hat{\theta}_{1,\lambda}, \ldots, \hat{\theta}_{n_0,\lambda})$ corresponding to the set that is known to be nominal which are assumed to be i.i.d. with distribution function $F_\lambda^{(0)}$, estimated from the features $\tilde{\boldsymbol{\theta}}_{0\lambda}$, and $\hat{\boldsymbol{\theta}}_{1\lambda} = (\hat{\theta}_{n_0+1,\lambda}, \ldots, \hat{\theta}_{n,\lambda})$ for the other values, assumed to be i.i.d. with distribution function $F_\lambda^{(1)}$. In order to select interesting features, we use the 2-Wasserstein test to test, for each feature, the null hypothesis

$$H_{0,\lambda} : \{F_\lambda^{(0)} = F_\lambda^{(1)}\}.$$

We are therefore dealing with a multiple testing problem since we test in both cases (wavelet or PCA decomposition), $|\Lambda| = p$ null hypotheses. Under $H_{0,\lambda}$, the p-value p_λ of the test is expected to be (asymptotically) uniformly distributed on $[0, 1]$ whereas, under the alternative, it is expected to be close to zero. The null hypothesis is rejected at level α for a p-value smaller than α. Nevertheless, it is well known that, when we deal with many

	Declared non significant	Declared significant	Total
True Null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	$m - m_0$
Total	$m - R$	R	m

Table 3.3 – Multiple testing procedure.

hypotheses, the probability to have at least one p-value smaller than the level α can be very large. Here, for both types of features, we test p hypotheses and for any $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\bigcup_{\lambda\in\Lambda}\{p_\lambda < t\}\right) &= 1 - \mathbb{P}\left(\bigcap_{\lambda\in\Lambda}\{p_\lambda > t\}\right)\\ &\sim 1 - (1-t)^p\\ &\xrightarrow[p\to\infty]{} 1\end{aligned}$$

We have used the independence of the random variables $(p_\lambda, \lambda \in \Lambda)$. Indeed, in both cases described in Sections 3.2.1 and 3.2.2, the basis $(\phi_\lambda)_{\lambda\in\Lambda}$ is orthonormal for the scalar product $\langle\cdot,\cdot\rangle_p$ in $\mathbb{R}^p$. Since the vectors $\mathbf{X}_i$ are Gaussian vectors and since $\hat{\theta}_{i\lambda} = \langle \mathbf{X}_i, \phi_\lambda\rangle_p$ for any $\lambda \in \Lambda$, we get that in both cases, the p vectors $(\hat{\boldsymbol{\theta}}_{\cdot\lambda})$ for $\lambda \in \Lambda$ are stochastically independent.

For example, we know that for a desired level $\alpha = 5\%$, the probability to reject at least one test among 50 if the null hypotheses are all true is already larger than 90%. Thus, we will use the procedure proposed by Benjamini and Hochberg [10] to control the false discovery rate. Let us first give some definitions. Consider that we have m hypotheses to test where m_0 hypotheses are true. R is the total number of rejections. The table 3.3 summarizes the situation. In this table, only R and m are known. The false discovery rate (FDR) is defined by

$$FDR = \mathbb{E}\left(\frac{V}{max(R,1)}\right).$$

The Benjamini and Hochberg [10] procedure allows to control the FDR.

3.3.3.2 Control the FDR with the Benjamini-Hochberg procedure

Benjamini and Hochberg [10] proposed a simple way to control the false discovery rate. Assume that all the p-values $(p_k)_{1\le k\le m}$ are independent random variables, and let $p_{(1)} \le p_{(2)} \le \ldots \le p_{(m)}$ be the ordered p-values. Let k^* be the largest k for which $p_{(k)} \le k\alpha/m$, then reject all the hypotheses $H_{0,(k)}, k = 1, ..., k^*$. Benjamini and Hochberg proved that

the FDR for this procedure does not exceed the level α.
Other procedures exist, and many of them are detailed in [71] but we limit ourselves to this procedure since it is very easy to implement. Note that it is justified in our cases (wavelet or PCA decompositions) since, as explained above, our p-values are independent. This procedure has been chosen since it is really easy to implement, but other procedures can be chosen, where many of them can be found in [71].

3.4 Outlier detection with the Local Outlier Factor

Once we have selected the features that isolate the best the anomalies in the sense of the comparison of the distributions, we are ready to apply an outlier detection technique on those features. There exists many unsupervised outlier detection methods. Chandola et al. [19] detailed many of them, and Gupta [37] specifies the outlier detection techniques for application to temporal data. They are often categorized into four categories: distance-based, density-based, as it was applied in [54] and clustering-based, as in Wu [89] and He [41]. For instance, the One-Class SVM developed by Schölkopf et al. [75] is a density-based method that is widely used for outlier detection purposes. However, most of time, the telemetry data we are dealing with evolve slowly because of seasonality effects. As a result, there is no immediate separation between the anomalies and the nominal data. Hence, the optimal conditions to use the One-Class SVM are not satisfied.
The Local Outlier Factor (LOF) is a score introduced by Breuning et al. [14] to detect outlier data. In addition of detecting outliers, it returns a score of anomalousness. This method is mixing the density-based and distance-based points of view, and has already been tested on space telemetry data since this method inspired the ESA in the Novelty software [57]. It is a local method since this factor depends on how the object is isolated with respect to the surrounding neighbourhood.
Suppose we have n objects $x_1, ..., x_n$ to cluster. To simplify the notations, we suppose that for x_1, x_2, x_3 all different, $d(x_1, x_2) \neq d(x_1, x_3)$.
Let $x \in \{x_1, ..., x_n\}$ and let $k < n$ be the number of neighbours that we consider. Choosing the best value for k is not so easy. In [14], different values are tested from 10 to 50, and the performance depends on how the data is distributed (different clusters, statistical fluctuations...). As our data depends on seasonnality effects, we chose to consider small values for k.

- Let $d^k(x)$ be the k-distance of x, which means that for k objects among $x_1, ..., x_n$, the distance to x is closer than $d^k(x)$, and the other $n-k$ points are situated further. Let $N_k(x)$ be the set of k nearest neighbours of x.
- The reachability distance of x with respect to an object o is then defined as $r_k(x, o) = \max\{d^k(o), d(x, o)\}$. If x and o are sufficiently close, the distance between them is

replaced by the k-distance of o.

- Then the local reachability density of x is defined as

$$lr_k(x) = \frac{k}{\sum_{o \in N_k(x)} r_k(x, o)}.$$

- The Local Outlier Factor is then defined as

$$LOF_k(x) = \frac{1}{k} \sum_{o \in N_k(x)} \frac{lr_k(o)}{lr_k(x)}.$$

In other words, the LOF compares the nearest neighbours density distance of a given object with the nearest neighbours density distance of its nearest neighbours. When this score is close to 1, it means that the object is distributed in the same way as its neighbours. When having a large LOF, the corresponding object is likely to be an outlier.

3.5 Applications

3.5.1 Application to benchmark data

We apply our outlier detection method firstly on a simulated telemetry that is really close to what we observe on real space telemetries. This simulated example has been created in order to ease the validation of each method.
We simulated $n = 480$ days of telemetry to get a significant number of signals after splitting the signal into days. We have $p = 256$ measurements per day. Each day of telemetry corresponds to an observation. The total signal symbolizes two year of telemetry, where a year lasts 240 days in this example. We added a Gaussian noise to our observations.
We introduced eight anomalies of several types, that represent a complete panel to what can be observed on real telemetries. These anomalies are situated only in the first 240 days.
The anomalies that are introduced are the following: 4 pattern anomalies (change in the pattern or in the amplitude of the data), 3 local anomalies (noise, spikes, data set to default value...) and one periodicity anomaly (two patterns instead of one). The pattern anomalies occur on days 6, 26, 70 and 220, the local anomalies on days 134, 156 and 201, and the periodicity anomaly on day 98.
The figures 3.4 and 3.5 show the portions of the signal containing anomalies, where the pattern anomalies are all in Figure 3.4. Some of the anomalies seem obvious (days $219, 97$), and some other are less pronounced (days $6, 26, 134$).
The aim of the study is to retrieve the days with abnormal behaviors.
For each day of telemetry, we compute the features based on the Haar wavelet basis and the PCA basis, and keep the raw-data as a reference result. We then compute the Local

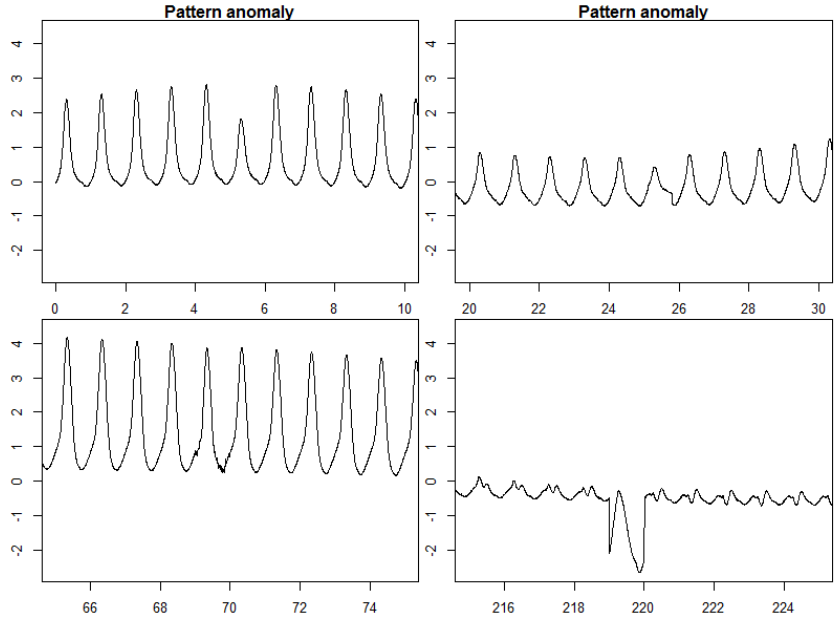

Figure 3.4 – Portions of the signal containing the 4 pattern anomalies.

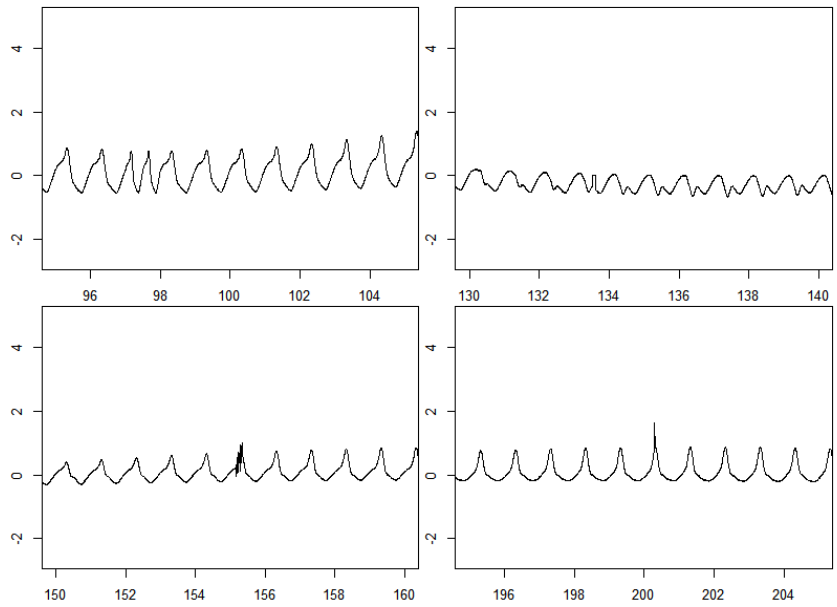

Figure 3.5 – Portions of the signal containing the periodicity anomaly (top left) and the 3 local anomalies.

Outlier Factor on the features that are not labeled as nominal. We will test several feature sets:

0. The raw data.
1. The full PCA coefficients.
2. The first d coefficients of the PCA representing at least 95% of the variance.
3. The full Haar wavelet coefficients.
4. The 8 coefficients from the levels $0, 1, 2$ of a Haar wavelet basis.
5. The 8 coefficients from the third level of a Haar wavelet basis.
6. The 16 coefficients from the fourth level of a Haar wavelet basis.
7. The PCA components resulting from the feature selection based on the 2-Wasserstein test.
8. The PCA components resulting from the feature selection based on the ∞-Wasserstein test.
9. The wavelet coefficients resulting from the feature selection based on the 2-Wasserstein test.
10. The wavelet coefficients resulting from the feature selection based on the ∞-Wasserstein test.

We do not present the feature selection on Kolmogorov-Smirnov test, because this test is really close to the ∞-Wasserstein test, hence returns the same feature selection. After selecting our feature sets, we compute the Local Outlier Factor. We have two parameters to calibrate : the number of neighbours to compute the Local Outlier Factor. and the threshold for the value of the Local Outlier Factor to detect outliers. For the first parameter, we choose $k = 10$ neighbours. As mentionned earlier, $k = 10$ seems to be a good choice since the data has a yearly trend, and it is better to consider a number of neighbours that is not too large.
For the threshold, we report the days where the LOF is greater than 2, and the ones that are larger than 4. The results are summarized in the table 3.4. There are 8 anomalies to detect, and the local anomalies are expected to be harder to spot than the other anomalies.

A first constat is that the Local Outlier Factor computed on the raw-data gives bad results, the same as the ones provides on the full PCA coefficients. There is too many redondant information in the raw-data, and the values are maybe too close to each other in general, hence do not allow to detect the outliers in a proper way. This reinforces the fact that projections are really helpful for highlighting outliers.

Feature	Anom LOF>2	False alarm	Anom LOF>4	False alarm	Nb features
0 - Raw data	2/8	7	1/8	0	256
1 - PCA - full	2/8	7	1/8	0	256
2 - PCA - 95%	2/8	8	1/8	0	3
3 - Haar - full	3/8	0	1/8	0	256
4 - Haar (lev 0,1,2)	3/8	0	1/8	0	8
5 - Haar (lev 3)	3/8	0	2/8	0	8
6 - Haar (lev 4)	5/8	0	3/8	0	16
7 - PCA - W2	8/8	0	7/8	0	17
8 - PCA - W∞	8/8	0	7/8	0	14
9 - Haar - W2	6/8	2	5/8	0	4
10 - Haar - W∞	6/8	2	5/8	0	4

Table 3.4 – Anomaly found for each feature set.

3.5.1.1 Comments on the PCA results

At first, one can notice that both two-sample tests generate almost the same feature selection, with, consequently, the same results. The set 2, containing the first PCA coefficients, does not contain major information on the outliers. In fact, the 4 first coefficients are not selected when we use the novel feature selection (sets 7 and 8). The selected features that are common from both tests are the ones for which

$$\lambda \in \{10, 17, 21, 29, 31, 27, 41, 57, 58, 65, 68, 69, 88, 94\}\,.$$

It indicates that resuming the full data is not the best way to detect outliers. In fact, the information contained on outliers is unlikely to appear in the first components since the anomaly is rare, thus not representative of a large portion of the variance of the data. The figure 3.6 shows clearly how performant our procedure is comparatively to the common way to proceed. It enables to isolate better the anomalies to the nominal data. Without controlling the FDR, we would have retained 20 features for the 2-Wasserstein test, and 19 for the ∞-Wasserstein test, instead of 17 and 14. As we have chosen $p = 256$, it enables to reduce even more the dimension of the data, with better results. In fact, without controlling the FDR, we would have missed one anomaly.
An important advantage of our procedure based on the feature selection is that it allows to well separate the values of the LOF of the outliers from the values of the LOF for nominal date and therefore it is not very sensitive to the threshold : with the value 2 and 4 of the threshold, we detect almost the same anomalies for the sets $7, 8, 9, 10$. Our procedure enlarges the margin between the nominal data and the outliers. This is an important result because for the other cases, the set of detected outliers is very sensitive to the threshold.

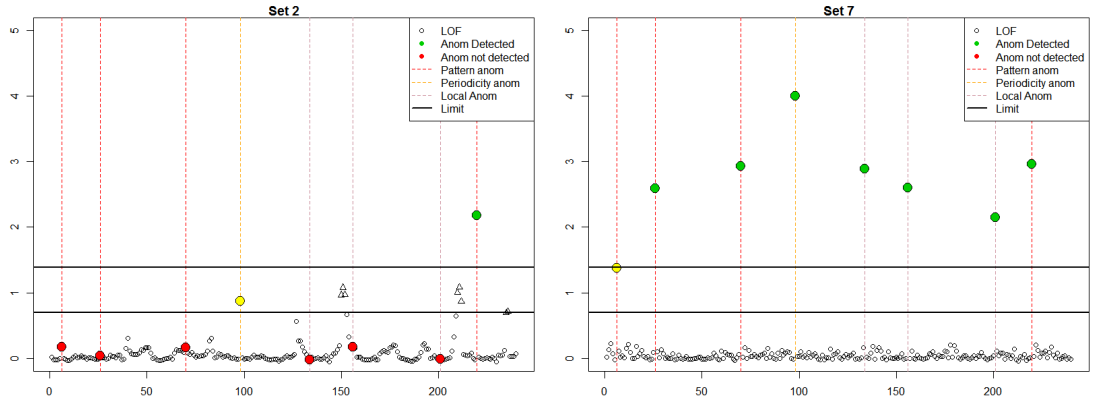

Figure 3.6 – LOF for both PCA coefficients selection - common (set 2) and novel (set 7), and limit (LOF=2,4).

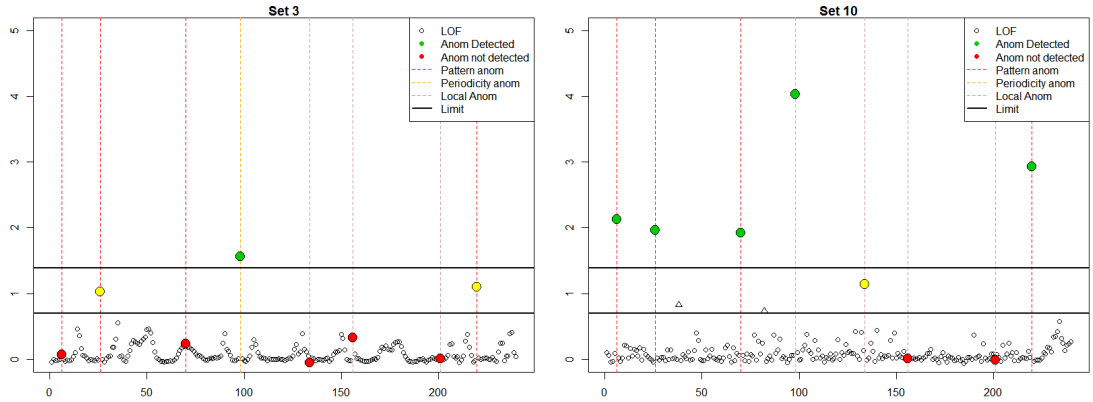

Figure 3.7 – LOF for two Wavelet sets coefficients selection -levels where $j \leq 2$ (set 3), novel procedure (set 10), and limit ($LOF > 2, LOF > 4$).

3.5.1.2 Comments on wavelet results

We have similar results with the wavelet decomposition. The levels $l \leq 2$ do not capture any information on the local events. The larger levels, like the level $l = 4$ exhibit easier the local events. However, the results are even better thanks to the automatic selection of wavelet coefficients, as we can see on Figure 3.7. Once again, our procedure to select the features enables to isolate better the anomalies to the nominal data. Such results can help to calibrate the level of rejection of the Local Outlier Factor, since a larger margin will lead to the optimum of the method. If we look at the 4 wavelet coefficients that were selected after the novel feature selection procedure, with the BH procedure at level 5%, we have:

- 1 position over 4 for $l = 2$,
- 1 position over 8 for $l = 3$,
- 1 position over 16 for $l = 4$,
- 1 position over 32 for $l = 5$,

The medium levels are well represented here. Without the BH procedure to control the FDR, we would have retained 12 features, instead of 4. The results are better without controlling the FDR. Indeed, we would have detected one additional anomaly at the level $LOF = 4$, while removing the false alarms for $LOF > 2$.

3.5.2 Application to real telemetry data

We apply our methods on two years of a real telemetry data. As the data is not labeled in this example we consider the first year as the nominal one. In fact, this can be justified by the fact that during their first year, the satellites exhibit nominal behaviors since each of its equipments have been checked recently during the test phases. As the PCA seems to be the basis for which the selection of the coefficients is the best, we apply the automatic coefficients selection on the PCA coefficients thanks to the 2-Wasserstein test. We also compare the results with the first PCA coefficients representing 95% of the variance. The Figure 3.8 represents on the x-axis the LOF computed on the PCA coefficients representing 95% of the variance, against the LOF computed on the coefficients selected thanks to the multiple testing, for each day. For confidentiality reasons, we have changed the dates.
As a first remark, we can notice that for this example, both LOF computations are quite similar. Secondly, we can see that the LOF computed on the PCA coefficients resuming 95% of the variance returns more extreme values : there are more days situated in the bottom right square of the graphic than on the upper left square. Consequently, there are more anomalies detected with the standard PCA.
We have highlighted some days that seem interesting.

- In blue we have represented a day for which the LOF is greater on the coefficients selected with our procedure than on the standard first principal components.
- In orange we have highlighted the highest LOF value deriving from the standard selection, which was not raised as so much abnormal on the coefficient selected thanks to multiple testing.
- The day in green represent the opposite situation where the LOF computed on the coefficients selected with multiple testing is greater than the LOF computed on the standard principal components.
- The days in red and yellow correspond to the biggest LOF for both selections.

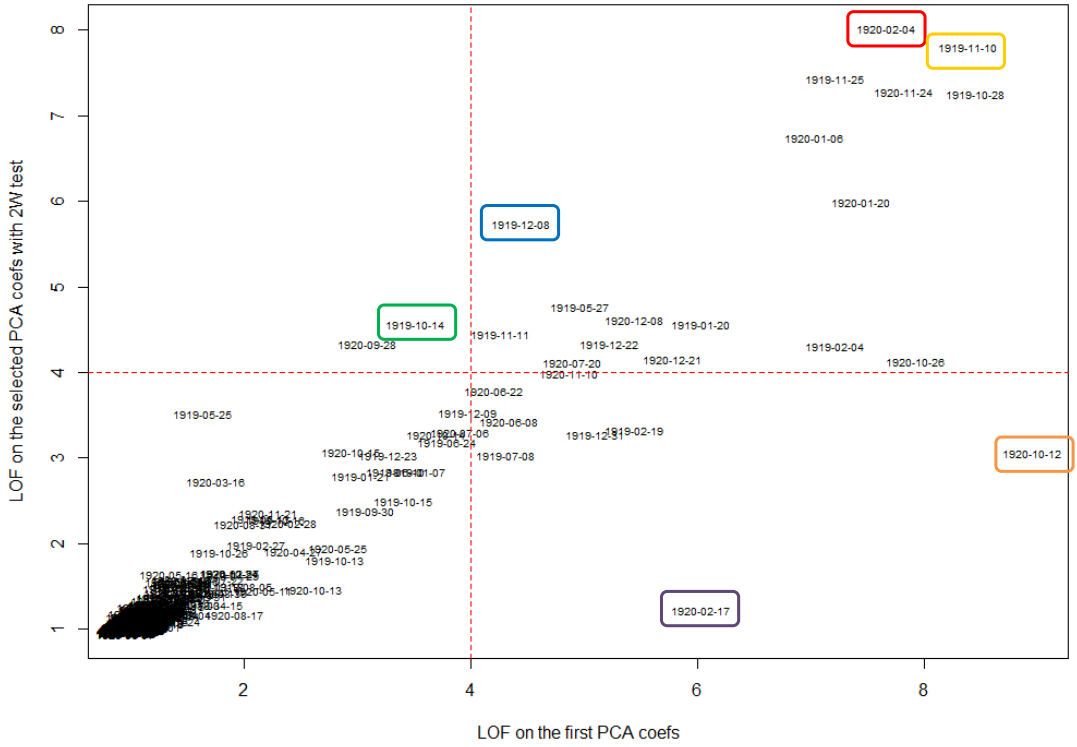

Figure 3.8 – LOF values on a real telemetry computed on PCA coefficients, representing 95% of the variance (on the x-axis) and selected thanks to the 2-Wasserstein test (on y-axis).

- The day in purple was not detected as an outlier by the multiple testing selection.

The corresponding days are represented on Figure 3.9. As we can see, for each of the days we have highlighted, we are able to catch by eye some changes. The day that was not detected as an outlier by our new procedure (in purple) correspond to a minor change, than can be found elsewhere on this telemetry, hence corresponding to a false detection. It reinforces the idea that our new procedure limits the false detections. For this example, we remark that both selection exhibit mostly the same days as outliers, which can be explained by the fact that our procedure selected some of the highest principal components $\lambda \in \{1, 3, 4, ...\}$.

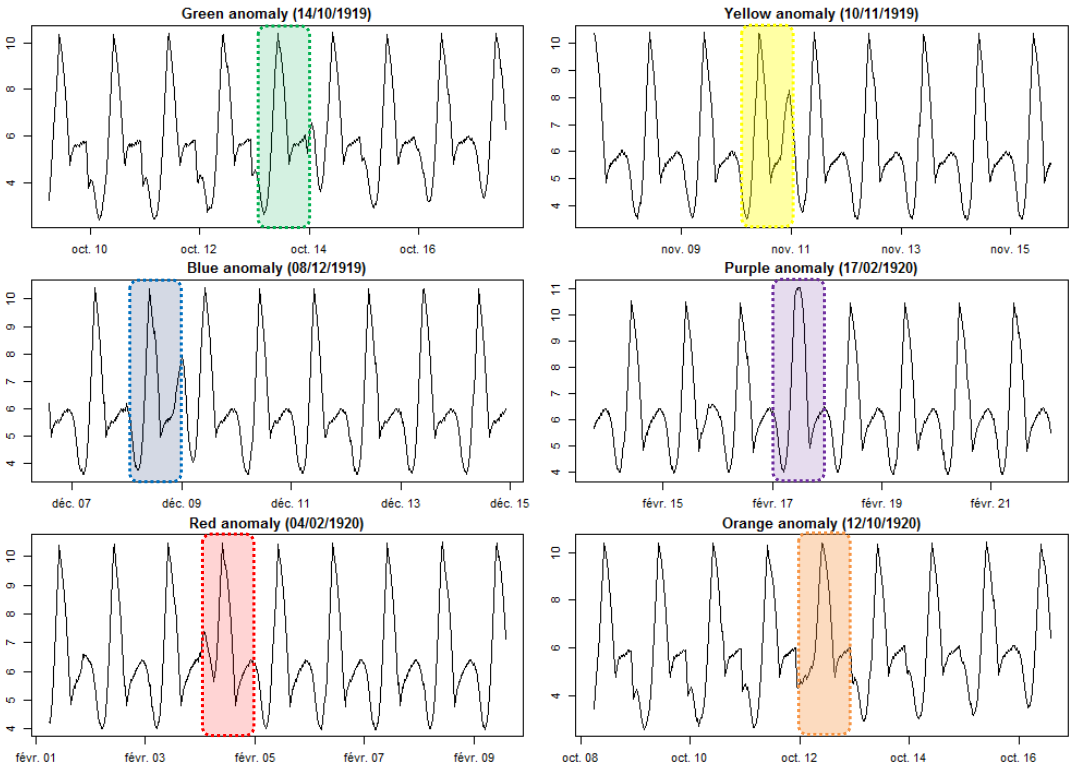

Figure 3.9 – Corresponding outlier detection.

3.6 Conclusion

This original features selection procedure is really relevant for the detection of outliers. It reinforces the idea that clustering and outlier detection cannot be addressed by similar methodologies. In the case of the PCA, only the first principal components are usually retained for clustering. What we have shown in this paper is that, even if it is the best method to reduce the dimension of the data, it is not the best one to exhibit abnormal behaviors. In fact, those unexpected events do not represent a large portion of the variance of the data, since the outliers are rare and do not have a repetitive signature. Consequently, the anomalies are unlikely to appear clearly as outliers in the very first components, but some further components can get such information. This method is really adapted to the application to the Space telemetry since it enables to target obvious anomalies as well as lighter local anomalies.
We have also mentioned the case of the Haar wavelet basis. In fact, it is not easy to know which are the levels to retain for such analysis. One can be interested by concentrating the global information, and some other to get details as well. The selection we develop enables to guarantee both types of information on the data.
The PCA decomposition with our feature selection procedure gives the best results in this study. The wavelet decomposition combined with the original feature selection procedure gives also good results. An important advantage of the wavelet decomposition compared with the PCA decomposition is that it is a fixed basis, whereas the PCA is based on a data dependent basis and to keep independence properties, we had to isolate some data to compute the principal components. Moreover, if we want to implement online procedures, which is often the case when we deal with big datasets, a fixed basis is much more relevant.

Our approach requires to have some knowledge on the data because a set of data that does not contain any anomaly has to be isolated. The difficulty to implement a total unsupervised approach comes from the fact that the distribution of the features for nominal data is unknown.
In this direction, a test has been developed by Candelon et al. [17], their method is based on bootstrap permutations, which increases a lot the computation time, and it is not suitable for multiple testing, as in our situation.
One further work would be to adapt this procedure to the online context as it was treated for instance in [79], where additional data may change in our case the feature selection online, and the nominal behavior can be learned thanks to the numerical simulations that can be provided in order to reproduce how the satellite behaves in various situations.

3.A Appendix : Robustness to dataset contamination

We have built our procedure based on the assumption that a subset of curves do not contain any anomaly. Now we would like to check what happens if some anomalies remain in this subset. In fact, some anomalies can have been missed by the operator who did the classification, or we could for example compare two years of telemetries on the assumption that no anomaly occured in the satellite during its first year in Space since its behavior has been checked recently thanks to the test phases.
For this application, we simulate two years of a new telemetry, where 2% of the days in the first year contains anomalies. We would like to compare the outlier detection in this first year, for two different simulations for the second year.

- In the first simulation, we do not add any anomalies in the second year. This is the usecase we used in the paper, for which the procedure has been proposed. The objective is to compare results of the second simulations to this reference example results.

- In the second simulation, we add anomalies in 5% of the days of observations in the second year. We add both local and pattern anomalies. This correspond to an extreme case where the "clean" set contains more outlier than the set to determine.

- Then, we apply the LOF computation on the full two years of data, instead of the first one as we did on Section 3.5.1, for both simulations. The first simulation is supposed to return similar results to what has been proposed in the paper, and the objective is to compare the results on the first year in both simulations.

We represent in the figure 3.10 the LOF results for the first simulation on the PCA coefficients selected by the 2-Wasserstein test, as it was shown that it was the most powerful test in the paper. The figure 3.11 provides the same representation in the case where 5% of the days contain anomalies in the second year. As we can see, the fact that both subsets contain anomalies does not deteriorate the detection of the outliers in the first year. In fact, for both simulations, the multiple testing procedure returns approximately the same levels of coefficients. We also represent the same example on the selected Haar wavelet coefficients in Figure 3.12 and 3.13. We acknowledge similar performances on the Haar coefficients. In this application, the simulation where there are anomalies in both years enables to highlight one of the local anomalies in the first year that was missed in the first simulation where the second year was fully clean.
As we can see, even when there are a few anomalies in the set that is not supposed to contain outliers, we remark that our procedure remains really efficient.
Such results may seem surprising. In fact, in the application where both years contain anomalies, additional levels of coefficients are selected in our multiple testing procedure. For instance, in this application, the 11th principal component was selected in the second

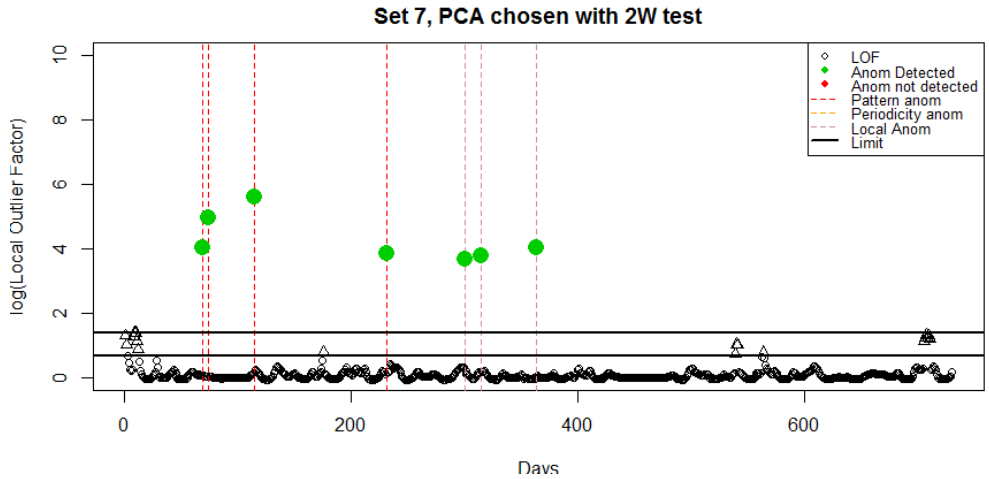

Figure 3.10 – LOF computed on the PCA coefficients selected thanks to the 2-Wasserstein test, first application where no anomalies occured in the second year.

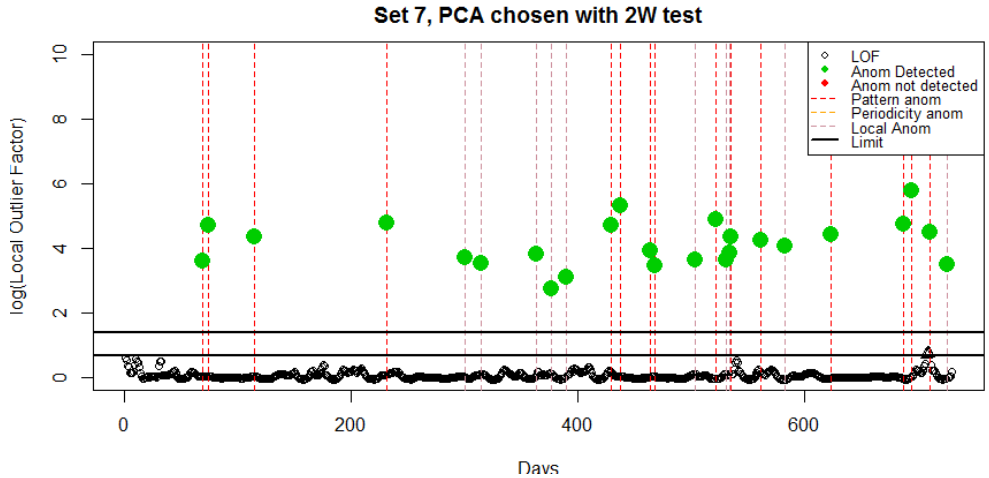

Figure 3.11 – LOF computed on the PCA coefficients selected thanks to the 2-Wasserstein test, second application where 5% of the days contain anomalies in the second year.

simulation, despite it was not selected in the first simulation. We also acknowledge the same fact with the Haar wavelet bases, with the Haar coefficients of level $\lambda = (5, 11)$ for instance.

This can be explained by the fact that, in practice, the anomalies are all different, hence generate different changes in the distributions of the coefficients. In the case where an anomaly appear identically on two years of telemetries, then it will affect the distributions of some coefficients in the same way for both years, hence the features on which this anomaly occur are unlikely to be selected. However, in practice, this situation never

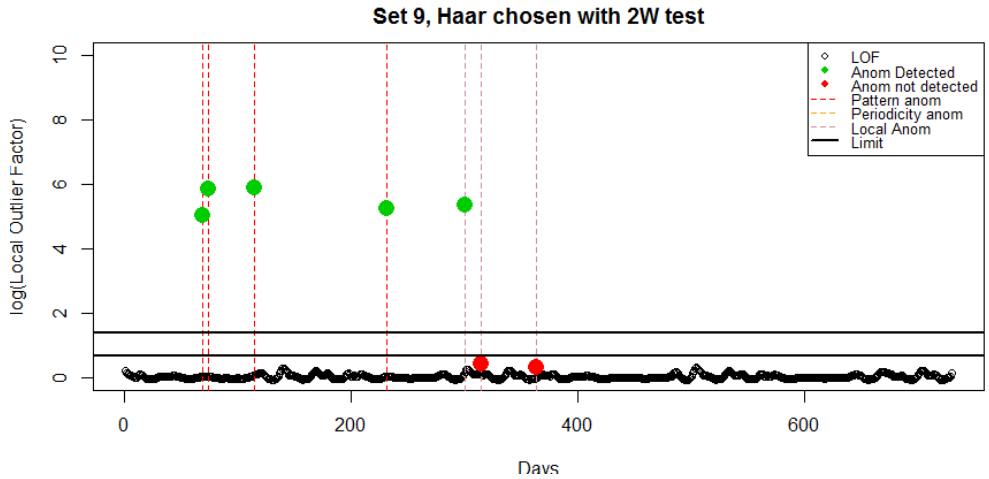

Figure 3.12 – LOF computed on the Haar coefficients selected thanks to the 2-Wasserstein test, first application where no anomalies occured in the second year.

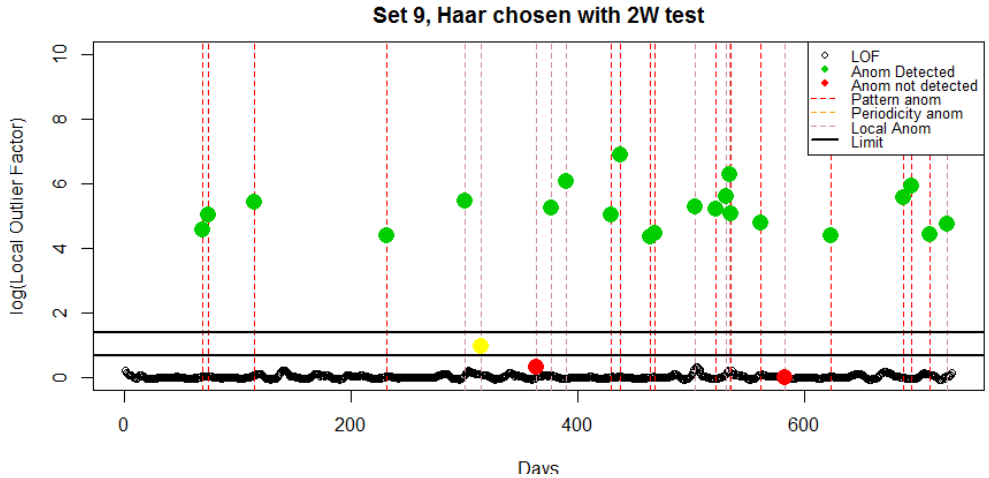

Figure 3.13 – LOF computed on the Haar coefficients selected thanks to the 2-Wasserstein test, second application where 5% of the days contains anomalies in the second year.

happens, hence we are still able to catch the coefficients on which the anomalies appear, even if the second year is contaminated by unexpected anomalies.
In our case, the anomalies are all different, hence impact the distributions of both years in different ways. Consequently, when there are outliers in both years rather than on a single year, we select additional levels for the outlier detection.
With this property, we may be able to apply this test more easily without labelling the data.

Chapter 4

Outlier Detection in Multivariate Space Telemetries based on Covariance Matrices Equality Test

Dans ce chapitre, nous nous intéressons à l'aspect multivarié des télémesures spatiales. Dans la pratique, si une réelle panne apparaît sur le satellite, il est fort probable qu'elle engendre des changements de comportement sur plusieurs télémesures. Dans ce but nous nous intéressons à la matrice de covariance entre plusieurs télémesures. Nous cherchons à mettre en lumière les périodes où des changements inattendus ont changé significativement la structure de covariance des télémesures.
Pour cela, nous appliquons trois tests statistiques, auxquels nous ajoutons un nouveau test statistique pour comparer l'égalité de deux matrices de covariance.
Nous comparons ces quatre tests sur des données simulées, des vecteurs Gaussiens et des télémesures simulées, ainsi que sur des données réelles.
Ce papier, relatant des résultats récents, est toujours un travail en cours, et fera l'objet d'un papier soumis au cours de l'année 2018. Une partie de ces résultats a d'ores-et-déjà été présentée dans le cadre de la conférence *Big Data From Space* organisée par l'ESA à Toulouse en novembre 2017, pour lequel un article court a été publié.

Outlier Detection in Multivariate Space Telemetries based on covariance matrices equality test

Clémentine Barreyre[1], Béatrice Laurent[2], Jean-Michel Loubes[3], Bertrand Cabon[1], Loïc Boussouf[1]

Abstract

In this paper, we tackle the issue of outlier detection in multivariate telemetries. For this purpose, we would like to test the hypothesis that two covariance matrices are equal, and raise an alarm as soon as there is a rejection. We introduce three existing tests and a novel statistical test. The statistical tests are firstly compared on Gaussian simulations, and then on satellite telemetry data, firstly simulated, and then on a set of real telemetry data.

4.1 Introduction

In order to monitor a satellite, thousands of telemetries are sent back to Earth almost real-timed. Most of these parameters are observed every thirty seconds, leading to consider terabytes of historical data. Some of them are followed up automatically thanks to experts knowledge, based on threshold rules under specific conditions. It is known that sometimes, abnormal events that occur in one or more telemetries may be due to a misbehavior of the satellite that must be detected. This issue has already been treated by the ESA [57], the CNES [32] and the JAXA [33]. In their framework, the anomaly detection is done only on one single telemetry on which well-chosen features are computed to highlight anomalies. However, if a real anomaly appears, it is likely to affect more than one telemetry, changing the way the telemetries are structured the ones with the others. Hence major events can sometimes generate subtle changes in several telemetries. That is why the analysis of the covariance between the telemetries is suitable to detect and to anticipate failures.

[1]Airbus Defence and Space, Z.I. Palays, 31 rue des Cosmonautes, 31400 Toulouse, France

[2]Institut de Mathématiques de Toulouse (UMR 5219), INSA Toulouse, Université de Toulouse, 135 avenue de Rangueil, 31400 Toulouse, France

[3]Institut de Mathématiques de Toulouse (UMR 5219), Université Paul Sabatier, Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse Cedex 9, France

The ESA has already dealt with the multivariate aspect with DrMUST [58] software, which was designed for investigation when it is known that a real anomaly occured.
Our approach is to develop an unsupervised anomaly detection algorithm on multivariate telemetries based on the equality-test of covariance matrices. The covariance computation handles dimension reduction when it is needed.
This paper will be structured as follows. In the first section, we introduce the mathematical model and the covariance computation, in particular in a dimension reduction framework. In the next section, we introduce four tests to compare covariance matrices. The first test was developed by Fremdt [30], the second one by Ilea [45], the third one by Cai [16] and the last one is an original test, inspired by both first tests. In this section, we introduce this test and provide its demonstration, that we consolidate with simulations. In a last section, we apply this test firstly on simulated data to compare the performances of each test, and then on a real set of twenty telemetries.

4.2 Mathematical representation

4.2.1 Assumptions

We consider the framework where the data are scarcely distinguishable in terms of intensity but yet exhibit different behaviors in their variability, which implies that the novelty can sometimes be detected only by monitoring the changes in the covariance matrix between all the telemetries. We observe m telemetries $X_1, \dots, X_m$, where their means $\mu_1, \dots, \mu_m$ are supposed to be constant with respect to a time t :

$$\mu = \mathbb{E}[X(t)] = (\mu_1, \dots, \mu_m) \in \mathbb{R}^m.$$

Hence after proper normalization we can model the observations as a sequence of i.i.d random variables with respect to a time index t,

$$X(t) = (X_1(t), \dots, X_m(t)) \in \mathbb{R}^m.$$

The time is taken as a discrete sample $t_1, \dots, t_p$. The covariance structure of the random vector $X(t)$ is given for all t by

$$\Gamma = \mathbb{E}([X(t) - \mu]^T [X(t) - \mu]) \in \mathbb{R}^{m \times m}. \tag{4.1}$$

Since the observations $\Big(X(t_k)\Big)_{1 \leq k \leq p}$ are assumed to be i.i.d, the empirical estimator is given by

$$\hat{\Gamma} = \frac{1}{p} \sum_{k=1}^{p} [X(t_k) - \bar{X}]^T [X(t_k) - \bar{X}]. \tag{4.2}$$

$$= \frac{1}{p} \sum_{k=1}^{p} X(t_k)^T X(t_k) - \bar{X}^T \bar{X} \tag{4.3}$$

where $\bar{X} = \frac{1}{p}\sum_{i=1}^{p} X(t_k)$.

We consider a Gaussian model where for all $k = 1, \ldots, p$, $\left(X(t_k) - \bar{X}\right)$ are i.i.d centered Gaussian vector with covariance Γ. In this paper, we assume that for each $k = 1, \ldots, (p-1)$,

$$X(t_k) - \bar{X} \perp\!\!\!\perp \bar{X}(t_{k+1}) - \bar{X}. \tag{4.4}$$

4.2.2 Empirical computation of the covariance matrix

The covariance $\mathbf{\Gamma}_{j,j'}$ between two telemetries X_j and $X_{j'}$ can be estimated by

$$\hat{\mathbf{\Gamma}}_{j,j'} = \frac{1}{p}\sum_{k=1}^{p}(X_{j,k} - \bar{X}_j)(X_{j',k} - \bar{X}_{j'}),$$

where

$$\bar{X}_j = \frac{1}{p}\sum_{k=1}^{p} X_{j,k}.$$

One can also denote Y_j as the centered signal $Y_j = X_j - \bar{X}_j$, for all $j = 1, ..., m$. As supposed in the equation (4.4), the $Y_j(t_1), \ldots, Y_j(t_p)$ are supposed to be i.i.d. Then, if $\boldsymbol{Y} = [Y_1, \ldots, Y_m] \in \mathbb{R}^{p\times m}$, the covariance matrix $\hat{\mathbf{\Gamma}}$ can be computed by

$$\hat{\mathbf{\Gamma}} = \frac{1}{p}\boldsymbol{Y}^T\boldsymbol{Y}. \tag{4.5}$$

In this framework, we estimate one covariance matrix per day, that we denote $\hat{\mathbf{\Gamma}}^{(i)}$, for each day $i = 1, ..., n$.
Detecting some changes in the behavior of the telemetries can be done by testing, for each day $2 \leq i \leq n$, the equality of the covariance matrices corresponding to the day i, $\mathbf{\Gamma}^{(i)}$ and $i-1$, $\mathbf{\Gamma}^{(i-1)}$. In this paper, three approaches will be introduced.

4.2.3 Reduced dimension

In some situations, it may be useful to use dimension reduction to compute the covariance, for example when it is needed to manipulate a smaller amount of data for storage issues. We observe $\left(Y_1(t_k), \ldots, Y_m(t_k)\right)_{1\leq k\leq p}$ p i.i.d centered vectors which covariance matrix is $\mathbf{\Gamma}$. As the telemetries are observations of functional data, it is possible to represent the functions by using projections onto orthonormal bases. The functional approach can be in fact really relevant as soon as the sampling of the telemetries is not regular, or when some portions of data are missing, which can be observed in real telemetry data.
Given a function $f \in \mathbb{L}^2([0,1])$, if $(\phi_\lambda)_{\lambda\in\mathbb{N}^*}$ is an orthonormal basis in $\mathbb{L}^2([0,1])$, then the function f can be decomposed in this basis,

$$f(t) = \sum_{\lambda\in\mathbb{N}^*} \theta_\lambda \phi_\lambda(t).$$

where $\theta_\lambda = \int_0^1 f(t)\phi_\lambda(t)dt$. We want to introduce such a decomposition for Y_j. We assume that $Y_j \in \mathbb{L}^2([0,1])$. Since we only observe $\left(Y_j(t_k)\right)_{1\leq k\leq p}$ we use the decomposition

$$\hat{Y}_j(t) = \sum_{\lambda\in\Lambda} \hat{\theta}_{j,\lambda}\phi_\lambda(t),$$

where $\Lambda \subset \mathbb{N}^*$ and $\hat{\theta}_{j,\lambda} = \frac{1}{p}\sum_{k=1}^p Y_j(t_k)\phi_\lambda(t_k)$. This representation enables us to reduce the dimension of the data by considering a reduced number of coefficients, leading to select the levels $\lambda \in \Lambda = \{1, ..., q\}$, where $q \leq p$ is the number of components to retain.
Dimension reduction using projections is really common when we deal with functional data, see for example [4], [6], where many orthonormal bases are tested, for a clustering application.
Let $\mathbf{\Phi}_q \in \mathbb{R}^{p\times q}$ be the matrix representation of the q first functions of the functional basis, where $(\mathbf{\Phi}_q)_{\lambda,k} = \phi_\lambda(t_k)$, and $\hat{\mathbf{\Theta}}_q \in \mathbb{R}^{m\times q}$ is the matrix of the estimated coefficients of $\boldsymbol{Y}$ in this basis, where $(\hat{\mathbf{\Theta}}_q)_{j,\lambda} = \hat{\theta}_{j,\lambda}$. Consequently, we can approach $\boldsymbol{Y}$ by $\tilde{\boldsymbol{Y}}_q = \mathbf{\Phi}_q\hat{\mathbf{\Theta}}_q^T$. The approached covariance matrix $\tilde{\mathbf{\Gamma}}_q$ can be computed by applying the equation (4.5) to $\tilde{\boldsymbol{Y}}_q$, and if $(\phi_\lambda)_{\lambda\in\mathbb{N}^*}$ is orthonormal in $\mathbb{R}^p$, then $\mathbf{\Phi}_q^T\mathbf{\Phi}_q = I_q$, and we get

$$\tilde{\mathbf{\Gamma}}_q = \frac{1}{p}\hat{\mathbf{\Theta}}_q\hat{\mathbf{\Theta}}_q^T. \tag{4.6}$$

This property enables to use dimension reduction for computing the covariance.

4.2.4 Example of covariance computation by reducing the dimension

We consider a set of $m = 50$ simulated curves sampled on 256 points. We choose to reduce the dimension of the curves thanks to a Haar wavelet basis, as this basis is orthonormal both in $\mathbb{L}^2([0,1])$ and $\mathbb{R}^p$. Let us first recall its definition. We set $\psi = \mathbb{1}_{[0,1/2[} - \mathbb{1}_{[1/2,1[}$. For all $l \geq 0$, $k \in \Lambda(l) = \{0, 1, \ldots, 2^l - 1\}$, let $\phi_{l,k}(x) = 2^{l/2}\psi(2^l x - k)$. The functions $(\phi_0, \phi_{l,k}, l \geq 0, k \in \Lambda(l))$ form the orthonormal Haar basis of $\mathbb{L}^2([0,1])$. The index l is recalled as the scale, and k as the position.
For this application, we retain the $(2^{L+1} - 1)$ levels corresponding to the wavelet levels for which $0 \leq l \leq L$, for a fixed upper scale L. If we represent the observations before and after dimension reduction, with $L = 4$ as upper scale, we obtain the results represented in Figure 4.1. In order to get the sensitivity of the dimension reduction on the computation of the covariance matrix, we compute the empirical covariance matrix from the raw-data $\hat{\mathbf{\Gamma}}$ by using the equation (4.5), then we compute $\tilde{\mathbf{\Gamma}}_L$ by using the dimension reduction from the equation (4.6). We estimate the relative error ϵ_L between the two covariance matrices by

$$\epsilon_L = \sqrt{\frac{1}{m^2}\sum_{j,j'=1}^m \left(\frac{\hat{\mathbf{\Gamma}}_{j,j'} - (\tilde{\mathbf{\Gamma}}_L)_{j,j'}}{\hat{\mathbf{\Gamma}}_{j,j'}}\right)^2}.$$

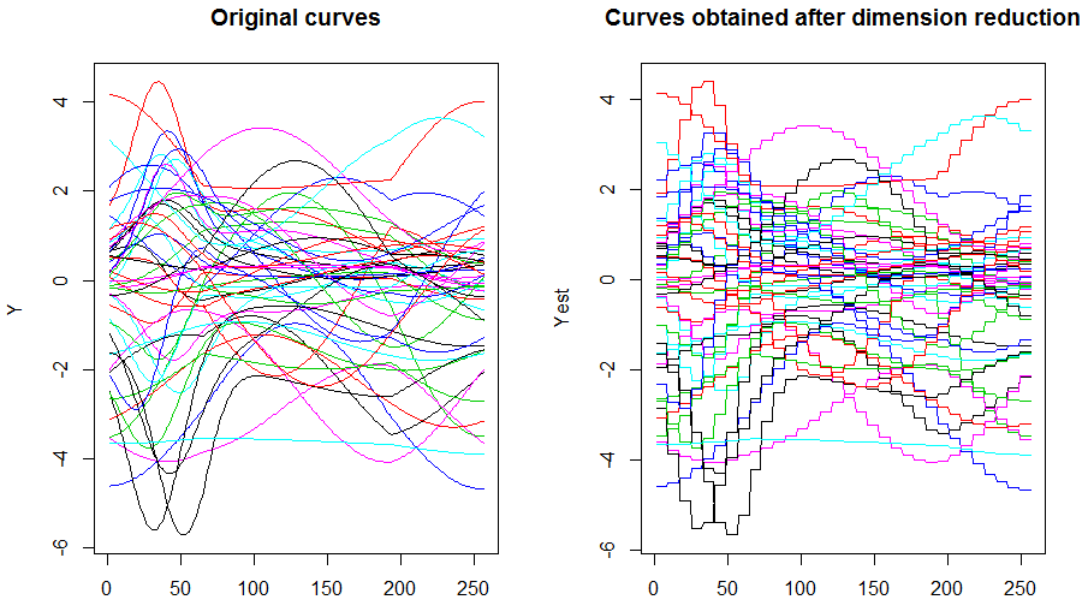

Figure 4.1 – Dimension reduction with Haar wavelets, where only the levels up to $L = 4$ are selected. Initial functions (on the left) are sampled on $p = 256$ time stamps, whereas the reduced dimension curves (on the right) are obtained from q =31 coefficients.

From the example presented in figure 4.1, we get $\epsilon_4 = 0.052$, which represents a relative error which is really small. With $L = 5$, then the error decreases to $\epsilon_5 = 0.013$, whereas $\epsilon_3 = 0.2$ for $L = 3$. For this application we should keep $L \geq 4$ to keep an error lower than 5%. The covariance computed after dimension reduction is really close to the covariance computed on the raw-data as soon as the dimension reduction is not too hard.
In the next section, we introduce several approaches to test the equality of two covariance matrices.

4.3 Covariance equality test

Suppose we have only two days $\boldsymbol{Y}^{(1)} \in \mathbb{R}^{p\times m}$ and $\boldsymbol{Y}^{(2)} \in \mathbb{R}^{p\times m}$ of a zero-mean daily-repetitive telemetry that we want to compare, under the hypothesis that the instants $\boldsymbol{Y}^{(1)}_{\cdot,k}$, $k = 1, \ldots, p$ are i.i.d, as it was introduced in Section 4.2.1. The repetiveness assumption is justified by the natural periodicity of the geostationary satellites. We would like to test the following hypothesis :

$$H_0 : \{\boldsymbol{\Gamma}^{(1)} = \boldsymbol{\Gamma}^{(2)}\}. \tag{4.7}$$

For this purpose, we will apply three tests that already exist, and an original one that will be introduced later in this section. The two first tests consist in comparing the eigenvectors of both matrices, as for the test we have developed. The third test is built on the difference of both matrices. We will challenge the two approaches with simulations.
For all the tests introduced in this section, $\hat{\mathbf{\Gamma}}^{(1)}$ and $\hat{\mathbf{\Gamma}}^{(2)}$ indicate the empirical covariance matrices $\hat{\mathbf{\Gamma}}^{(1)} = (\boldsymbol{Y}^{(1)})^T\boldsymbol{Y}^{(1)}/p$ and $\hat{\mathbf{\Gamma}}^{(2)} = (\boldsymbol{Y}^{(2)})^T\boldsymbol{Y}^{(2)}/p$.

4.3.1 Fremdt test

The first test is a non-parametric test introduced by Fremdt et al. [30] that was primary defined to test the equality of the covariance structures in two functional samples. The authors noticed that testing the equality of the covariance is equivalent to test that both samples have the same functional principal components. We adapted it to our problem, where the functional samples here are two days of periodical telemetries. The test consists into the following steps.

- Let us consider a reference matrix $\mathbf{\Gamma}$. It can be for instance the mean matrix between the two considered covariance matrices. In practice, $\mathbf{\Gamma}$ is estimated by $\hat{\mathbf{\Gamma}}$

$$\hat{\mathbf{\Gamma}} = \frac{1}{2}(\hat{\mathbf{\Gamma}}^{(1)} + \hat{\mathbf{\Gamma}}^{(2)}).$$

  One can also consider the covariance matrix computed on the barycenter distribution in the Wasserstein space, as it is proposed by Le Gouic et al. [50]. We assume that $\mathbf{\Gamma}$ admits the SVD decomposition $\mathbf{\Gamma} = \boldsymbol{\varphi}\mathbf{\Lambda}\boldsymbol{\varphi}^T$, where $\Lambda_{i,j} = \lambda_i \times \mathbb{1}_{i=j}$ are the eigenvalues if $i = j$. In practice, as the real covariance matrix $\mathbf{\Gamma}$ is unknown, we apply this decomposition to $\hat{\mathbf{\Gamma}}$, getting $\hat{\mathbf{\Gamma}} = \hat{\boldsymbol{\varphi}}\hat{\mathbf{\Lambda}}\hat{\boldsymbol{\varphi}}^T$.

- Let $\hat{\boldsymbol{\varphi}}_d = (\hat{\boldsymbol{\varphi}}_{\cdot,\lambda})_{\lambda=1,\dots,d}$, be the matrix containing the d first ordered eigenvectors of $\hat{\boldsymbol{\varphi}}$, and $\hat{a}^{(1)}$, $\hat{a}^{(2)} \in \mathbb{R}^{p\times d}$ be the coefficients obtained by projecting the telemetries from both days in this family. They can be computed by

$$\hat{a}^{(1)} = \boldsymbol{Y}^{(1)}\hat{\boldsymbol{\varphi}}_d \text{ and } \hat{a}^{(2)} = \boldsymbol{Y}^{(2)}\hat{\boldsymbol{\varphi}}_d.$$

  They are the d first principal components, and we assume that the eigenvalues $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_d)$ are strictly decreasing.

- The matrix $\mathbf{\Lambda} = \boldsymbol{\varphi}^T\mathbf{\Gamma}\boldsymbol{\varphi} = \boldsymbol{\lambda}I_d$ is the eigenvalues matrix, hence it is diagonal. Under H_0, $\mathbf{\Gamma}^{(1)} = \mathbf{\Gamma}^{(2)} = \mathbf{\Gamma}$, hence $\mathbf{\Lambda} = \mathbf{\Lambda}^{(1)} = \mathbf{\Lambda}^{(2)}$, where $\mathbf{\Lambda}^{(1)} = \boldsymbol{\varphi}^T\mathbf{\Gamma}^{(1)}\boldsymbol{\varphi}$ and $\mathbf{\Lambda}^{(2)} = \boldsymbol{\varphi}^T\mathbf{\Gamma}^{(2)}\boldsymbol{\varphi}$. We estimate the first d eigenvalues matrices $\mathbf{\Lambda}^{(1)}$ and $\mathbf{\Lambda}^{(2)}$ by

$$\hat{\mathbf{\Delta}}_d^{(1)} = \hat{\boldsymbol{\varphi}}_d^T\hat{\mathbf{\Gamma}}^{(1)}\hat{\boldsymbol{\varphi}}_d = \frac{1}{p}(\hat{a}^{(1)})^T\hat{a}^{(1)} \in \mathbb{R}^{d\times d}.$$

$$\hat{\mathbf{\Delta}}_d^{(2)} = \hat{\boldsymbol{\varphi}}_d^T\hat{\mathbf{\Gamma}}^{(2)}\hat{\boldsymbol{\varphi}}_d = \frac{1}{p}(\hat{a}^{(2)})^T\hat{a}^{(2)} \in \mathbb{R}^{d\times d}.$$

- Denote $\xi = \text{Vech}(\hat{\mathbf{\Delta}}_d^{(2)} - \hat{\mathbf{\Delta}}_d^{(1)})$, which is the vectorised version of the matrix $(\hat{\mathbf{\Delta}}_d^{(2)} - \hat{\mathbf{\Delta}}_d^{(1)})$ on which only the indices corresponding to the upper triangle matrix are included. Consequently, we have $\xi \in \mathbb{R}^{\frac{d(d+1)}{2}}$. The hypothesis (4.7) is then equivalent to the hypothesis $H_0 : \{\xi = 0\}$.

- Denote $L \in \mathbb{R}^{\frac{d(d+1)}{2} \times \frac{d(d+1)}{2}}$ as the covariance matrix of ξ. The computation details are provided in [30].

- Then, it can be shown thanks to [30] that, under the null hypothesis H_0 the statistics

$$T_1 = p\xi^T L^{-1} \xi \xrightarrow[p \to +\infty]{\mathcal{D}} \chi^2_{\frac{d(d+1)}{2}}$$

- The hypothesis is then rejected at the level α if

$$T_1 \geq \chi^2_{d(d+1)/2,1-\alpha}$$

  where $\chi^2_{d(d+1)/2,1-\alpha}$ is the $1-\alpha$ level of a chi-squared distribution with $\frac{d(d+1)}{2}$ levels of freedom.

In our application, we apply this test every two-consecutive days in order to catch the daily changes in the covariance structure.

4.3.2 Ilea test

According to another paper proposed by Ilea [45], if $\boldsymbol{Y}^{(1)} \in \mathbb{R}^{p \times m}$ and $\boldsymbol{Y}^{(2)} \in \mathbb{R}^{p \times m}$ are two zero-mean Gaussian distributions, then the Rao geodesic distance between their two covariance matrices can be computed according to

$$GD(\hat{\mathbf{\Gamma}}^{(1)} || \hat{\mathbf{\Gamma}}^{(2)}) = \left(\frac{1}{2} \sum_{i=1}^{m} (\ln(\lambda_i^*))^2 \right)^{\frac{1}{2}},$$

where $\lambda_i^*, i = 1, .., m$ are the principal components of the matrix

$$\boldsymbol{M} = (\hat{\mathbf{\Gamma}}^{(2)})^{-1} \hat{\mathbf{\Gamma}}^{(1)} \tag{4.8}$$

According to the same paper, it can be shown that the statistics

$$T_2 = \frac{n}{4} \left(\sum_{i=1}^{m} (\ln(\lambda_i^*))^2 \right) \xrightarrow[p \to +\infty]{\mathcal{D}} \chi^2 \left(\frac{m(m+1)}{2} \right). \tag{4.9}$$

The hypothesis is then rejected at a level α if

$$T_2 \geq \chi^2_{m(m+1)/2,1-\alpha}$$

where $\chi^2_{m(m+1)/2,1-\alpha}$ is the $1-\alpha$ level of a chi-squared distribution with $\frac{m(m+1)}{2}$ levels of freedom.
In our application, we apply this test every two-consecutive days, as for the test developed by Fremdt.

Remark If $\hat{\mathbf{\Gamma}}^{(2)}$ is ill-conditioned, for example if some of its eigenvalues are really close to zero, then the computation of $\boldsymbol{M}$ is compromised because the invert of $\hat{\mathbf{\Gamma}}^{(2)}$ is numerically unstable. Consequently the test may be not applicable. It is also possible to take a reduced number of eigenvalues $d < m$ in the test (4.9). However, unlike the previous test, it is not a solution when some of the eigenvalues of $\hat{\mathbf{\Gamma}}^{(2)}$ are really close to zero, because the eigenvalues λ_i^*, $i = 1, ..., m$ are the eigenvalues of the matrix $\boldsymbol{M}$, nor $\hat{\mathbf{\Gamma}}^{(2)}$. Consequently, $\hat{\mathbf{\Gamma}}^{(2)}$ has always to be inverted in the first place.

4.3.3 Cai test

Cai et al. [16] proposed a test in another framework which is designed to be powerful against sparse alternatives. The test statistics is built on the maximum of the differences between the two covariance matrices. In addition to test the equality of the covariance matrices, it highlights the couple of telemetries that has changed the most. In the same framework as the Fremdt test where we consider two the covariance between m random variables, we compute the following test statistics

$$T_3 = \max_{1\leq i\leq j\leq m} p \times \left(\frac{(\hat{\Gamma}^{(1)}_{i,j} - \hat{\Gamma}^{(2)}_{i,j})^2}{\hat{\beta}^{(1)}_{i,j} + \hat{\beta}^{(2)}_{i,j}} \right), \tag{4.10}$$

where $\hat{\beta}^{(1)}_{i,j}$ (resp. $\hat{\beta}^{(2)}_{i,j}$) is an estimation of the variance of $\boldsymbol{Y}_i^{(1)T}\boldsymbol{Y}_j^{(1)}$ (resp. $\boldsymbol{Y}_i^{(2)T}\boldsymbol{Y}_j^{(2)}$) defined according to

$$\hat{\beta}^{(1)}_{i,j} = \frac{1}{p}\sum_{k=1}^{p}(Y^{(1)}_{k,i}Y^{(1)}_{k,j} - \hat{\Gamma}^{(1)}_{i,j})^2$$

$$\hat{\beta}^{(2)}_{i,j} = \frac{1}{p}\sum_{k=1}^{p}(Y^{(2)}_{k,i}Y^{(2)}_{k,j} - \hat{\Gamma}^{(2)}_{i,j})^2.$$

As demonstrated in Cai [16], the statistics T_3 is closely linked to a type-I extreme-value distribution, leading to reject the hypothesis H_0 at the level α when

$$T_3 \geq -\ln(8\pi) - 2\ln\left(\ln\left(\frac{1}{1-\alpha}\right)\right) + 4\ln(m) - \ln\Big(\ln(m)\Big).$$

4.3.4 New test

4.3.4.1 Definition of the test

We propose a new test which is inspired by the two first tests. It is easier to set up and faster to compute and to run on simulations. The test uses the following notations.

- Assume that $\boldsymbol{Y}^{(1)} \in \mathbb{R}^{p\times m}$ and $\boldsymbol{Y}^{(2)} \in \mathbb{R}^{p\times m}$ are two independent zero-mean Gaussian distributions, which covariance matrix are respectively $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$.
- As the true covariance matrices are unknown, they are estimated by their empirical covariance $\hat{\boldsymbol{\Gamma}}^{(1)}$ and $\hat{\boldsymbol{\Gamma}}^{(2)}$, as defined in the equation (4.5).
- Consider $\boldsymbol{\Gamma}$ as a reference matrix resuming the information of both covariance matrices $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$, which can be the barycenter matrix of $\boldsymbol{\Gamma}^{(1)}$ and $\boldsymbol{\Gamma}^{(2)}$ or their mean, for example.
- $\boldsymbol{\Gamma}$ can be expressed by its SVD decomposition $\boldsymbol{\Gamma} = \boldsymbol{\varphi}\boldsymbol{\Lambda}\boldsymbol{\varphi}^T$, where the d first eigenvalues $\Lambda_{i,i} = \lambda_i, i = 1, ..., d$ are supposed to be strictly decreasing.
- In practice, $\boldsymbol{\Gamma}$ is estimated by $\hat{\boldsymbol{\Gamma}}$, as for the Fremdt test, leading to consider estimations of its SVD decomposition $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\varphi}}^T$.
- Denote $\hat{\boldsymbol{\Delta}}_1 = \hat{\boldsymbol{\varphi}}^T\hat{\boldsymbol{\Gamma}}^{(1)}\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{\Delta}}_2 = \hat{\boldsymbol{\varphi}}^T\hat{\boldsymbol{\Gamma}}^{(2)}\hat{\boldsymbol{\varphi}}$ as the covariance matrices in the $\hat{\boldsymbol{\varphi}}$ basis. Their diagonal terms are denoted $\hat{\delta}_i^{(1)} = (\hat{\boldsymbol{\Delta}}_1)_{i,i}$ and $\hat{\delta}_i^{(2)} = (\hat{\boldsymbol{\Delta}}_2)_{i,i}$.

The test consists into the following proposition.

Proposition 2. *Under* $H_0 : \{\boldsymbol{\Gamma}^{(1)} = \boldsymbol{\Gamma}^{(2)} = \boldsymbol{\Gamma}\}$,

$$T_4 = \frac{p}{4}\left(\sum_{i=1}^{d}\left[ln\left(\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}}\right)\right]^2\right) \underset{p\to+\infty}{\overset{\mathcal{D}}{\longrightarrow}} \chi^2(d), \tag{4.11}$$

where $d \leq m$ *is the number of components to keep. Consequently, for large values of* p*, and if* d *is fixed and* $p \to +\infty$*, we can reject the hypothesis* H_0 *at the level* α *if*

$$T_4 \geq \chi^2_{d,1-\alpha}$$

where $\chi^2_{d,1-\alpha}$ *is the* $1-\alpha$ *level of a chi-squared distribution with* d *levels of freedom.*

Proof. Consider the eigenvalues decompositions for both matrices $\boldsymbol{\Gamma}^{(1)} = \boldsymbol{\varphi}_1\boldsymbol{\Lambda}_1\boldsymbol{\varphi}_1^T$ and $\boldsymbol{\Gamma}^{(2)} = \boldsymbol{\varphi}_2\boldsymbol{\Lambda}_2\boldsymbol{\varphi}_2^T$.

Under H_0, we have $\boldsymbol{\varphi}_1 = \boldsymbol{\varphi}_2 = \boldsymbol{\varphi}$. Consequently, $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ can both be estimated by $\hat{\boldsymbol{\varphi}}$, and the matrix $\boldsymbol{M}$ defined in (4.8) can be estimated by

$$\hat{\boldsymbol{M}} = (\hat{\boldsymbol{\Gamma}}^{(2)})^{-1}\hat{\boldsymbol{\Gamma}}^{(1)} \tag{4.12}$$

$$= (\hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\Delta}}_2\hat{\boldsymbol{\varphi}}^T)^{-1}(\hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\Delta}}_1\hat{\boldsymbol{\varphi}}^T) \tag{4.13}$$

$$= \hat{\boldsymbol{\varphi}}^T\hat{\boldsymbol{\Delta}}_2^{-1}\hat{\boldsymbol{\Delta}}_1\hat{\boldsymbol{\varphi}}. \tag{4.14}$$

From Ilea test, we can conclude that the eigenvalues λ^* of the matrix $\boldsymbol{M}$ can be estimated by $\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}}$ in this framework, even if the matrices $\hat{\boldsymbol{\Delta}}_2^{-1}$ and $\hat{\boldsymbol{\Delta}}_1$ are not diagonal.
In addition, we know that $\hat{\boldsymbol{\Gamma}}^{(1)}$ and $\hat{\boldsymbol{\Gamma}}^{(2)}$ have a Wishart distribution. As both sets are Gaussian and independent, we know thanks to the Cochran's theorem that, for each $z \in \mathbb{R}^p$, $z^T\hat{\boldsymbol{\Gamma}}^{(1)}z = (\boldsymbol{Y}^{(1)}z)^T(\boldsymbol{Y}^{(1)}z)$ and $z^T\hat{\boldsymbol{\Gamma}}^{(2)}z = (\boldsymbol{Y}^{(2)}z)^T(\boldsymbol{Y}^{(2)}z)$ are independent. Consequently, for each vector $z \in \mathbb{R}^p$,

$$\frac{z^T\hat{\boldsymbol{\Gamma}}^{(1)}z}{z^T\hat{\boldsymbol{\Gamma}}^{(2)}z} \sim \mathcal{F}(p,p).$$

If $z_i = \boldsymbol{\varphi}_{\cdot,i}$, then we get

$$\frac{z_i^T\hat{\boldsymbol{\Gamma}}^{(1)}z_i}{z_i^T\hat{\boldsymbol{\Gamma}}^{(2)}z_i} = \frac{\tilde{\delta}_i^{(1)}}{\tilde{\delta}_i^{(2)}} \sim \mathcal{F}(p,p)$$

where $\tilde{\delta}_i^{(1)}$ and $\tilde{\delta}_i^{(2)}$ are the diagonal terms of the matrices $\tilde{\boldsymbol{\Delta}}_1$ and $\tilde{\boldsymbol{\Delta}}_2$ such that

$$\tilde{\boldsymbol{\Delta}}_1 = \boldsymbol{\varphi}^T\hat{\boldsymbol{\Gamma}}^{(1)}\boldsymbol{\varphi}$$

and resp. for $\tilde{\boldsymbol{\Delta}}_2$. In practice, as the eigenvectors $\boldsymbol{\varphi}_{\cdot,i}, i = 1, ..., m$ of the true covariance matrix are unknown, the random variables $\tilde{\delta}_i^{(1)}$ and $\tilde{\delta}_i^{(2)}$ are unknown as well. However, if $\hat{z}_i = \hat{\boldsymbol{\varphi}}_{\cdot,i}$ is the i^{th} eigenvector of the matrix $\hat{\boldsymbol{\Gamma}}$ then $\hat{z}_i^T\hat{\boldsymbol{\Gamma}}^{(1)}\hat{z}_i = \hat{\delta}_i^{(1)}$. Besides from the results of Lemma 1, we have $\hat{\delta}_i^{(1)} = \tilde{\delta}_i^{(1)}(1+\epsilon_i') + O_P\left(\frac{1}{p}\right)$, and resp. for $\hat{\delta}_i^{(2)}$, where $\epsilon_i' = 2(\epsilon_i - 1)$, $c_i = sign\langle\hat{\varphi}_i, \varphi_i\rangle$ and $\epsilon_i = \hat{\varphi}_i - c_i\varphi_i = O_P(\frac{1}{\sqrt{p}})$. Therefore,

$$\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}} = \frac{\tilde{\delta}_i^{(1)}}{\tilde{\delta}_i^{(2)}} \times \frac{1+\epsilon_i' + O_P(\frac{1}{p})}{1+\epsilon_i' + O_P(\frac{1}{p})} \tag{4.15}$$

As demonstrated in [47], it can be shown that if $\frac{\tilde{\delta}_i^{(1)}}{\tilde{\delta}_i^{(2)}} \sim \mathcal{F}(p,p)$, then

$$\tilde{L}_i = \ln\left(\frac{\tilde{\delta}_i^{(1)}}{\tilde{\delta}_i^{(2)}}\right) \sim \text{III-GL}(\frac{p}{2}),$$

where III-GL($\frac{p}{2}$) denotes a Type-III Generalized Logistic distribution, of parameter $p/2$. This distribution is also a log-Fisher distribution.

In practice, $\tilde{L}_i$ is unknown since $\tilde{\delta}_i^{(1)}$ and $\tilde{\delta}_i^{(2)}$ are unknown. Instead we use the statistics $\hat{L}_i$, that we build thanks to the results from the equation (4.15).

$$\begin{aligned}
\hat{L}_i &= \ln\left(\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}}\right) \\
&= \ln\left(\frac{\tilde{\delta}_i^{(1)}}{\tilde{\delta}_i^{(2)}}\right) + \ln\left(\frac{1+\epsilon_i' + O_P(\frac{1}{p})}{1+\epsilon_i' + O_P(\frac{1}{p})}\right) \\
&= \tilde{L}_i + \ln\Big(1+\epsilon_i' + O_P(\frac{1}{p})\Big) - \ln\Big(1+\epsilon_i' + O_P(\frac{1}{p})\Big) \\
&= \tilde{L}_i + \epsilon_i' - \epsilon_i' + O_P(\frac{1}{p}) \\
&= \tilde{L}_i + O_P(\frac{1}{p})
\end{aligned}$$

Furthermore, thanks to Aroian [5], as $\tilde{L}_i$ has a log-Fisher distribution, we know that

$$\frac{\sqrt{p}}{2}\tilde{L}_i \xrightarrow[p\to+\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

for all $i = 1, ..., m$. Hence,

$$\begin{aligned}
\frac{\sqrt{p}}{2}\hat{L}_i &= \frac{\sqrt{p}}{2}\Big(\tilde{L}_i + O_P(\frac{1}{p})\Big) \\
&= \frac{\sqrt{p}}{2}\tilde{L}_i + O_P\Big(\frac{1}{\sqrt{p}}\Big) \\
&\xrightarrow[p\to+\infty]{\mathcal{D}} \mathcal{N}(0,1)
\end{aligned}$$

Moreover, as the initial data are Gaussian random variables and the eigenvectors $(\varphi_i)_{i=1,...,p}$ are orthonormal, then $\tilde{\delta}_i^{(1)} \perp\!\!\!\perp \tilde{\delta}_j^{(1)}$ and $\tilde{\delta}_i^{(2)} \perp\!\!\!\perp \tilde{\delta}_j^{(2)}$, for all $i \neq j$. Consequently, the statistics $(\tilde{L}_i)_{i=1,...,d}$ are independent as well. Hence we have that $\sum_{i=1}^{d}(\frac{\sqrt{p}}{2}\tilde{L}_i)^2 \xrightarrow[p\to+\infty]{\mathcal{D}} \chi^2(d)$, and

$$T_4 = \sum_{i=1}^{d}\left(\frac{\sqrt{p}}{2}\hat{L}_i\right)^2 \tag{4.16}$$

$$= \sum_{i=1}^{d}\left(\frac{\sqrt{p}}{2}\tilde{L}_i\right)^2 + O_P\Big(\frac{1}{\sqrt{p}}\Big). \tag{4.17}$$

As d is fixed, we have, as $p \to +\infty$

$$T_4 = \sum_{i=1}^{d}\left(\frac{\sqrt{p}}{2}\hat{L}_i\right)^2 = \sum_{i=1}^{d}\frac{p}{4}\left[\ln\left(\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}}\right)\right]^2 \xrightarrow[p\to+\infty]{\mathcal{D}} \chi^2(d).$$

□

Lemma 1. *In this lemma, we use the same notations as the ones defined for the Proposition 2.*
The empirical covariance matrices $\hat{\mathbf{\Gamma}}^{(1)}$ *and* $\hat{\mathbf{\Gamma}}^{(2)}$ *can be written as an expansion into the real eigenvectors basis* $\boldsymbol{\varphi}$ *and its estimation* $\hat{\boldsymbol{\varphi}}$ *according to*

$$\hat{\mathbf{\Gamma}}^{(1)} = \hat{\boldsymbol{\varphi}}\hat{\mathbf{\Delta}}_1\hat{\boldsymbol{\varphi}}^T = \boldsymbol{\varphi}\tilde{\mathbf{\Delta}}_1\boldsymbol{\varphi} \tag{4.18}$$

$$\hat{\mathbf{\Gamma}}^{(2)} = \hat{\boldsymbol{\varphi}}\hat{\mathbf{\Delta}}_2\hat{\boldsymbol{\varphi}}^T = \boldsymbol{\varphi}\tilde{\mathbf{\Delta}}_2\boldsymbol{\varphi} \tag{4.19}$$

For each $k \in \{1,2\}$, *both matrices* $\hat{\mathbf{\Delta}}_k = (\hat{\delta}_{i,j}^{(k)})_{i,j=1,...,m}$ *and* $\tilde{\mathbf{\Delta}}_k = (\tilde{\delta}_{i,j}^{(k)})_{i,j=1,...,m}$ *are symmetric. We denote* $\hat{\delta}_i^{(k)} = \hat{\delta}_{i,i}^{(k)}$ *and* $\tilde{\delta}_i^{(k)} = \tilde{\delta}_{i,i}^{(k)}$ *as the diagonal terms of both matrices. Then, for large values of* p, *it can be shown that, for all* $i = 1, ..., m$

$$\hat{\delta}_i^{(k)} = \tilde{\delta}_i^{(k)}(1 + \epsilon_i' + O_P\Big(\frac{1}{p}\Big)).$$

where $\epsilon_i' = 2(\epsilon_i - 1)$, $\epsilon_i = \hat{\varphi}_i - c_i\varphi_i = O_P\Big(\frac{1}{\sqrt{p}}\Big)$ *and* $c_i = \langle\varphi_i, \hat{\varphi}_i\rangle$.

Proof. If the eigenvalues $\lambda_i, i = 1, ..., m$ of the matrix $\mathbf{\Gamma}$ are strictly decreasing, then, according to Bosq [13], we know that for all $i = 1, ..., m$ $||\varphi_i - c_i\hat{\varphi}_i|| = O_P\Big(\frac{1}{\sqrt{p}}\Big)$, where $\varphi_i = \boldsymbol{\varphi}_{\cdot,i}$ and $\hat{\varphi}_i = \hat{\boldsymbol{\varphi}}_{\cdot,i}$, $c_i = \text{sign}(\langle\hat{\varphi}_i, \varphi_i\rangle)$ such that $c_i^2 = 1$ and $||\cdot||$ denotes the 2-norm. Consequently, for all $i = 1, ..., d$, there exist a vector $\epsilon_i \in \mathbb{R}^m$, where $||\epsilon_i|| = O_P\Big(\frac{1}{\sqrt{p}}\Big)$ for which $\hat{\varphi}_i = c_i\varphi_i + \epsilon_i$. Consequently, for $k \in \{1,2\}$,

$$\begin{aligned}
\hat{\delta}_i^{(k)} &= \hat{\varphi}_i^T\hat{\mathbf{\Gamma}}^{(k)}\hat{\varphi}_i && (4.20)\\
&= (c_i\varphi_i + \epsilon_i)^T\hat{\mathbf{\Gamma}}^{(k)}(c_i\varphi_i + \epsilon_i) && (4.21)\\
&= \varphi_i^T\hat{\mathbf{\Gamma}}^{(k)}\varphi_i + c_i\varphi_i^T\hat{\mathbf{\Gamma}}^{(k)}\epsilon_i + c_i\epsilon_i^T\hat{\mathbf{\Gamma}}^{(k)}\varphi_i + \epsilon_i^T\hat{\mathbf{\Gamma}}^{(k)}\epsilon_i && (4.22)\\
&= \varphi_i^T\boldsymbol{\varphi}\tilde{\mathbf{\Delta}}_k\boldsymbol{\varphi}^T\varphi_i + 2c_i\varphi_i^T\boldsymbol{\varphi}\tilde{\mathbf{\Delta}}_k\boldsymbol{\varphi}^T\epsilon_i + \sum_{j=1}^m\sum_{l=1}^m\epsilon_{i,l}\epsilon_{i,j}\hat{\Gamma}_{l,j}^{(k)} && (4.23)\\
&= \tilde{\delta}_i^{(k)} + 2c_i\sum_{j=1}^m\sum_{l=1}^m\tilde{\delta}_{i,l}^{(k)}\varphi_{l,j}\epsilon_{i,j} + O_P\Big(\frac{1}{p}\Big) && (4.24)\\
&= \tilde{\delta}_i^{(k)} + 2c_i\tilde{\delta}_i^{(k)}\sum_{j=1}^m\varphi_{i,j}\epsilon_{i,j} + 2c_i\sum_{j=1}^m\sum_{l\neq i=1}^m\tilde{\delta}_{i,l}^{(k)}\varphi_{l,j}\epsilon_{i,j} + O_P\Big(\frac{1}{p}\Big) && (4.25)
\end{aligned}$$

Denote $|||.|||$ as a matrix norm, for example the 2-matrix-norm, defined as follows. If $\boldsymbol{M} \in \mathbb{R}^{m\times n}$

$$|||\boldsymbol{M}||| = \max_{j=1,..,n}\sqrt{\sum_{i=1}^m M_{i,j}^2}.$$

We know from the central limit theorem that under H_0, $|||\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}^{(k)}||| = O_P\Big(\frac{1}{\sqrt{p}}\Big)$, leading to have

$$\begin{aligned}
|||\mathbf{\Lambda} - \tilde{\mathbf{\Delta}}_k||| &= |||\boldsymbol{\varphi}^T(\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}^{(k)})\boldsymbol{\varphi}||| \\
&\leq |||\boldsymbol{\varphi}^T||| \times |||\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}^{(k)}||| \times |||\boldsymbol{\varphi}||| \\
&= |||\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}^{(k)}||| \\
&= O_P\Big(\frac{1}{\sqrt{p}}\Big)
\end{aligned}$$

as $|||\boldsymbol{\varphi}||| = |||\boldsymbol{\varphi}^T||| = 1$. Consequently, as $\Lambda_{i,l} = \lambda_i \times \mathbb{1}_{i=l}$ for all $i, j = 1, ..., m$, then for each $k \in \{1, 2\}$ there exist a vector $\eta_{i,l}^{(k)} \in \mathbb{R}^m$ such that

$$\tilde{\delta}_{i,l}^{(k)} = \Lambda_{i,l} + \eta_{i,l}^{(k)} \tag{4.26}$$

$$= \lambda_i \times \mathbb{1}_{i=l} + \eta_{i,l}^{(k)}, \tag{4.27}$$

with $\eta_{i,j}^{(k)} = O_P\Big(\frac{1}{\sqrt{p}}\Big)$ for both $k \in \{1, 2\}$. In the equation (4.25), we have split the sum in two by isolating the case where $l = i$ to the case where $l \neq i$. Then, we can replace $\tilde{\delta}_{i,l}^{(k)}$ in the sum where $l \neq i$ by $\eta_{i,j}^{(k)}$ thanks to the results of (4.27).

$$\hat{\delta}_i^{(k)} = \tilde{\delta}_i^{(k)} + 2c_i\tilde{\delta}_i^{(k)}\sum_{j=1}^{m}\varphi_{i,j}\epsilon_{i,j} + 2c_i\sum_{j=1}^{m}\sum_{l\neq i=1}^{m}\eta_{i,l}^{(k)}\varphi_{j,l}\epsilon_{i,j} + O_P\Big(\frac{1}{p}\Big) \tag{4.28}$$

$$= \tilde{\delta}_i^{(k)} + \tilde{\delta}_i^{(k)}\epsilon_i' + O_P\Big(\frac{1}{p}\Big) \tag{4.29}$$

$$= \tilde{\delta}_i^{(k)} \times \Big(1 + \epsilon_i' + O_P\Big(\frac{1}{p}\Big)\Big) \tag{4.30}$$

where $\epsilon_i' = 2\langle c_i\varphi_i, \epsilon_i\rangle = 2\langle c_i\varphi_i, \hat{\varphi}_i\rangle - 2\langle c_i\varphi_i, c_i\varphi_i\rangle = 2(\epsilon_i - 1)$. □

Remark As for Fremdt test, we can choose to consider only a reduced number of eigenvalues. This gives us the opportunity to consider the case where the variables are strongly correlated, by neglecting the smallest eigenvalues that are really close to zero, and for which the computation of the T_4 statistics may be numerically unstable.

4.3.4.2 Example

We illustrate the results of this new test thanks to Gaussian simulations. We simulate two sets of $m = 10$ random Gaussian vectors of length $p = 200$.

$$Y_k^{(1)} = \mathcal{N}_m(0, \Gamma),\ k = 1, ..., p$$

$$Y_k^{(2)} = \mathcal{N}_m(0, \Gamma), \ k = 1, ..., p$$

and we suppose that $Y_1^{(1)}, \ldots, Y_p^{(1)}$ are independent (resp. for $Y^{(2)}$).
We repeat the operation $N = 1000$ times, and we change randomly the empirical covariance matrix Γ at each iteration.
We choose to represent the distribution of l_i, where

$$l_i = \frac{\sqrt{p}}{2} \times L_i = \frac{\sqrt{p}}{2} \ln\left(\frac{\hat{\delta}_i^{(1)}}{\hat{\delta}_i^{(2)}}\right),$$

for all $i = 1, ..., m$. Its distribution is expected to be close to a $\mathcal{N}(0, 1)$ distribution. The distributions for all $i = 1, \ldots, m$ are represented in Figure 4.2. The darker red curve represents the distribution of l_1. Then, the color of the estimated pdf becomes lighter when we increase the level i until $i = m$. The blue curve represents the pdf of a $\mathcal{N}(0, 1)$ distribution. Visually, the distribution of l_i seems reall close to a $\mathcal{N}(0, 1)$, for

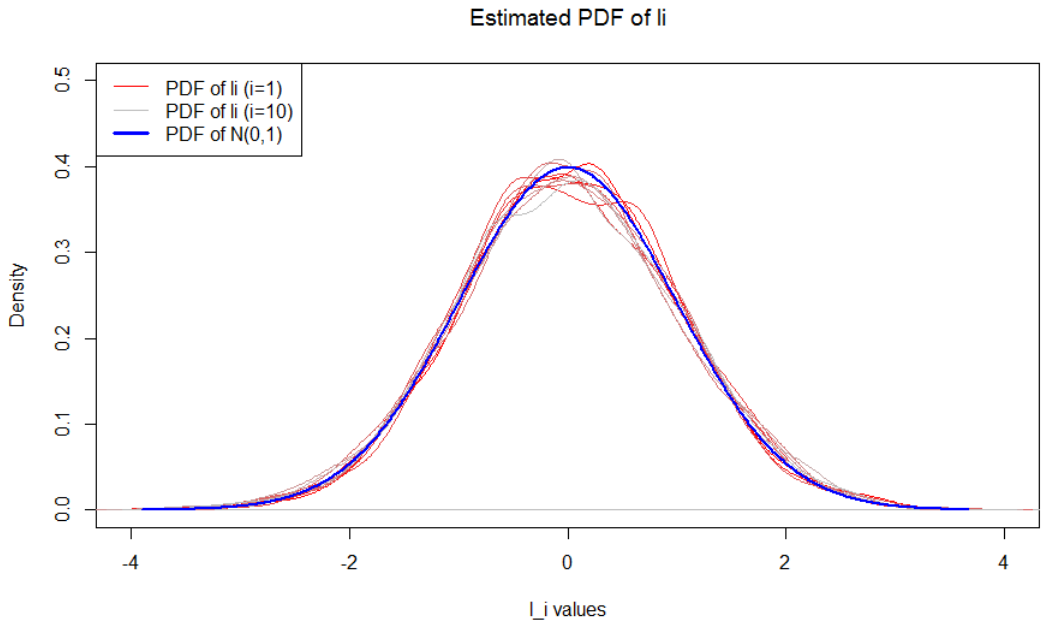

Figure 4.2 – Density of l_i, $i = 1, ..., 10$ and $\mathcal{N}(0, 1)$ distribution.

all $i = 1, ..., m$, with $m = 10$. In order to consolidate this assumption, we also apply the Kolmogorov-Smirnov goodness-of-fit test on l_i for each level $i = 1, ..., 10$. We accept the hypothesis that $l_i \sim \mathcal{N}(0, 1)$, for each $i = 1, ..., 10$ at the level $\alpha = 0.05$. The smallest p-value is obtained for $i = 3$ and is worth 0.26.
We also represent the pdf of T_4 in the figure 4.3. We can see that its distribution is really

close to a $\chi^2(d)$ distribution, for $d = 10$. The Kolmogorov-Smirnov goodness-of-fit test on the hypothesis $H_0 : \{T_4 \sim \chi^2(10)\}$ returns a p-value equal to 0.19, which corresponds to the acceptance of the test at the level $\alpha = 0.05$.
This simulation enhances the proposition 2 in the case of Gaussian distributions, for $p = 200$.

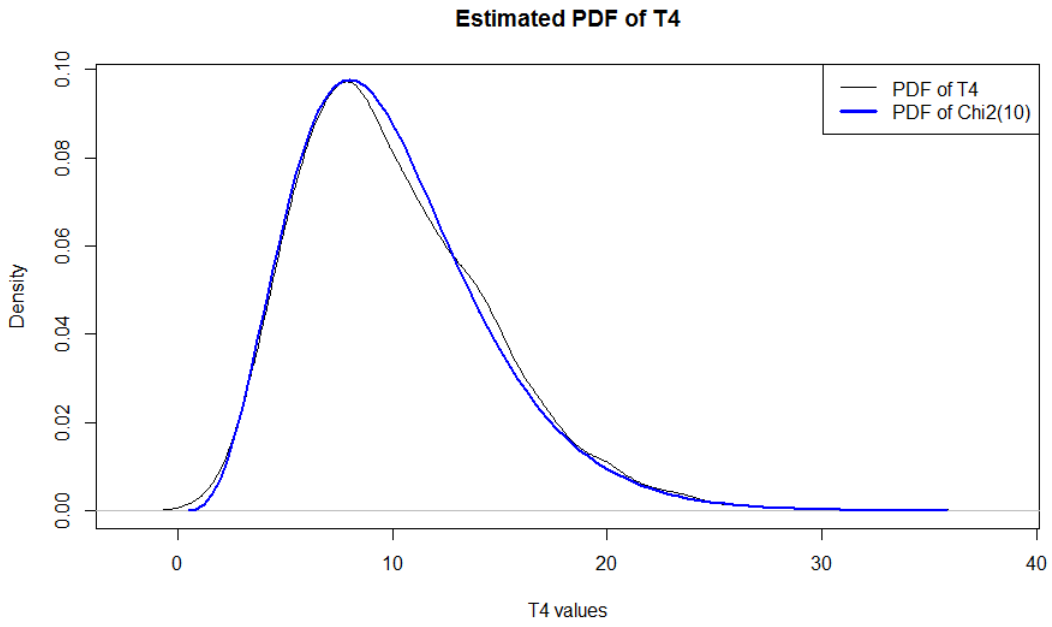

Figure 4.3 – Density of T_4, $i = 1, ..., 10$ and $\chi^2(10)$ distribution.

4.3.4.3 Choice of the parameter d

As mentionned in the definition of the test, it is sometimes relevant to consider only a reduced number of eigenvalues d. The choice of $d \leq m$ may be important. Here are some ideas to set it up automatically.

- Choose d such as the d first eigenvalues of $\hat{\mathbf{\Gamma}}$ represent at least a proportion of $1-\alpha$ of the variance, for a given $\alpha > 0$.

$$d = \underset{d'=1..m}{\operatorname{argmin}} \left(\frac{\sum_{i=1}^{d'} \hat{\lambda}_i}{\sum_{i=1}^{m} \hat{\lambda}_i} > 1 - \alpha \right)$$

- d can be defined thanks to a multiple testing procedure. This can be done in a further study.

4.4 Application

In this section, we apply the four tests on several types of data. The goal is to rank these tests by comparing their power and their adaptability on several sources of data.

4.4.1 On Gaussian distributions

We repeat $N = 1000$ times the following operations.

1. Simulate three sets of $m = 10$ Gaussian random variables.

$$Y_k^{(1)} = \mathcal{N}_m(0, \Gamma_1),\ k = 1, ..., p$$

$$Y_k^{(2)} = \mathcal{N}_m(0, \Gamma_2),\ k = 1, ..., p$$

$$Y_k^{(3)} = \mathcal{N}_m(0, \Gamma_3),\ k = 1, ..., p$$

   with $p = 100$, and we suppose that $Y_1^{(1)}, \ldots, Y_p^{(1)}$ are independent (resp. for $Y^{(2)}$ and $Y^{(3)}$). In this example we have $\Gamma_1 = \Gamma_2$. We fix Γ_1 and Γ_2 as the covariance of a known set of telemetries.
   Γ_3 is chosen in order to change lightly two rows and columns of Γ_1 hence Γ_2, which can be interprated in our application by changes affecting two among the ten telemetries.

2. Estimate Γ_1, Γ_2 and Γ_3 for each set $\hat{\Gamma}_1 = (Y^{(1)})^T Y^{(1)}/p$, $\hat{\Gamma}_2 = (Y^{(2)})^T Y^{(2)}/p$ and $\hat{\Gamma}_3 = (Y^{(3)})^T Y^{(3)}/p$.

3. Compute the test statistics T_1, T_2, T_3, T_4 corresponding to the four tests previously defined, to test the two hypothesis $H_{0,1} : \{\Gamma_1 = \Gamma_2\}$ and $H_{0,2} : \{\Gamma_1 = \Gamma_3\}$. The first test should be accepted since $\Gamma_1 = \Gamma_2$, whereas the second one should be rejected. In this example, we keep all the levels, leading to have $d = 10$.

4. Draw the ROC curve from the outputs of all the simulations.

The figure 4.4 shows the ROC curves resulting from these simulations. We can see that the novel test is approximately as powerful on Gaussian simulations than the Ilea test. The Cai test gives really bad results on this simulation. In the next section, we apply those tests on our data and similar simulated data, to check if the tests have the same power in a different framework on data that are not always Gaussian.

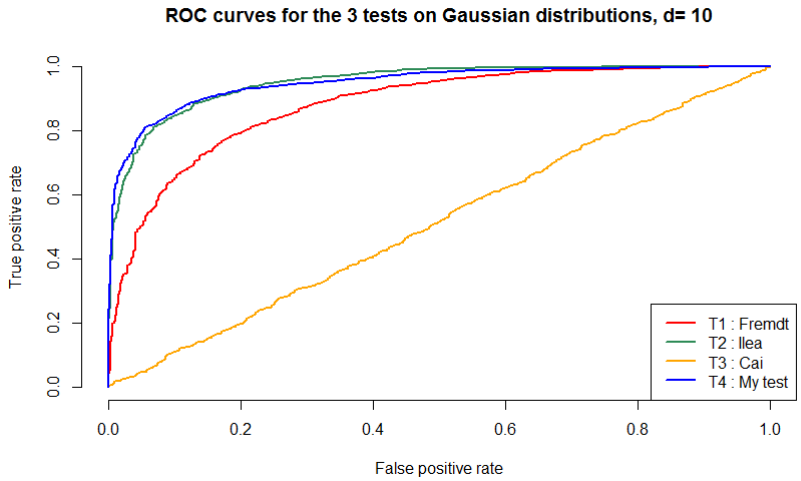

Figure 4.4 – Comparison of the power of the three tests on Gaussian simulations, with $m = 10$, $d = 10$ and $p = 100$.

4.4.2 On simulated telemetries with no anomalies

We apply the three tests previously introduced on simulated telemetries, simulated in the same way as the ones generated in the figure 4.1. We simulate these telemetries by fixing the following parameters.

- We consider $m = 10$ telemetries.
- They are observed $p = 100$ times a day.
- We consider $n = 365$ days, leading to test, for each $i = 2, ..., 365$ the hypothesis $H_0 : \{\mathbf{\Gamma}^{(i)} = \mathbf{\Gamma}^{(i-1)}\}$.

In this example, we do not add any anomalies on the data. However, it is important to notice that those data are not always Gaussian, and exhibit periodic patterns. Those telemetries can be impacted by seasonal effects, changing a little the patterns every day. In fact, we challenge the statistical tests previously defined on these data to see how the tests interprate the seasonnality aspect, knowing that we expect to have no rejections for this set of telemetries.
Several values of d for the Fremdt test and the new test are tested. We can see on the figures 4.5 and 4.6 that the choice of d has a strong influence on the results of the Fremdt test, whereas the new test remains really stable : every day, the value of T_4 corresponds to a p-value really close to one, leading to have no rejections, whereas the Ilea test always returns a p-value equal to zero, which correspond to a strong rejection of the hypothesis. The Cai test detects some false alarms.

We can notice that the test we introduced in this paper seems to be the most appropriate on this type of data.
The Fremdt test can have eather no rejections or a large number of rejections depending on the choice of the parameter d. The Ilea test always rejects the hypothesis for every single simulation we have made in this framework. In fact, the telemetries are often strongly correlated, hence this test is not applicable because the covariance matrices are ill-conditioned. Finally, the new test has returned the results that were expected.
On the next application, we will apply the Fremdt test, the Cai test and the new test on simulated telemetries that contain anomalies. The Ilea test can be abandonned since we expect to have only rejections, as for this application.

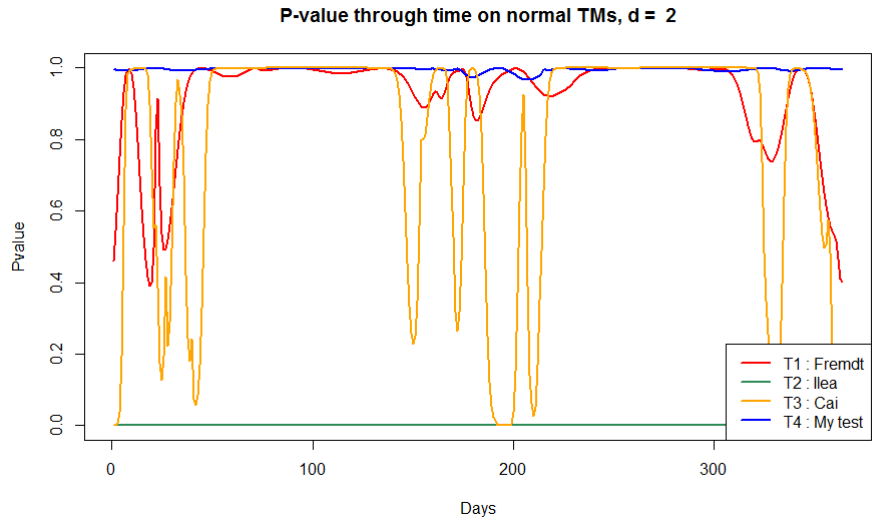

Figure 4.5 – P-value through the days on simulated TM with no anomalies, d=2.

4.4.3 On simulated telemetries with random anomalies

We simulate random telemetries on which we add anomalies, that can be eather local, such as spikes, noise, or global, such as changes in the pattern. We simulate them by fixing the following parameters.

- We consider $m = 10$ telemetries.
- There are observed $p = 200$ times a day.
- We consider $n = 365$ days, leading to test, for each day $i = 2, ..., 365$ the hypothesis $H_0 : \{\mathbf{\Gamma}^{(i)} = \mathbf{\Gamma}^{(i-1)}\}$.

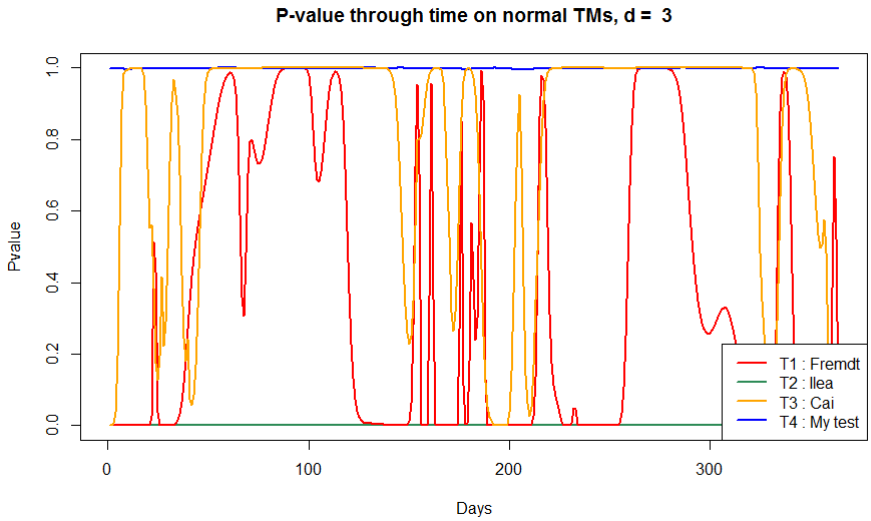

Figure 4.6 – P-value through the days on simulated TM with no anomalies, d=3.

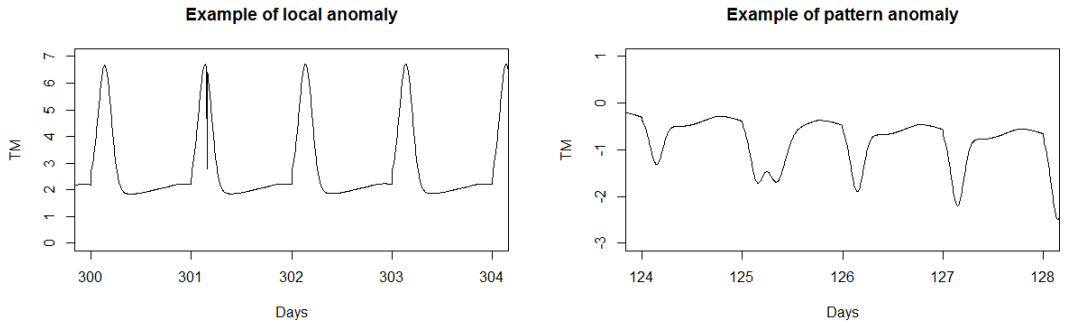

Figure 4.7 – Example of a local anomaly on day 301 (on the left) and a pattern anomaly on day 125 (on the right).

On the figure 4.7 are represented examples of inserted anomalies. These anomalies are inserted randomly, hence there is no family effect on the anomalies : they are inserted on different days in each telemetry.
We apply the Fremdt test, the Cai test and the new test on those data, and we choose to test $d \in \{2, 3, 4\}$ for the Fremdt test, as the previous application showed that the high levels for d were returning too many false alarms. We choose to set $d \in \{5, 6, 7\}$ for the novel test, to see if the choice of d as an influcence of the number of anomalies that we can detect. The results are represented in the figure 4.8. As we can see, the new test is much more efficient on this type of data. We can see that larger values of d enable to detect lighter anomalies such as the local anomalies. In fact, the local anomalies cannot

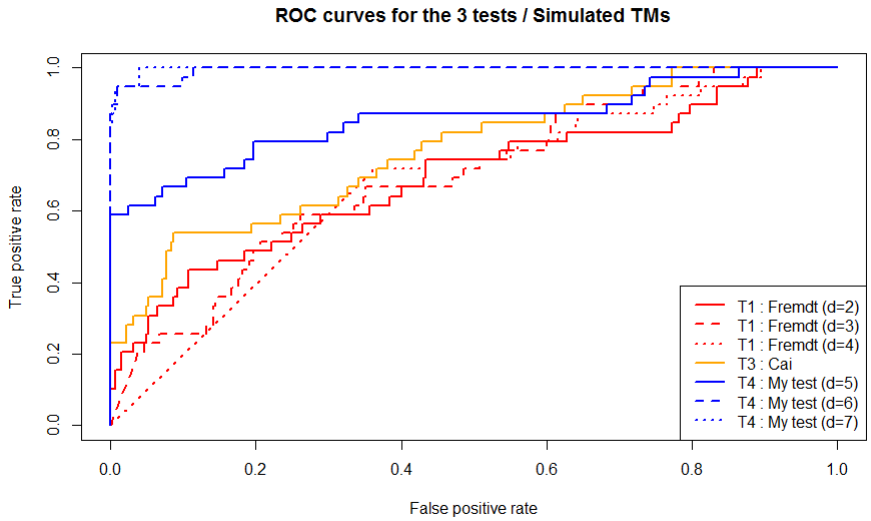

Figure 4.8 – ROC curve for the Fremdt test, the Cai test and the new test on simulated anomalies.

be detected with $d = 5$ but they can be mostly detected with $d = 7$.

On the figures 4.9, 4.10 and 4.11, we can see that almost all the days on which we

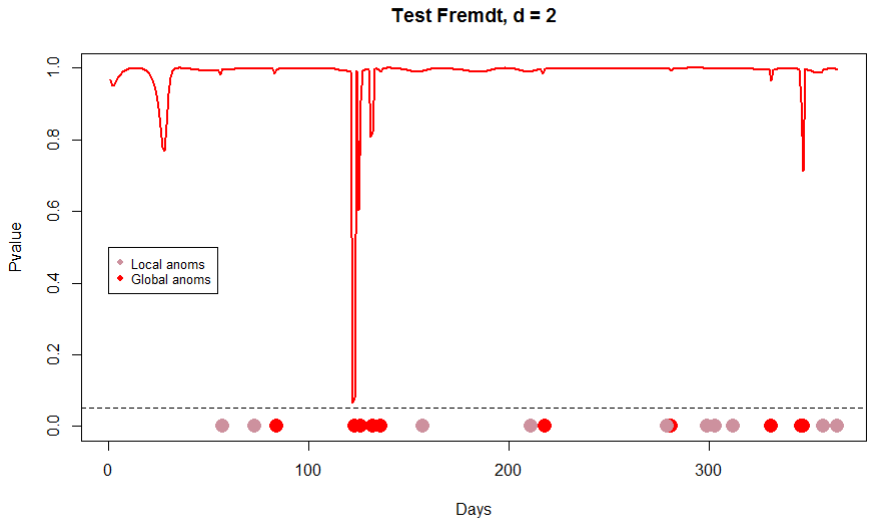

Figure 4.9 – Evolution of the p-value deriving from the Fremdt test, with $d = 2$ through the days and real anomalies: local anomalies are in pink, and pattern anomalies are in red. The algorithm misses almost all the detections.

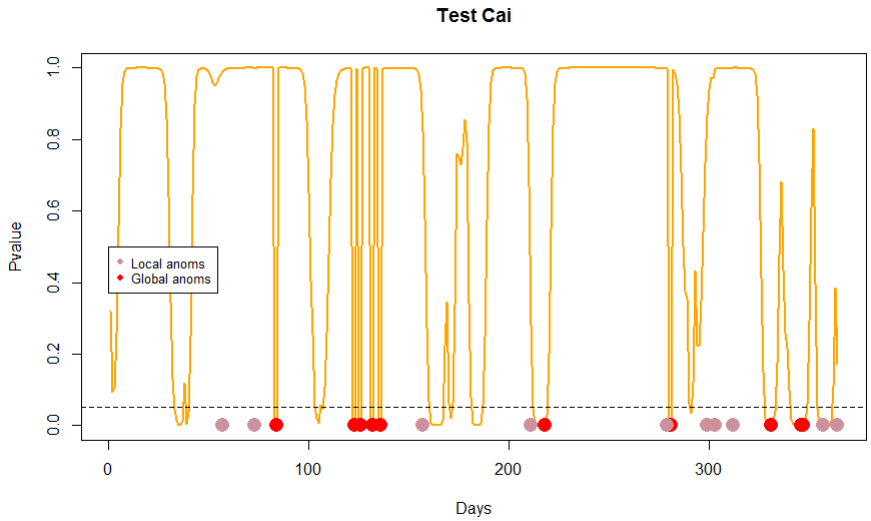

Figure 4.10 – Evolution of the p-value deriving from the Cai test through the days and real anomalies: local anomalies are in pink, and pattern anomalies are in red. The algorithm generates many false alarms.

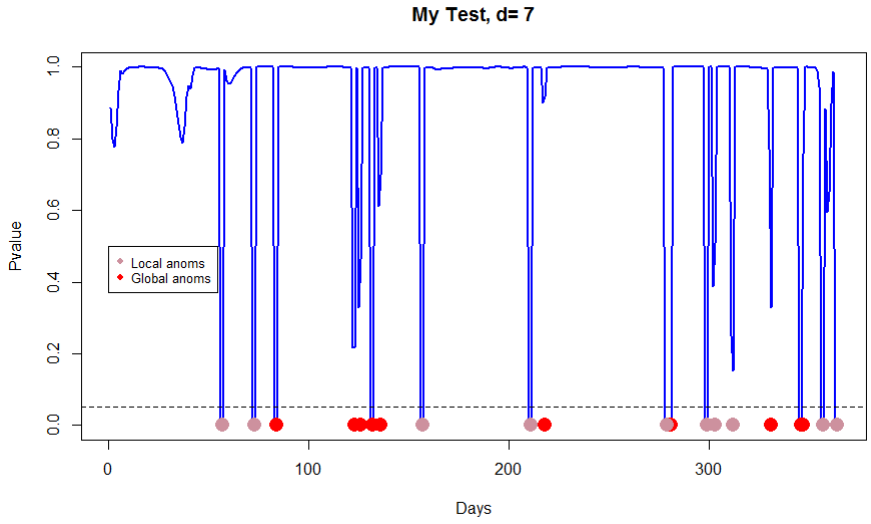

Figure 4.11 – Evolution of the p-value deriving from the new test with $d = 7$ through the days and real anomalies : local anomalies are in pink, and pattern anomalies are in red. Almost all the anomalies are detected with no false alarms at the level 5%. this test is the most robust in this application.

inserted anomalies were detected as abnormal by the new test. The Fremdt test is also able to detect some of these misbehaviors, but it misses most of them. In fact, it consider as abnormal some of the days where no anomaly occured if we increase the parameter d, as we have seen in the previous application : the dashed red line for which d=4 is mostly degenerated. The ROC curves in figure 4.8 shows clearly that the test we have developed is much more powerful on this dataset.
The Cai test is a little bit more powerful than the Fremdt test but it also returns many false alarms and misses most of the local anomalies.
As the anomalies inserted are not linked together, the results here must be validated on real telemetries where known anomalies impact several telemetries. This will be done in the next section.

4.4.4 On real telemetries

4.4.4.1 On the raw-data

We choose to apply the test on a group of 20 correlated telemetries that are closely linked. Those telemetries are daily periodical and can be cut into days. We have observed those telemetries on 365 days, then we can apply 364 tests, for each couple of days $(i-1, i)$, for $i = 2, ..., n$, to catch the changes in the behavior of those telemetries.
The choice of d is done in order to keep at least 99% of the variance, for each covariance matrix. For most of the days, the first 2 eigenvalues represent already more than 99% of the variance, and for some other days we need to keep the 3 first eigenvalues. We apply the Fremdt test and the new test, at first with $d = 2$ and then with $d = 3$. We also apply the Cai test. In the figure 4.12 are represented the evolutions of the test statistics through time and their sensitivity to the choice of the parameter d. We have represented in red three days for which we observe the maximum rejections in at least one of our scenarii : the days 151, 319 and 363.
We can see that the days for which the tests are rejected change a little bit if we consider $d = 2$ or $d = 3$, specially for the Fremdt test. For example, the day 363 that represents the biggest rejection of the Fremdt test with $d = 3$ was not rejected by the same test for $d = 2$. For the novel test, the rejections do not change so much, in particular for the days we highlight, which corresponds to some of the biggest rejections.
We choose to represent the telemetries for each of the selected day, and the days before, to highlight the changes in the behavior of the telemetries. The figure 4.13 shows that the patterns changed a little bit between the two days of telemetries 150 and 151, mostly on the first portion of the signal. The figure 4.14 exhibits changes in the patterns mostly on the second part of the signal, on the day 319. Finally, the day 363 exhibits major changes in the covariance structure, that corresponds actually to a real event that occured in the satellite. The figure 4.16 shows three main behaviors that occured this day on this set of telemetries. A red background has been added to identify the day corresponding to

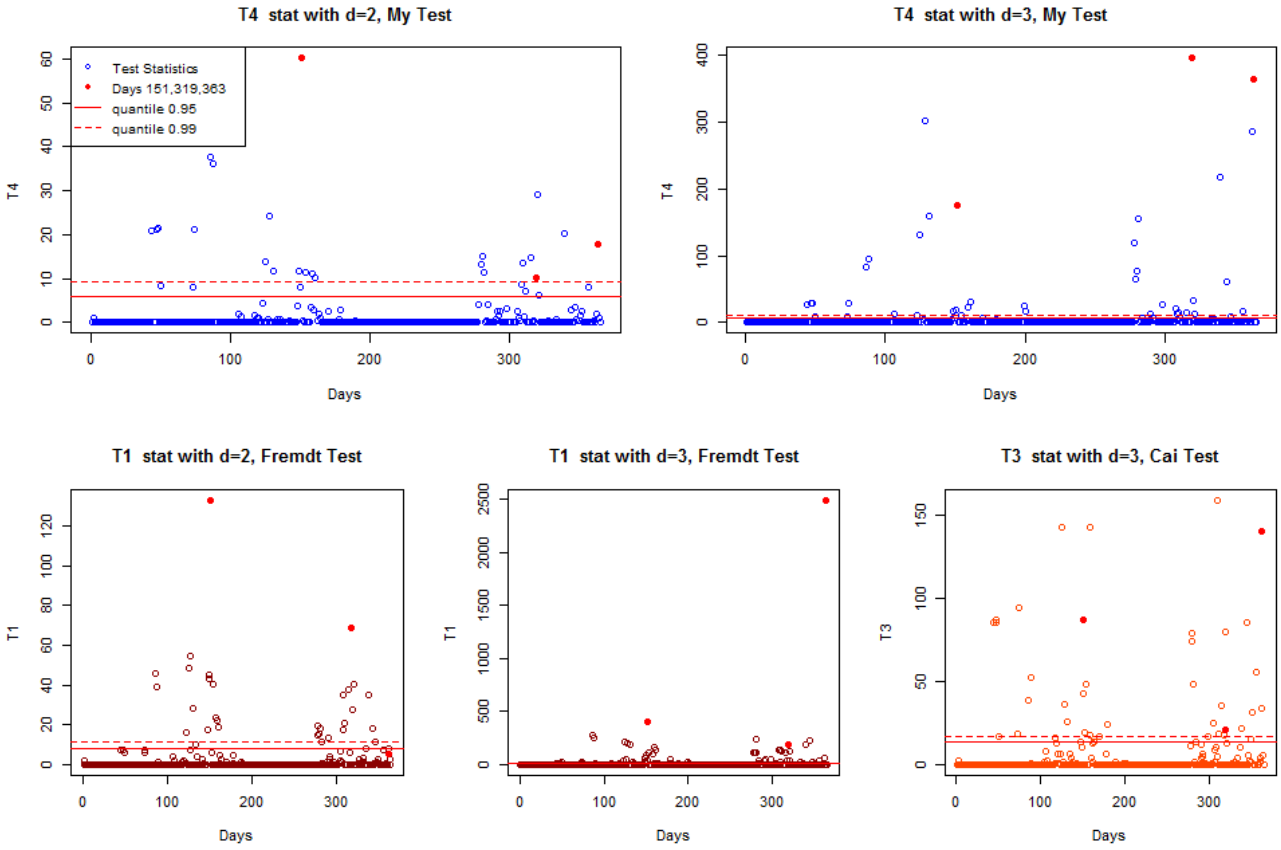

Figure 4.12 – Evolution of the test statistics through time, Novel test (in blue) and Fremdt test (in red) and Cai Test (in orange). The days in red are the values for three reference days : 151, 319 and 363.

the detected change. These events changed strongly the covariance structure, and all the telemetries were impacted by this event. The upper telemetries are the ones for which the telemetries start to decrease that day, not necessarily simultaneously. The telemetries in the center exhibit only light pattern changes. The telemetries in the bottom of the figure are the telemetries that increase.

4.4.4.2 Dimension Reduction using wavelets

In order to get the sensitivity in the dimension reduction framework, we apply the new test on these data on the coefficients arising from the Haar wavelet decomposition, until the upper scale level $L = 4$, as for the example 4.2.4. We can see that two among the three days we have highlighted still corresponds to major rejections, in particular for the day 363. Unlike the test on the raw-data, the choice of $d = 2$ is enough to detect the day

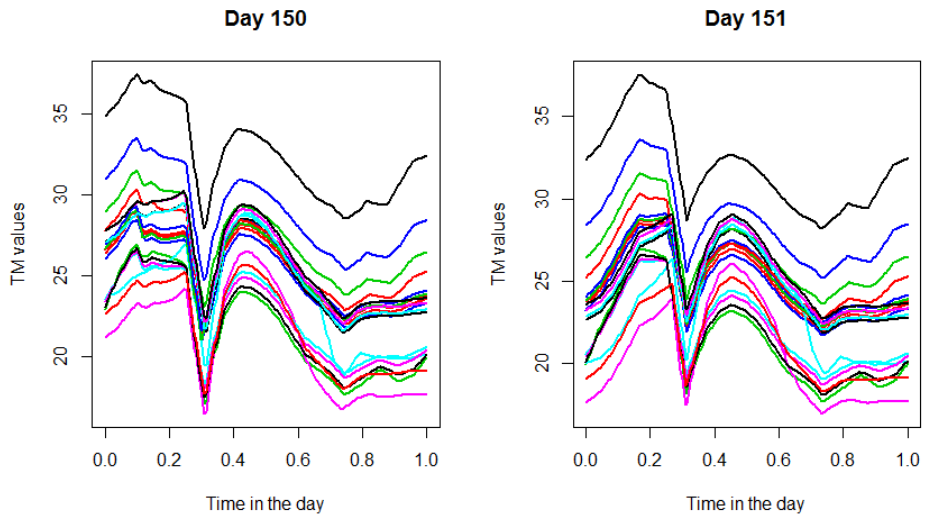

Figure 4.13 – Day 150 and 151, change detected in the covariance matrix.

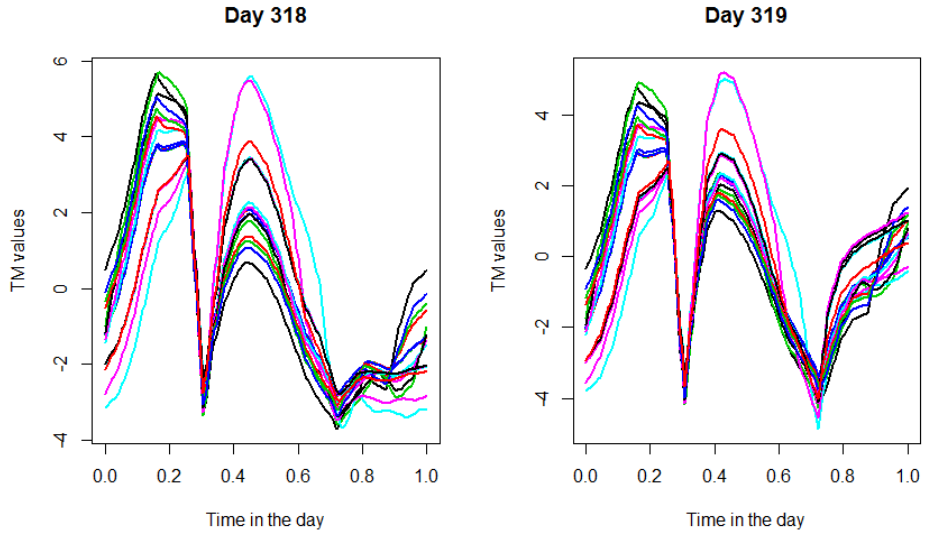

Figure 4.14 – Day 318 and 319, change detected in the covariance matrix.

363 as a major outlier.
Dimension reduction can be used in order to reduce the time computation and the volume of the data that is considered. As we can see on this example, the major events will be easier to raise. However, reducing the dimension makes the detection of local events more difficult to detect.

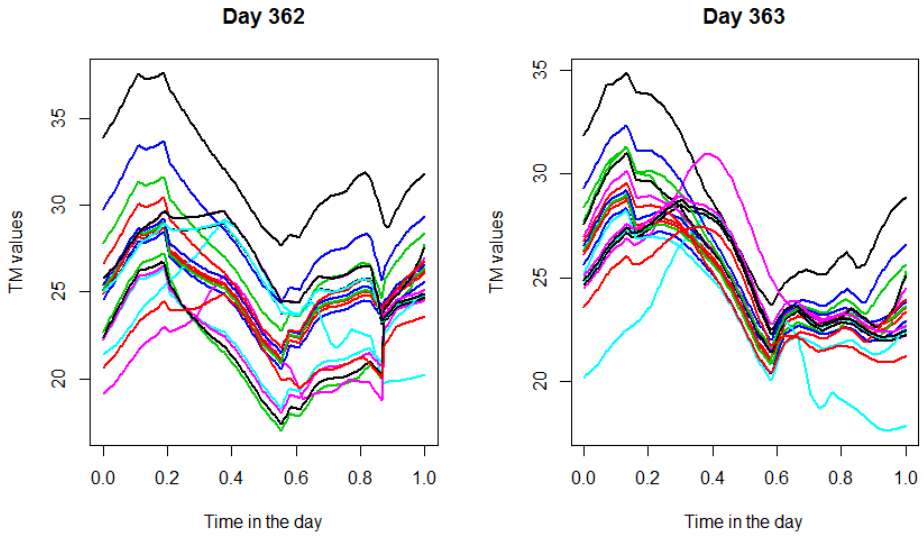

Figure 4.15 – Day 362 and 363, change detected in the covariance matrix.

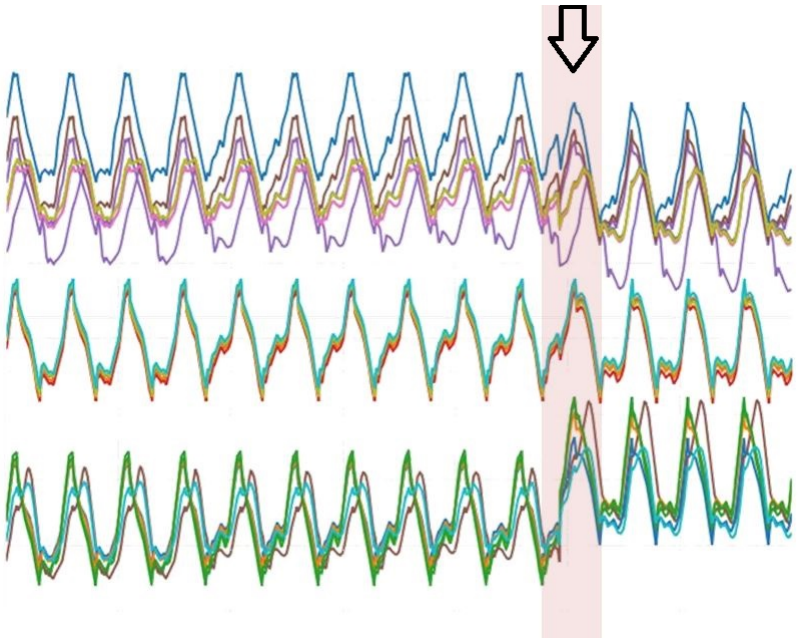

Figure 4.16 – Change at the day 363 represented in three clusters of telemetries.

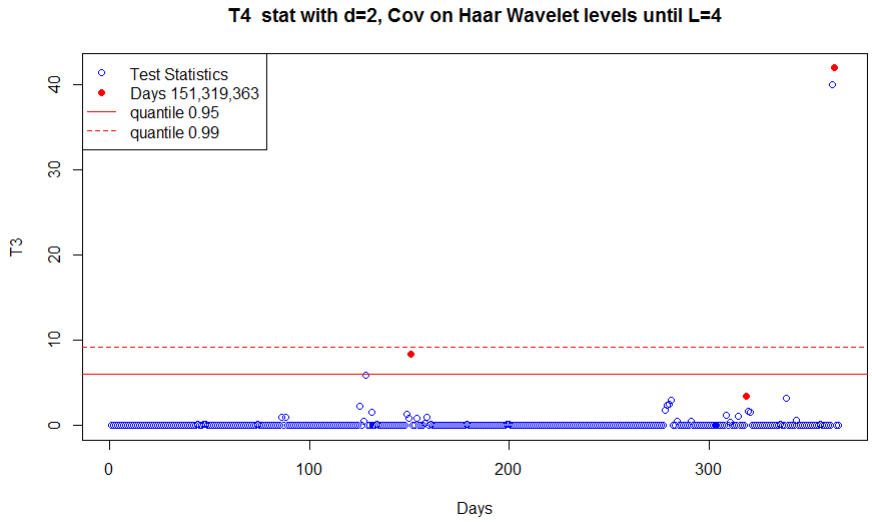

Figure 4.17 – Evolution of the test statistics deriving from the novel test through time. The days in red are the values for three reference days : 151, 319 and 363.

4.5 Conclusion

We have seen that we are able to detect events that change the behavior of groups of telemetries thanks to statistical tests applied on the daily covariance matrices. The methods we propose are really fast, and can be applied on a large number of telemetries. To follow-up the satellite in-flight, such methods can be applied on clusters of telemetries, to help for the daily monitoring as well as supporting deeper investigations.
Among the statistical tests we proposed in this paper, the one we have developed seems to be the best choice for this application.
In fact, it enables to highlight isolated anomalies as well as events in several telemetries. It is also really easy to set up : there is no need to invert any matrix. It is also the most robust to the choice of the parameter d. In fact, as we consider only the d parameters corresponding to the eigenvalues in a new basis, we neglect the values of the rest of the matrix, supposed to be equal to zero, that can be interpreted as useless information. One further work can be to apply different eigenvalues selections, thanks to multiple testing, for example.
In the case of telemetries, it is also really relevant on strongly correlated data, as the data we have in the application to space telemetries. The test developed by Ilea cannot be used on this type of data, despite it is the test that has the best power on Gaussian simulations.
These methods can be used in a dimension reduction framework, as we have seen in our simulations. Coupled with univariate methods, these algorithms are really efficient in highlighting the most abnormal behaviors almost real-time.
As a further work, one can develop multiple testing to set up automatically the choice of the parameter d, or generalize the tests by taking into account more than two days at each test. These methods are already implemented within a web application at Airbus Defence and Space, that has already been used successfully for anomaly investigations.

Conclusion

Les objectifs de cette thèse visaient initialement à développer des méthodes de détection d'anomalies univariées et multivariées applicables aux spécificités des données de tests et des télémesures satellites dans le but de détecter les divergences de comportement à bord du satellite et de détecter les pannes dès qu'elles surviennent.
Nous avons choisi de nous concentrer sur des méthodes particulièrement pertinentes dans le cadre de traitement de données fonctionnelles. En partant de ce postulat, les approches choisies sont variées, de même que les apports théoriques auxquels nous avons contribué.

Une première étape a consisté à comprendre et catégoriser l'anomalie dans les données dont nous disposons. Ainsi, nous avons pu comparer les résultats de chacune des méthodes en fonction de chaque typologie d'anomalie observée. Une fois cette définition établie, nous avons appliqué des méthodes issues de l'état de l'art pour détecter des comportements atypiques dans des données fonctionnelles. Cette étape nous a aidé à comprendre les mécanismes de la détection d'anomalies, et comment fonctionnaient ces méthodes dans le cadre de leur application aux données fonctionnelles. Nous avons testé une grande variété de projections sur des bases de fonctions et de méthodes de détection d'anomalies. Nous avons pu en déduire qu'il était très difficile de mettre en lumière toutes les familles d'anomalies attendues à l'aide d'une projection et d'une méthode de détection d'anomalies données. En agrégeant les différentes projections, on était plus à-même de détecter toutes les anomalies, en particulier les anomalies locales, moins évidentes en général à isoler des autres données.
Comme nous avons pu le constater, cette première partie de la thèse a surtout été applicative. Même si cette étape en soi n'a pas eu d'apport théorique majeur, elle a tout de même été essentielle puisqu'elle nous a ouvert des possibilités pour aller plus loin que l'état de l'art. Les réelles avancées théoriques de la thèse ont été acquises plus tard.
Nous avons tout d'abord créé une procédure de tests multiples pour sélectionner les coefficients les plus pertinents pour la détection d'anomalies, optimisant la création de ces indicateurs. Ainsi, les coefficients issus d'une base de projections bien choisie sont capables d'isoler toutes les familles d'anomalies. Nous améliorons très largement les résultats obtenus lors de l'étape précédente. De plus, dès lors qu'une partie de la donnée est labélisée comme étant nominale, nous savons qu'il est suffisant de s'intéresser à une base de projec-

tions, en particulier la base issue de l'analyse en composantes principales, ainsi qu'à une méthode de détection d'anomalies donnée. De cette manière, nous pouvons nous abstenir de réaliser une revue complète de toutes les méthodes telle que nous l'avons réalisé dans la première étape pour chaque nouvelle application.
Enfin, nous avons abordé l'aspect multivarié des données grâce à l'analyse de la covariance. Nous avons contribué à améliorer la connaissance de ce sujet puisque nous avons développé un test original, dont les performances ont été démontrées grâce à des applications à des données simulées et réelles. La convergence de ce test a également été démontrée théoriquement. De plus, nous démontrons que pour notre application, à savoir les télémesures spatiales, ce test est le plus performant pour détecter des anomalies.

Ces deux dernières contributions, à savoir la sélection des coefficients à l'aide d'une procédure de tests multiples et l'analyse de la covariance des télémesures nous ont permis d'obtenir de très bons résultats pour la détection d'anomalies dans les télémesures. D'ores-et-déjà, certaines améliorations peuvent être apportées à ces deux méthodes. Concernant la procédure de tests multiples, nous avons vu qu'elle reposait en premier lieu sur l'identification d'un sous-ensemble d'individus connus comme ne comportant pas d'anomalie. Nous avons montré par des simulations que même si l'ensemble de référence contenait quelques anomalies, la procédure permettait de bien identifier les coefficients relatant de l'information pertinente sur les anomalies. Cet aspect peut faire l'objet d'investigations plus approfondies. Nous pourrions également développer un test non supervisé pour détecter la présence ou non de données aberrantes dans un vecteur de données. Certains tests existent déjà dans ce but, mais s'appuyent sur des hypothèses paramétriques fortes ou sont construits à l'aide de permutations bootstrap très coûteuses en temps de calcul.
Enfin, pour améliorer la détection d'anomalies dans un contexte multivarié, nous pourrions généraliser cette procédure de tests en comparant plus de deux journées isolées à la fois. Par exemple, on pourrait comparer la nouvelle journée de télémesures avec l'ensemble des 10 dernières journées. De cette manière, on pourrait être robuste aux journées atypiques, et ainsi ne pas déclencher d'anomalies lors du retour au comportement nominal.
Dans ce manuscrit de thèse, nous ne nous sommes pas focalisés sur l'analyse des télémesures discrètes. Ces télémesures sont également très informatives sur l'état du système, et peuvent être prises en compte dans un environnement multivarié pour la recherche de précurseurs.
Nous avons également abordé la mise en œuvre opérationnelle de ces algorithmes par le biais de développements d'applications. Cette approche pourra à l'avenir nous ouvrir d'autres possibilités, notamment pour prendre en compte le retour utilisateur sur nos algorithmes. Cet aspect renforce le caractère semi-supervisé que l'on a abordé dans le cadre de la procédure de tests multiples. Pour aller plus loin encore, on peut également envisager d'identifier sur les applications les zones jugées intéressantes par les utilisateurs, par exemple grâce aux zooms réalisés sur la télémesure lors certaines périodes récurrentes,

pour certains types de motifs. Cet aspect, s'approchant des pratiques réalisées dans le domaine du webmarketing, sont pour le moment très hypothétiques.

A l'inverse, plutôt que de s'inspirer d'autres applications pour notre problématique, on constate également que d'autres domaines pourraient être intéressés par nos approches. On peut bien entendu penser à toutes les applications autour de l'industrie, de la simulation et des tests, mais aussi pour des applications à d'autres domaines traitant des données fonctionnelles et des séries temporelles.

Comme nous l'avons évoqué en introduction, l'industrie des satellites et de l'Espace est en pleine transformation. Nous avons vu que cette thèse s'est inscrite donc dans un contexte changeant, engendré notamment par l'émergence des constellations de satellites, la mise en orbite effectuée grâce à la propulsion électrique, et la digitalisation facilitant le suivi de la donnée tout au long de la vie du satellite. Le contexte actuel est par conséquent très favorable pour populariser des initiatives autour de l'apprentissage statistique et de l'analyse de données, telle que l'a été cette thèse.
En dehors de la problématique spécifiquement abordée dans la thèse, de nombreux sujets relatifs à l'analyse de données ont émergé au cours des dernières années. On peut citer notamment les applications à l'imagerie satellite qui a fait de très grandes avancées depuis l'émergence du deep-learning. Plus proche du sujet de cette thèse, nous avons découvert au cours de ces trois années une multitude de nouveaux besoins, démontrant ainsi la légitimité de ces approches dans un environnement tel que l'industrie spatiale.
Nous constatons donc qu'en plus des résultats propres à la thèse, nous avons pu contribuer à démontrer le gain potentiel de la fouille de données à de nouveaux métiers autour de l'industrie des satellites.
Je pense personnellement que cela fait partie des contributions importantes que peut apporter une thèse CIFRE que de pouvoir promouvoir sa spécialité au sein d'une entreprise pour l'encourager à investir dans cette voie. Concernant cet aspect, je pense qu'il a été acquis puisqu'on m'a offert la possibilité de rester au sein de l'entreprise.

Bibliographie

[1] Mina Abdel-Sayed, Daniel Duclos, Gilles Faÿ, Jérôme Lacaille, and Mathilde Mougeot. Nmf-based decomposition for anomaly detection applied to vibration analysis. 6 :73–81, 09 2016.

[2] Jake Allen. *Comparison of Time Series and Functional Data Analysis for the Study of Seasonality*. PhD thesis, East Tennessee State University, 2011.

[3] Fabrizio Angiulli and Fabio Fassetti. Detecting distance-based outliers in streams of data. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 811–820. ACM, 2007.

[4] Anestis Antoniadis, Xavier Brossat, Jairo Cugliari, and Jean-Michel Poggi. Clustering functional data using wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(01) :1350003, 2013.

[5] Leo A Aroian. A study of ra fisher's z distribution and the related f distribution. *The Annals of Mathematical Statistics*, 12(4) :429–448, 1941.

[6] Benjamin Auder and Aurélie Fischer. Projection-based curve clustering. *J. Stat. Comput. Simul.*, 82(8) :1145–1168, 2012.

[7] Peter B. de Selding. Airbus and oneweb form joint venture to build 900 satellites. *Space News*, 2016.

[8] Vic Barnett, Toby Lewis, et al. *Outliers in statistical data*, volume 3. Wiley New York, 1994.

[9] Clémentine Barreyre, Béatrice Laurent, Jean-Michel Loubes, and Bertrand Cabon. Détection d'événements atypiques dans des données fonctionnelles. In *Les journées de la Statistique*, 2016.

[10] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1) :289–300, 1995.

[11] Guilhem Boltz. Airbus : le premier satellite haute puissance tout électrique, EUTELSAT 172B, a été lancé par ariane 5. *Décryptagéo*, juin 2017.

[12] Bernard Bonnard, Ludovic Faubourg, and Emmanuel Trélat. *Mécanique céleste et contrôle des véhicules spatiaux*, volume 51. Springer Science & Business Media, 2005.

[13] Denis Bosq. *Linear processes in function spaces : theory and applications*, volume 149. Springer Science & Business Media, 2012.

[14] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF : identifying density-based local outliers. In *ACM sigmod record*, volume 29-2, pages 93–104. ACM, 2000.

[15] Benoît Cadre, Bruno Pelletier, and Pierre Pudlo. Estimation of density level sets with a given probability content. *Journal of Nonparametric Statistics*, 25(1) :261–272, 2013.

[16] Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501) :265–277, 2013.

[17] Bertrand Candelon and Norbert Metiu. A distribution-free test for outliers. *Deutsche Bundesbank, Research Centre*, (02/2013), 2013.

[18] Miguel Cárdenas-Montes. Depth-based outlier detection algorithm. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 122–132. Springer, 2014.

[19] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection : A survey. *ACM computing surveys (CSUR)*, 41(3) :15, 2009.

[20] Ronald R Coifman and M Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*, 38(2) :713–718, 1992.

[21] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[22] Rémy Decourt. Airbus va se doter d'une constellation inédite de satellites d'observation de la terre. *Futura sciences*, 2016.

[23] Eustasio del Barrio, Juan A. Cuesta-Albertos, Carlos Matrán, and Jesús M. Rodrí guez Rodrí guez. Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.*, 27(4) :1230–1239, 1999.

[24] Eustasio Del Barrio and Jean-Michel Loubes. Central limit theorems for empirical transportation cost in general dimension. *arXiv preprint arXiv :1705.01299*, 2017.

[25] Chloé Dimeglio, Santiago Gallón, Jean-Michel Loubes, and Elie Maza. A robust algorithm for template curve estimation based on manifold embedding. *Computational Statistics & Data Analysis*, 70 :373–386, 2014.

[26] Xuemei Ding, Yuhua Li, Ammar Belatreche, and Liam P Maguire. An experimental evaluation of novelty detection methods. *Neurocomputing*, 135 :313–327, 2014.

[27] C. Désir, S. Bernard, C. Petitjean, and L. Heutte. One class random forests. *Pattern Recognition*, 2013.

[28] Manuel Febrero, Pedro Galeano, and Wenceslao González-Manteiga. Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19(4) :331–345, 2008.

[29] Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis : theory and practice*. Springer Science & Business Media, 2006.

[30] Stefan Fremdt, Josef G Steinebach, Lajos Horváth, and Piotr Kokoszka. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1) :138–152, 2013.

[31] Magalie Fromont, Béatrice Laurent, Matthieu Lerasle, Patricia Reynaud-Bouret, et al. Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. In *COLT*, pages 23–1, 2012.

[32] Sylvain Fuertes, Gilles Picart, Jean-Yves Tourneret, Lotfi Chaari, André Ferrari, and Cédric Richard. Improving spacecraft health monitoring with automatic anomaly detection techniques. In *14th International Conference on Space Operations*, page 2430, 2016.

[33] Ryohei Fujimaki, Takehisa Yairi, and Kazuo Machida. An approach to spacecraft anomaly detection problem using kernel feature space. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 401–410. ACM, 2005.

[34] Pedro Galeano, Daniel Peña, and Ruey S Tsay. Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, 101(474) :654–669, 2006.

[35] Fabrice Gamboa, Jean-Michel Loubes, and Elie Maza. Semi-parametric estimation of shifts. *Electronic Journal of Statistics*, 1 :616–640, 2007.

[36] J. González and A. Muñoz. Representing functional data in reproducing kernel hilbert spaces with application to clustering and classification. *Statistics and Econometrics 013*, 2010.

[37] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. Outlier detection for temporal data : A survey. *IEEE Transactions on Knowledge and Data Engineering*, 26(9) :2250–2267, 2014.

[38] Ali S Hadi. Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 761–771, 1992.

[39] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[40] Douglas M Hawkins. *Identification of outliers*, volume 11. Springer, 1980.

[41] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern Recognition Letters*, 24(9) :1641–1650, 2003.

[42] Victoria Hodges. Designs on space : The lifecycle of a satellite. *Catalyst*, 2009.

[43] H. Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 2007.

[44] Mia Hubert, Peter J Rousseeuw, and Pieter Segaert. Multivariate functional outlier detection. *Statistical Methods & Applications*, 24(2) :177–202, 2015.

[45] Ioana Ilea, Lionel Bombrun, Christian Germain, Romulus Terebes, and Monica Borda. Statistical hypothesis test for robust classification on the space of covariance matrices. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 271–275. IEEE, 2015.

[46] Julien Jacques and Cristian Preda. Functional data clustering : a survey. *Advances in Data Analysis and Classification*, 8(3) :231–255, 2014.

[47] MC Jones. The logistic and the log f distribution. *Handbook of the Logistic Distribution*, 2006.

[48] Edwin M Knorr and Raymond T Ng. Finding intensional knowledge of distance-based outliers. In *VLDB*, volume 99, pages 211–222, 1999.

[49] H-P. Kriegel, P. Kröger, E. Schubert, and A.Zimek. Loop : Local outlier probabilities. In *Proceedings of the 18th ACM Conference on Information and Knowledge management*, 2009.

[50] Thibaut Le Gouic and Jean-Michel Loubes. Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields*, pages 1–17, 2016.

[51] Frank Lehot and Jean-François Clervoy. *Histoire de la conquête spatiale*. De Boeck Superieur, 2017.

[52] F.T. Liu, K.M. Ting, and Z-H. Zhou. Isolation forest. *Data Mining*, 2008.

[53] Regina Y Liu, Jesse M Parelius, Kesar Singh, et al. Multivariate analysis by data depth : descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The annals of statistics*, 27(3) :783–858, 1999.

[54] Elio Lozano and E Acufia. Parallel algorithms for distance-based and density-based outliers. In *Data Mining, Fifth IEEE International Conference on*, pages 4–pp. IEEE, 2005.

[55] Gérard Maral and Michel Bousquet. *Satellite communications systems : systems, techniques and technology*. John Wiley & Sons, 2011.

[56] Markos Markou and Sameer Singh. Novelty detection : a review—part 1 : statistical approaches. *Signal processing*, 83(12) :2481–2497, 2003.

[57] José-Antonio Martínez-Heras, Alessandro Donati, Marcus GF Kirsch, and Frederic Schmidt. New telemetry monitoring paradigm with novelty detection. In *SpaceOps 2012 Conference, Stockholm, Sweden*, pages 11–15, 2012.

[58] José-Antonio Martínez-Heras, Alessandro Donati, Bruno Sousa, and Jörg Fischer. Drmust—a data mining approach for anomaly investigation. In *12th International Conference on Space Operations*, pages 11–15, 2012.

[59] J. Mercer. Functions of positive and negative type and their connection with the theory of integral equations. *Philos. Trans. Roy. Soc. London*, 1909.

[60] Isabella Morlini. On the dynamic time warping for computing the dissimilarity between curves. *New developments in classification and data analysis*, pages 63–70, 2005.

[61] C Ordoñez, J Martínez, JR Rodríguez-Pérez, and A Reyes. Detection of outliers in gps measurements by using functional-data analysis. *Journal of Surveying Engineering*, 137(4) :150–155, 2011.

[62] Jun Pan, Jinglong Chen, Yanyang Zi, Yueming Li, and Zhengjia He. Mono-component feature extraction for mechanical fault diagnosis using modified empirical wavelet transform via data-driven adaptive fourier spectrum segment. *Mechanical Systems and Signal Processing*, 72 :160–183, 2016.

[63] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99 :215–249, 2014.

[64] Jean-François Poussin and Gérard Berger. Eurostar E3000 three-year flight experience and perspective. In *25th AIAA International Communications Satellite Systems Conference (organized by APSCC)*, page 3124, 2007.

[65] Tsirizo Rabenoro. *Outils statistiques de traitement d'indicateurs pour le diagnostic et le pronostic des moteurs d'avions*. PhD thesis, Université Paris 1 Panthéon Sorbonne, 2015.

[66] Svetlozar T Rachev. The monge-kantorovich problem on mass transfer and its applications in stochastics. *Teor. Veroyatnost. i Primenen*, 29(4) :625–653, 1984.

[67] Aaditya Ramdas, Nicolás Garcí a Trillos, and Marco Cuturi. On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2) :Paper No. 47, 15, 2017.

[68] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.

[69] Haojie Ren, Nan Chen, and Changliang Zou. Projection-based outlier detection in functional data. *Biometrika*, 104(2) :411–423, 2017.

[70] Joseph P Romano and Michael Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4) :1237–1282, 2005.

[71] Etienne Roquain. Type I error rate for testing many hypotheses : a survey with proofs. *Journal de la Société Francaise de Statistique*, 152(2) :3–38, 2011.

[72] F.S. Schlindwein and D.H. Evans. Autoregressive spectral analysis as an alternative to fast fourier transform analysis of doppler ultrasound signals. *Doppler physics and signal processing*, 1992.

[73] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2000.

[74] B. Schölkopf, A. J. Smola, and K-R. Müller. Kernel principal component analysis. *Artificial Neural Network*, 2005.

[75] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7) :1443–1471, July 2001.

[76] J. Shawe-Taylor and B.Zlicar. Novelty detection with one-class support vector machines. *Advances in Statistical Models for Data Analysis*, 2015.

[77] Rowland R Sillito and Robert B Fisher. Semi-supervised learning for anomalous trajectory detection. In *BMVC*, volume 27, pages 1025–1044, 2008.

[78] Ingo Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov) :67–93, 2001.

[79] Sharmila Subramaniam, Themis Palpanas, Dimitris Papadopoulos, Vana Kalogeraki, and Dimitrios Gunopulos. Online outlier detection in sensor data using non-parametric models. In *Proceedings of the 32nd international conference on Very large data bases*, pages 187–198. VLDB Endowment, 2006.

[80] Yufei Tao, Xiaokui Xiao, and Shuigeng Zhou. Mining distance-based outliers from large databases in any metric space. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 394–403. ACM, 2006.

[81] Thaddeus Tarpey and Kimberly K. J. Kinateder. Clustering functional data. *J. Classification*, 20(1) :93–114, 2003.

[82] Albert Thomas, Vincent Feuillard, and Alexandre Gramfort. Calibration of one-class svm for mv set estimation. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*, pages 1–9. IEEE, 2015.

[83] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63(2) :411–423, 2001.

[84] S. Vantini. On the definition of phase and amplitude variability in functional data analysis. *Test*, 21 :676–696, 2012.

[85] Tommi Vatanen, Mikael Kuusela, Eric Malmi, Tapani Raiko, Timo Aaltonen, and Yoshikazu Nagai. Semi-supervised detection of collective anomalies with an application in high energy particle physics. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8. IEEE, 2012.

[86] Cédric Villani. *Optimal transport : old and new*, volume 338. Springer Science & Business Media, 2008.

[87] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Ann. Statist.*, 25 :1251–1276, 1997.

[88] R Wilcox. Kolmogorov–smirnov test. *Encyclopedia of biostatistics*, 2005.

[89] L. Wu, K. John Wu, A. Sim, M. Churchill, J. Y. Choi, A. Stathopoulos, C. S. Chang, and S. Klasky. Towards real-time detection and tracking of spatio-temporal features : Blob-filaments in fusion plasma. *IEEE Transactions on Big Data*, 2(3) :262–275, Sept 2016.

Articles et Conférences

ARTICLES

- C. Barreyre, B. Laurent, J.M. Loubes, B. Cabon, L. Boussouf
  Statistical Methods for Space Application in Space Telemetries. *A paraître dans l'AIAA (2018)*

- C. Barreyre, B. Laurent, J.M. Loubes, B. Cabon, L. Boussouf
  Multiple Testing for Outlier Detection in Functional Data. *Soumis à IEEE Transactions on Big Data (2018)*

COMMUNICATIONS

- **Détection d'événements atypiques dans les données fonctionnelles**
  Les Journées de la Statistiques, Montpellier (juin 2016).
- **Multiple Testing for Outlier Detection in Functional Data**
  Groupe de travail Machine-Learning - Institut de Mathématiques de Toulouse, Toulouse (juin 2017).
- **Multiple Testing for Outlier Detection in Functional Data**
  European Meeting of Statisticians, Helsinki (juillet 2017).
- **Detecting Outlier Detection in Multivariate telemetries thanks to Covariance Analysis**
  Big Data from Space, Toulouse (novembre 2017).
- **Statistical Methods for Outlier Detection in Space Telemetries**
  SpaceOps, Marseille (mai 2018).

TITLE : High Dimension Statistics for Space Applications on functional data deriving from satellites

Abstract

In this PhD, we have developed statistical methods to detect abnormal events in all the functional data produced by the satellite all through its lifecycle.
The data we are dealing with come from two main phases in the satellite's life, telemetries and test data. A first work on this thesis was to understand how to highlight the outliers thanks to projections onto functional bases. On these projections, we have also applied several outlier detection methods, such as the One-Class SVM, the Local Outlier Factor (LOF). In addition to these two methods, we have developed our own outlier detection method, by taking into account the seasonality of the data we consider.
Based on this study, we have developed an original procedure to select automatically the most interesting coefficients in a semi-supervised framework for the outlier detection, from a given projection. Our method is a multiple testing procedure where we apply the two sample-test to all the levels of coefficients.
We have also chosen to analyze the covariance matrices representing the covariance of the telemetries between themselves for the outlier detection in multivariate data. In this purpose, we are comparing the covariance of a cluster of several telemetries deriving from two consecutive days, or consecutive orbit periods. We have applied three statistical tests targeting this same issue with different approaches. We have also developed an original asymptotic test, inspired by both first tests. In addition to the proof of the convergence of this test, we demonstrate thanks to examples that this new test is the most powerful.
In this PhD, we have tackled several aspects of the anomaly detection in the functional data deriving from satellites. For each of these methods, we have detected all the major anomalies, improving significantly the false discovery rate.

KEYWORDS : Outlier Detection, Functional Data, Unsupervised learning, Multiple Testing, Satellites Data

AUTEURE : Clémentine BARREYRE

TITRE : Statistique en Grande Dimension pour la Détection d'Anomalies dans les Données Fonctionnelles Issues des Satellites

DIRECTEURS DE THESE : Béatrice LAURENT & Jean-Michel LOUBES

LIEU ET DATE DE SOUTENANCE : GMM 13, Bâtiment GMM, RDC, INSA de TOULOUSE, 30 Mars 2018.

Résumé

Ce travail de thèse consiste au développement de méthodes statistiques pour détecter des comportements anormaux dans les données fonctionnelles que produit le satellite tout au long de sa vie. Un premier travail a été de comprendre comment mettre en évidence les anomalies grâce à des projections sur des bases de fonctions. En complément de cette revue des projections, nous avons appliqué plusieurs méthodes de détection d'anomalies, telles que la One-Class SVM et le Local Outlier Factor (LOF). En plus de ces deux méthodes, nous avons développé notre propre méthode pour prendre en compte la saisonnalité des courbes que nous considérons.
En se basant sur cette étude, nous avons développé une nouvelle procédure pour sélectionner automatiquement les coefficients les plus intéressants pour la détection d'anomalies dans un cadre semi-supervisé. Notre méthode est une procédure de tests multiples où nous appliquons un test à deux échantillons à tous les niveaux de coefficients.
Nous nous sommes également intéressés aux covariances des télémesures entre elles pour la détection d'anomalies. Pour cela, nous cherchons à comparer les covariances entre un groupe de télémesures pour deux journées, ou périodes consécutives. Nous avons appliqué trois tests statistiques ayant des angles d'approche différents. Nous avons également développé dans ce but un nouveau test asymptotique. Outre la démonstration de la convergence de notre test, nous démontrons par des exemples que ce test est dans la pratique le plus puissant sur les données dont nous disposons.
Dans cette thèse, nous avons abordé plusieurs aspects de la détection d'anomalies dans les données fonctionnelles issues des satellites. Pour chacune des méthodes, nous avons pu détecter toutes les anomalies, améliorant sensiblement le taux de fausses alarmes.

MOTS-CLES : Détection d'anomalies, Données Fonctionnelles, Apprentissage non Supervisé, Tests Multiples, Données Satellites

DISCIPLINE ADMINISTRATIVE : MITT : Domaine Mathématiques : Mathématiques appliquées

INTITULE ET ADRESSE DU LABORATOIRE : Université Toulouse 3 Paul Sabatier. Institut de Mathématiques de Toulouse (UMR 5219) 18 Route de Narbonne, 31400 Toulouse