

Estimation Parcimonieuse et Apprentissage de Dictionnaires pour la Détection d'Anomalies Multivariées dans des Données Mixtes de Télémétrie Satellites

Barbara PILASTRE

Encadrement:

Jean-Yves TOURNERET, Stéphane D'ESCRIVAN, Loïc BOUSSOUF

6 Novembre 2020

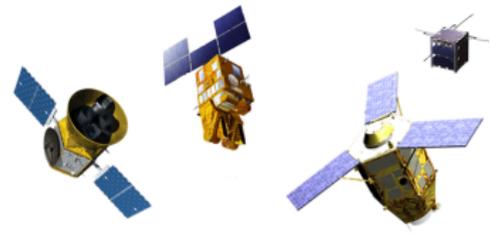
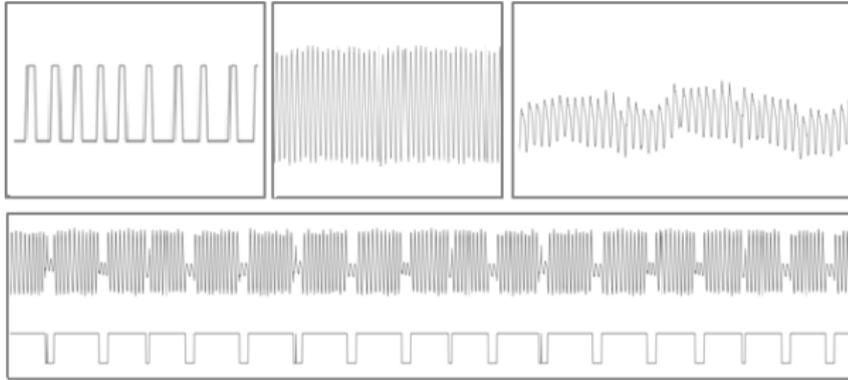
En collaboration avec Pierre-Baptiste Lambert, Clémentine Barreyre, Gustavo Silva et Paul Rodriguez.



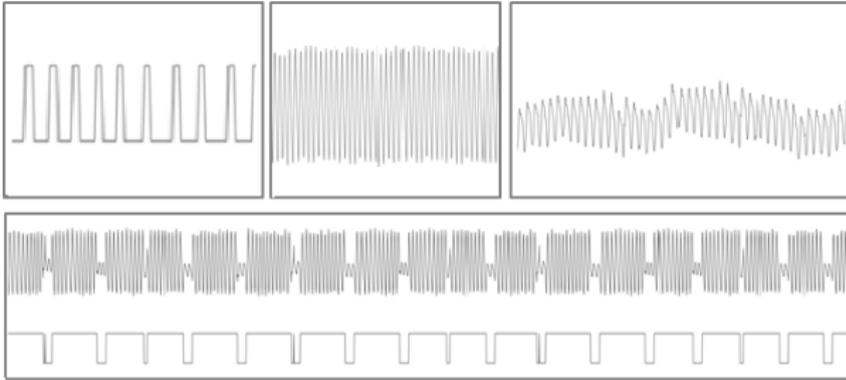
Table des Matières

- ① Contexte
- ② Estimation Parcimonieuse
- ③ Estimation Parcimonieuse Convulsive
- ④ Applications
- ⑤ Conclusions

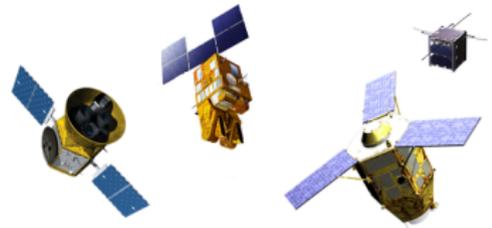
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



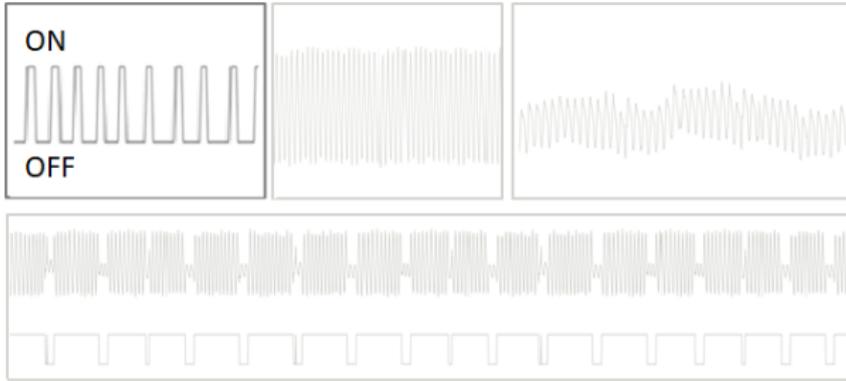
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



- Des centaines/milliers de signaux de télémétrie.



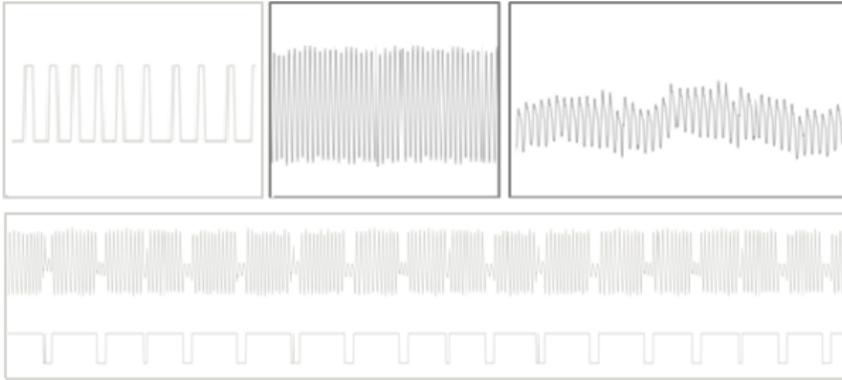
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



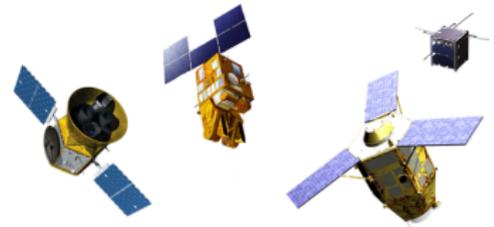
- Des centaines/milliers de signaux de télémétrie.
- Des données discrètes...



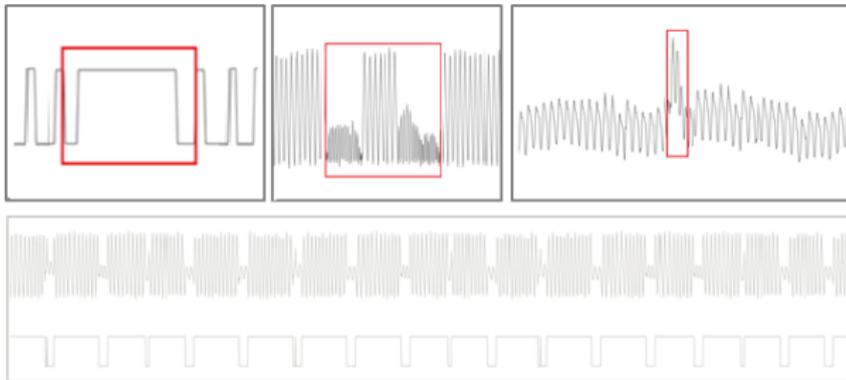
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



- Des centaines/milliers de signaux de télémétrie.
- Des données discrètes et continues.



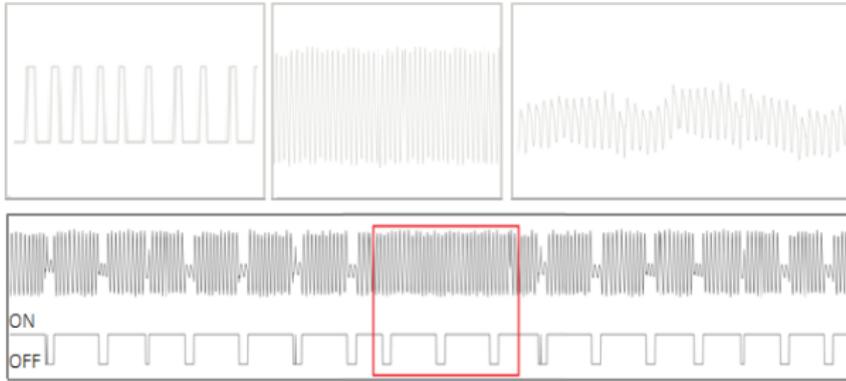
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



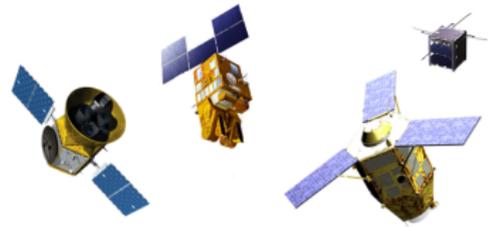
- Des centaines/milliers de signaux de télémétrie.
- Des données discrètes et continues.
- Des anomalies univariées...



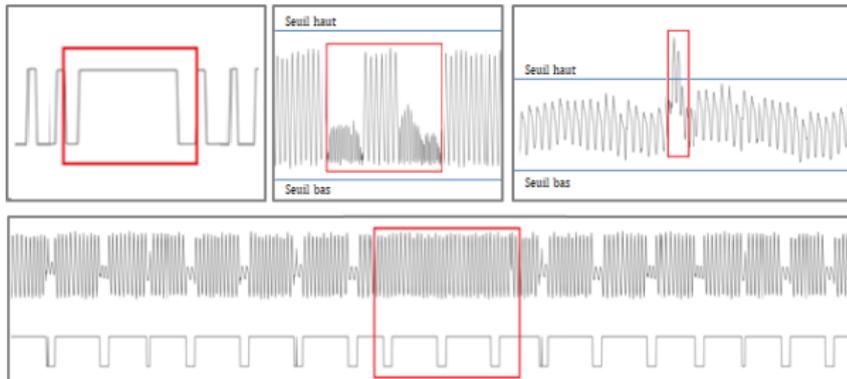
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



- Des centaines/milliers de signaux de télémétrie.
- Des données discrètes et continues.
- Des anomalies univariées et multivariées.



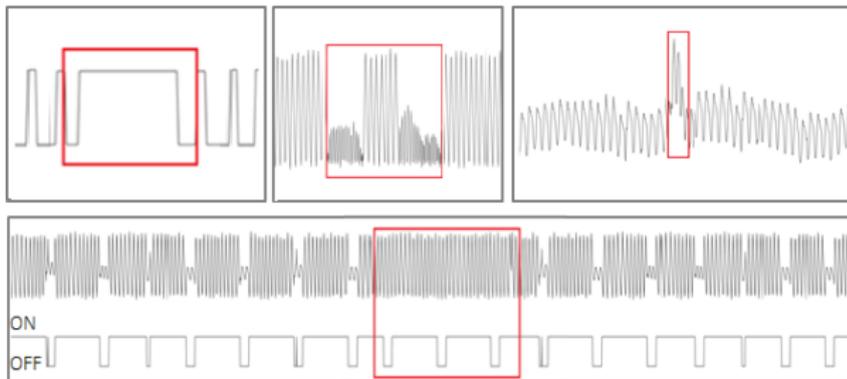
Suivi de l'État de Santé des Satellites : Surveillance de la Télémétrie



- Des centaines/milliers de signaux de télémétrie.
- Des données discrètes et continues.
- Des anomalies univariées et multivariées.
- Une surveillance par seuil (OOL) simple mais limitée.



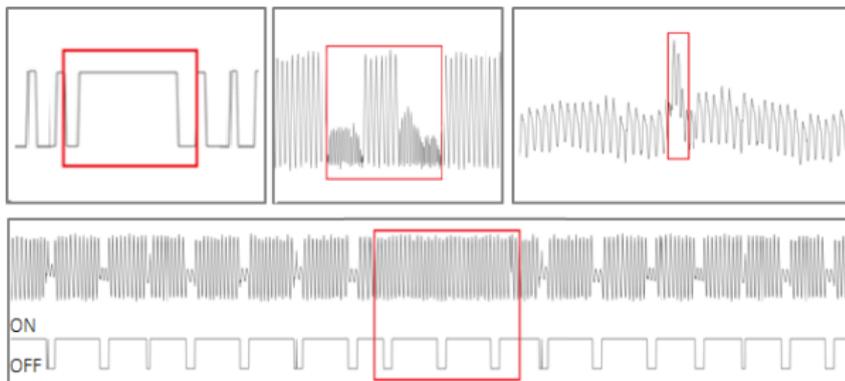
Problématique de Thèse



Enjeux de la thèse

- Traiter conjointement plusieurs signaux de télémétrie satellites.
- Traiter des données discrètes et continues.
- Détecter les anomalies univariées et multivariées.

Problématique de Thèse



Enjeux :

- Traiter conjointement plusieurs signaux de télémétrie satellites.
- Traiter des données discrètes et continues.
- Détecter les anomalies univariées et multivariées.

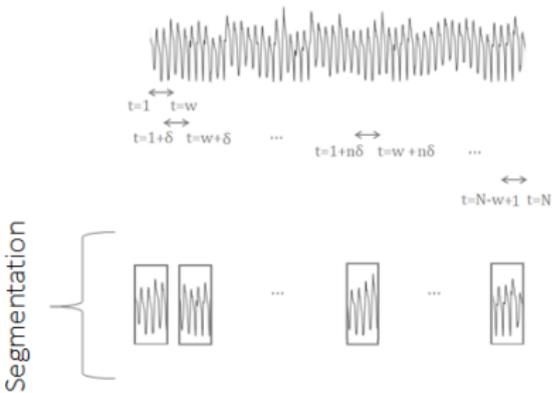
Hypothèse :

- La nouvelle télémétrie nominale peut être reconstruite à l'aide de peu de télémétrie nominale passée.
→ Application de méthodes d'estimation parcimonieuse.

Estimation Parcimonieuse

Estimation Parcimonieuse et Détection d'Anomalies Univariées

Pré-traitement



- **SEGMENTATION :**
Les signaux sont découpés de manière glissante en fenêtres de taille fixe W .

Estimation Parcimonieuse et Détection d'Anomalies Univariées¹

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

Si anomalie

Sinon

Élément nul
Élément non nul

- $\mathbf{y} \in \mathbb{R}^N$ est un signal test.
- $\Phi \in \mathbb{R}^{N \times L}$ est un dictionnaire appris à partir de données saines (K-SVD, ODL etc.).
- $\mathbf{x} \in \mathbb{R}^L$ est un vecteur parcimonieux de coefficients.
- $\mathbf{e} \in \mathbb{R}^N$ est un vecteur d'anomalie.
- $\mathbf{b} \in \mathbb{R}^N$ est un vecteur de bruit.

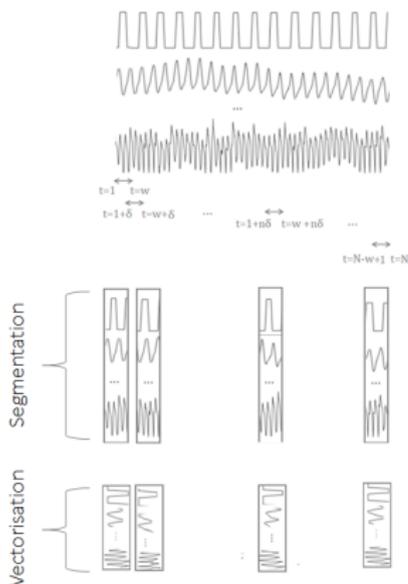
Problème de minimisation

$$\min_{\mathbf{x}, \mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \beta \|\mathbf{e}\|_2 \quad (1)$$

1. A. Adler and M. Elad and Y. Hel-Or and E. Rivlin, "Sparse Coding with Anomaly Detection" *Journal of Signal Processing Systems*, vol. 79, no.2, pp. 179-188, May 2015

Estimation Parcimonieuse et Détection d'Anomalies Multivariées² : ADDICT

Pré-traitement



- 1- SEGMENTATION :
Les signaux sont découpés de manière glissante en fenêtres de taille fixe W .
- 2- VECTORISATION :
Les matrices obtenues après segmentation sont vectorisées pour obtenir des vecteurs de données mixtes associées aux différents paramètres de télémétrie surveillés conjointement.

2. B. Pilastre, S. D'Escrivan, L. Boussouf and J-Y. Tourneret. "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning", Signal Processing, vol. 168, pp. 107320, March 2020

Estimation Parcimonieuse et Détection d'Anomalies Multivariées² : ADDICT

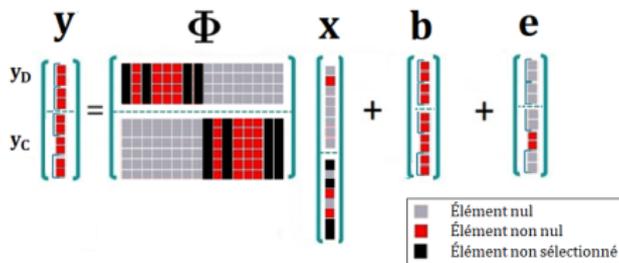
$$\begin{matrix} y_D \\ y_C \end{matrix} = \begin{matrix} \Phi \\ \Phi \end{matrix} \begin{matrix} x \\ x \end{matrix} + \begin{matrix} b \\ b \end{matrix} + \begin{matrix} e \\ e \end{matrix}$$

Legend:

- Élément nul
- Élément non nul
- Élément non sélectionné

2. B. Pilastre, S. D'Escrivan, L. Boussof and J-Y. Tourneret. "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning", Signal Processing, vol. 168, pp. 107320, March 2020

Estimation Parcimonieuse et Détection d'Anomalies Multivariées² : ADDICT



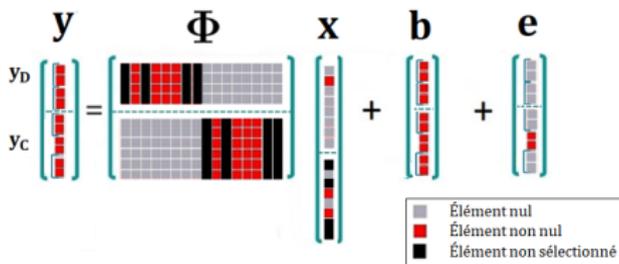
- Estimation parcimonieuse discrète

$$\min_{x_D \in \mathcal{B}, e_D \in \mathbb{R}^{N_D}} \|y_D - \Phi_D x_D - e_D\|_2^2 + b_D \sum_{k=1}^{K_D} \|e_{D,k}\|_2 \quad (2)$$

où \mathcal{B} est la base canonique ou base naturelle de \mathbb{R}^L

2. B. Pilastre, S. D'Escrivan, L. Boussouf and J-Y. Tourneret. "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning", Signal Processing, vol. 168, pp. 107320, March 2020

Estimation Parcimonieuse et Détection d'Anomalies Multivariées² : ADDICT



- Estimation parcimonieuse discrète

$$\min_{x_D \in \mathcal{B}, e_D \in \mathbb{R}^{N_D}} \|y_D - \Phi_D x_D - e_D\|_2^2 + b_D \sum_{k=1}^{K_D} \|e_{D,k}\|_2 \quad (2)$$

où \mathcal{B} est la base canonique ou base naturelle de \mathbb{R}^L

- Estimation parcimonieuse continue

$$\min_{x_C, e_C} \frac{1}{2} \|y_C - \Phi_{\mathcal{M}} x_C - e_C\|_2^2 + a_C \|x_C\|_1 + b_C \sum_{k=1}^{K_C} \|e_{C,k}\|_2 \quad (3)$$

où $\mathcal{M} = \{I \in \{1, \dots, L\} \mid \|\hat{e}_{D,I}\|_2 = 0\}$

2. B. Pilastre, S. D'Escrivan, L. Boussouf and J-Y. Tourneret. "Anomaly Detection in Mixed Telemetry Data Using a Sparse Representation and Dictionary Learning", Signal Processing, vol. 168, pp. 107320, March 2020

L'algorithme ADDICT : Estimation Parcimonieuse Continue

Problème de minimisation après intégration d'une variable auxiliaire \mathbf{z} et de la contrainte d'égalité $\mathbf{z} = \mathbf{x}_C$

$$\min_{\mathbf{x}_C, \mathbf{e}_C} \frac{1}{2} \|\mathbf{y}_C - \Phi_{\mathcal{M}} \mathbf{x}_C - \mathbf{e}_C\|_2^2 + a_C \|\mathbf{z}\|_1 + b_C \sum_{k=1}^{K_C} \|\mathbf{e}_{C,k}\|_2 \quad (4)$$

$$\text{s.c. } \mathbf{z} = \mathbf{x}_C \quad (5)$$

3. S. Boyd, "Distributed Optimization and Statistical Learning via Alternating Direction Method of Multipliers" *Foundations and Trends in Machine Learning*, vol. 3, pp. 1-222, 2011

L'algorithme ADDICT : Estimation Parcimonieuse Continue

Problème de minimisation après intégration d'une variable auxiliaire \mathbf{z} et de la contrainte d'égalité $\mathbf{z} = \mathbf{x}_C$

$$\min_{\mathbf{x}_C, \mathbf{e}_C} \frac{1}{2} \|\mathbf{y}_C - \Phi_{\mathcal{M}} \mathbf{x}_C - \mathbf{e}_C\|_2^2 + a_C \|\mathbf{z}\|_1 + b_C \sum_{k=1}^{K_C} \|\mathbf{e}_{C,k}\|_2 \quad (4)$$

$$\text{s.c. } \mathbf{z} = \mathbf{x}_C \quad (5)$$

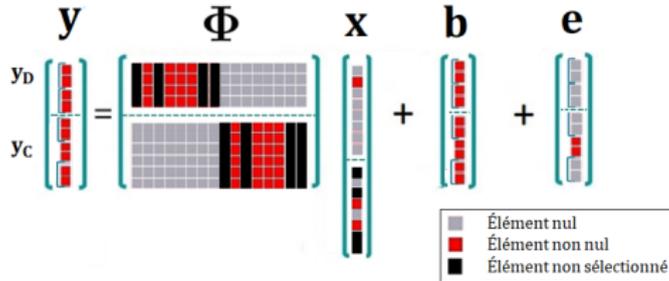
Mise à jour des vecteurs \mathbf{x}_C , \mathbf{e}_C et \mathbf{z} sont effectuées par minimisation du Lagrangien Augmenté à l'aide de l'algorithme ADMM³

$$\mathcal{L}_A(\mathbf{x}_C, \mathbf{z}, \mathbf{e}_C, \mathbf{m}, \mu) = \frac{1}{2} \|\mathbf{y}_C - \Phi_{\mathcal{M}} \mathbf{x}_C - \mathbf{e}_C\|_2^2 + a_C \|\mathbf{z}\|_1 \quad (6)$$

$$+ b_C \sum_{k=1}^{K_C} \|\mathbf{e}_{C,k}\|_2 + \mathbf{m}_C^T (\mathbf{z} - \mathbf{x}_C) + \frac{\mu_C}{2} \|\mathbf{z} - \mathbf{x}_C\|_2^2 \quad (7)$$

3. S. Boyd, "Distributed Optimization and Statistical Learning via Alternating Direction Method of Multipliers" *Foundations and Trends in Machine Learning*, vol. 3, pp. 1-222, 2011

L'algorithme ADDICT : Détection d'Anomalies



L'algorithme ADDICT : Détection d'Anomalies

$$\begin{array}{c} \mathbf{y} \\ y_D \\ y_C \end{array} = \begin{array}{c} \Phi \\ \begin{array}{c} \text{[Red blocks]} \\ \text{[Grey blocks]} \end{array} \end{array} \begin{array}{c} \mathbf{x} \\ \begin{array}{c} \text{[Red]} \\ \text{[Grey]} \\ \text{[Black]} \end{array} \end{array} + \begin{array}{c} \mathbf{b} \\ \begin{array}{c} \text{[Red]} \\ \text{[Grey]} \end{array} \end{array} + \begin{array}{c} \mathbf{e} \\ \begin{array}{c} \text{[Grey]} \\ \text{[Red]} \end{array} \end{array}$$

Élément nul
 Élément non nul
 Élément non sélectionné

Score d'anomalie

$$s(\mathbf{y}) = \begin{cases} -1 & \text{si } \|\hat{\mathbf{e}}_D\|_2 > 0 \text{ (i.e., } \mathcal{M} = \emptyset) \\ \|\hat{\mathbf{e}}_C\|_2 & \text{sinon.} \end{cases}$$

L'algorithme ADDICT : Détection d'Anomalies

$$\begin{array}{c} \mathbf{y} \\ y_D \\ y_C \end{array} = \begin{array}{c} \Phi \\ \begin{array}{c} \text{[Red blocks in top-left and bottom-right]} \\ \text{[Grey blocks in top-right and bottom-left]} \end{array} \end{array} \begin{array}{c} \mathbf{x} \\ \begin{array}{c} \text{[Red blocks in top-left and bottom-right]} \\ \text{[Grey blocks in top-right and bottom-left]} \end{array} \end{array} + \begin{array}{c} \mathbf{b} \\ \begin{array}{c} \text{[Red blocks in top-left and bottom-right]} \\ \text{[Grey blocks in top-right and bottom-left]} \end{array} \end{array} + \begin{array}{c} \mathbf{e} \\ \begin{array}{c} \text{[Red blocks in top-left and bottom-right]} \\ \text{[Grey blocks in top-right and bottom-left]} \end{array} \end{array}$$

Élément nul
 Élément non nul
 Élément non sélectionné

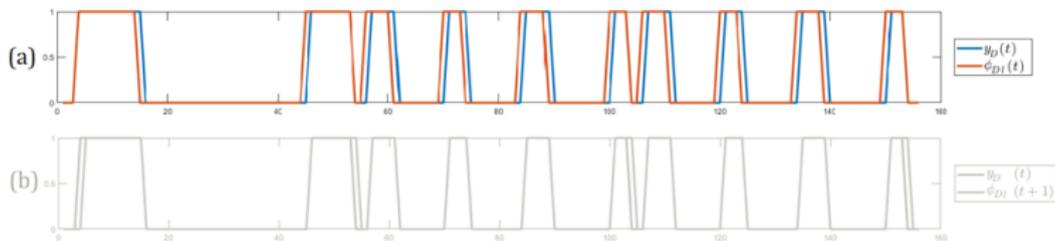
Score d'anomalie

$$s(\mathbf{y}) = \begin{cases} -1 & \text{si } \|\hat{\mathbf{e}}_D\|_2 > 0 \text{ (i.e., } \mathcal{M} = \emptyset) \\ \|\hat{\mathbf{e}}_C\|_2 & \text{sinon.} \end{cases}$$

Règle de décision

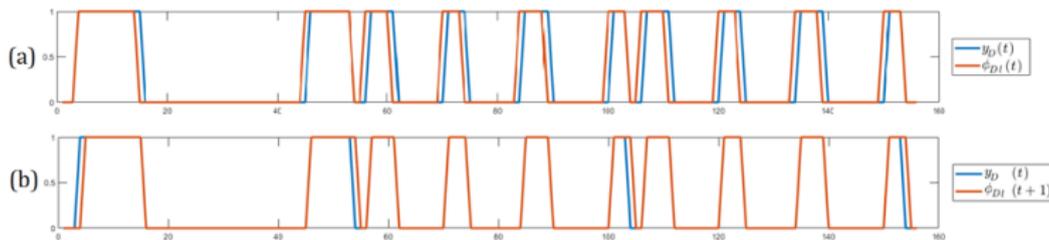
$$\text{Anomalie détectée si } \begin{cases} s(\mathbf{y}) = -1 & \text{(anomalie discrète)} \\ \text{ou} \\ s(\mathbf{y}) > S_{\text{PFA}} & \text{(anomalie continue/multivariée)} \end{cases}$$

L'algorithme ADDICT : Option d'Invariance par Translation ²



2. Q. Barthélemy, C. Gouy-Pailler, C. Isaac, A. Souloumiac, A. Larue, J.I. Mars "Multivariate Temporal Dictionary for EEG" *Journal of Neuroscience Methods*, vol. 215, pp. 19-28, 2013

L'algorithme ADDICT : Option d'Invariance par Translation²

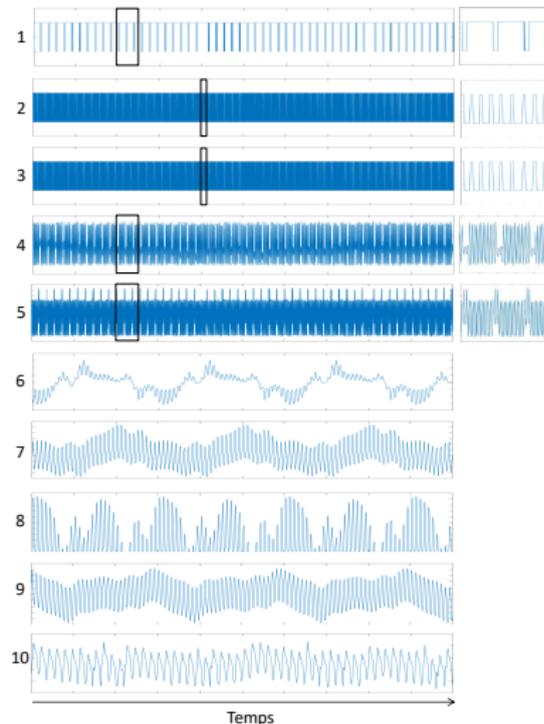


- Fonctionnement :
 - Décaler tous les atomes discrets de τ instants.
 - $\tau \in \{-\tau_{\max}, \dots, -1, 0, 1, \dots, \tau_{\max}\}$
- Intérêts :
 - Améliorer l'estimation.
 - Limiter les fausses alarmes.

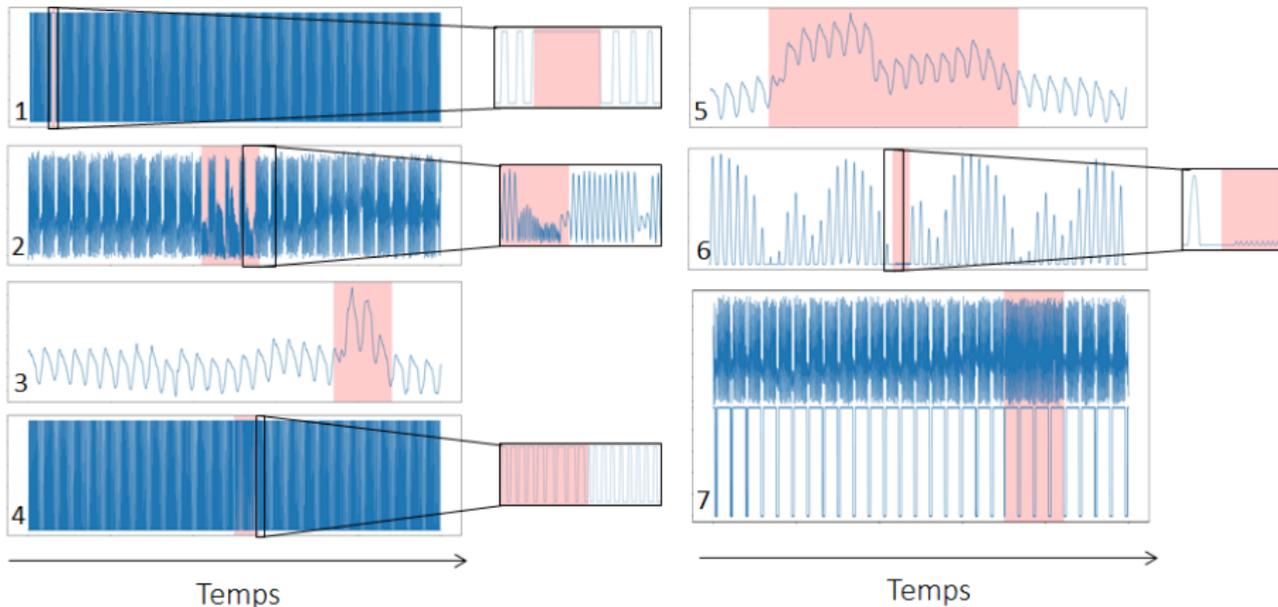
2. Q. Barthélemy, C. Gouy-Pailler, C. Isaac, A. Souloumiac, A. Larue, J.I. Mars "Multivariate Temporal Dictionary for EEG" *Journal of Neuroscience Methods*, vol. 215, pp. 19-28, 2013

Résultats Expérimentaux : Les Données d'Apprentissage

- 10 paramètres : 7 continus et 3 discrets.
- 2 mois de données saines : 30000 signaux d'apprentissage avec $w = 50$ (~ 30 minutes).
- $L = 2000$ atomes appris.



Résultats Expérimentaux : Les Données de Test



- 18 jours de télésurveillance : 1000 signaux de test.
- 90 signaux affectés associés à 7 périodes d'anomalie de durées variables.

Résultats Expérimentaux : Méthodes Concurrentes

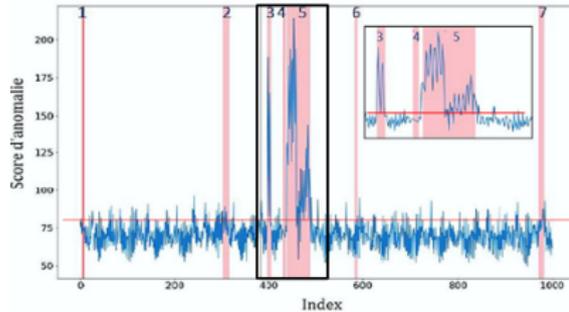
- MPPCAD⁴ (détection multivariée)
 - Données d'entrée : Vecteurs de données mixtes.
 - Approche : Probabiliste.
- OC-SVM⁵ (détection multivariée)
 - Données d'entrée : Vecteurs de données mixtes.
 - Approche : Machines à vecteurs supports.
- NOSTRADAMUS⁶ (détection univariée)
 - Données d'entrée : Vecteurs de descripteurs (moyenne, min, max, variance, etc.).
 - Approche : OC-SVM.

4. T. Yairi and *al.*, "A Data-Driven Health Monitoring Method for Satellite Housekeeping Data Based on Probabilistic Clustering and Dimensionality Reduction", *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no 3, pp. 1384-1401, Jun. 2017.

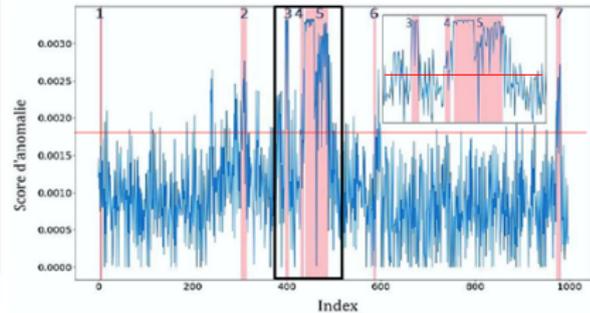
5. B. Schölkopf and *al.*, "Estimating the support of a high-dimensional distribution", *Neural Computation*, vol. 3, no 7, pp. 1443-1471, 2001.

6. S. Fuertes and *al.*, "Improving Spacecraft Health Monitoring with Automatic Anomaly Detection Techniques", *Proc. Int. Conf. Space Operations (SpaceOps'2016)*, South Korea, May 2016.

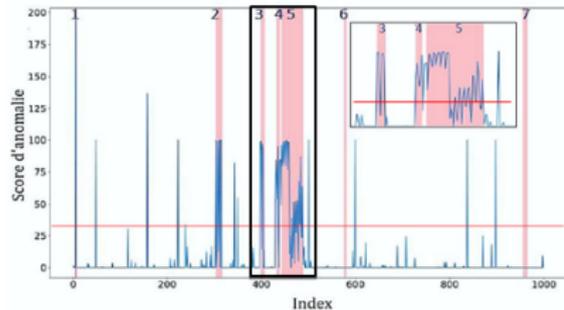
Résultats Expérimentaux : Étude Qualitative



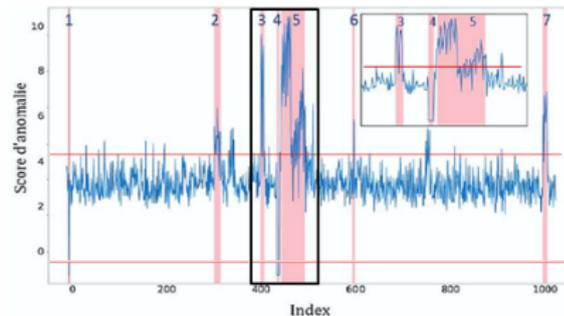
(b) MPPCAD



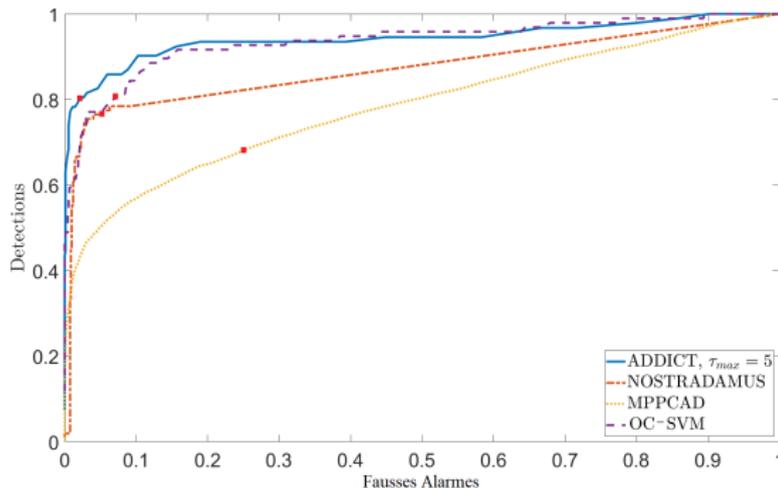
(c) OC-SVM



(a) NOSTRADAMUS

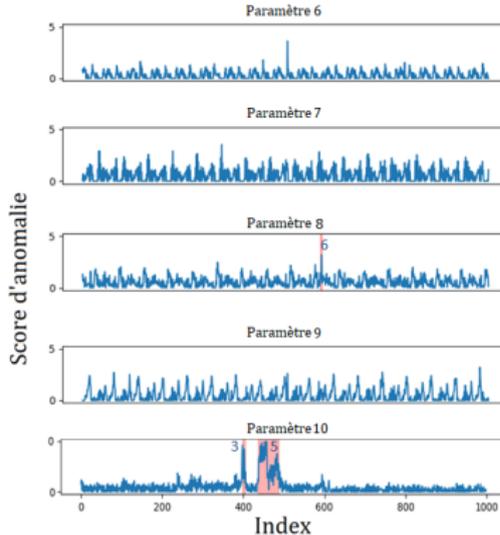
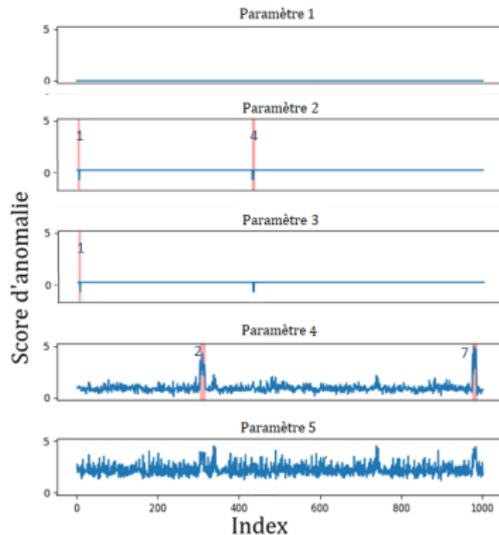
(d) ADDICT ($\tau = 5$)

Résultats Expérimentaux : Étude Quantitative

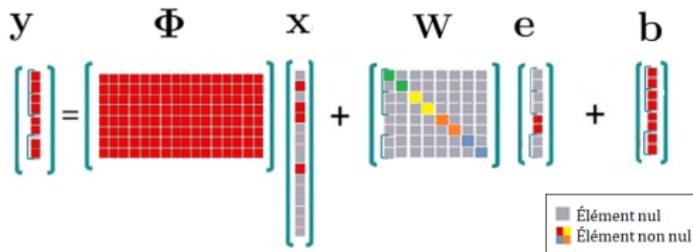


Méthode	Seuil	P_D	P_{FA}
OC-SVM	0.0018	80.85%	7%
MPPCAD	12	67%	25%
NOSTRADAMUS	29	77.26%	6%
ADDICT ($\tau_{max} = 5$)	4.2	80%	3%

Résultats Expérimentaux : Analyse Univariée



Les Extensions ADDICT : W-ADDICT⁷ et G-ADDICT

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


Élément nul
 Élément non nul

W-ADDICT

$$\min_{\mathbf{x}, \mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{x}\|_1 + \sum_{k=1}^K w_k \|\mathbf{e}_k\|_2 \quad (8)$$

- Poids univariés.
- Mise à jour à chaque itération.
- Intégration d'informations externes (ex : coefficient de corrélation).

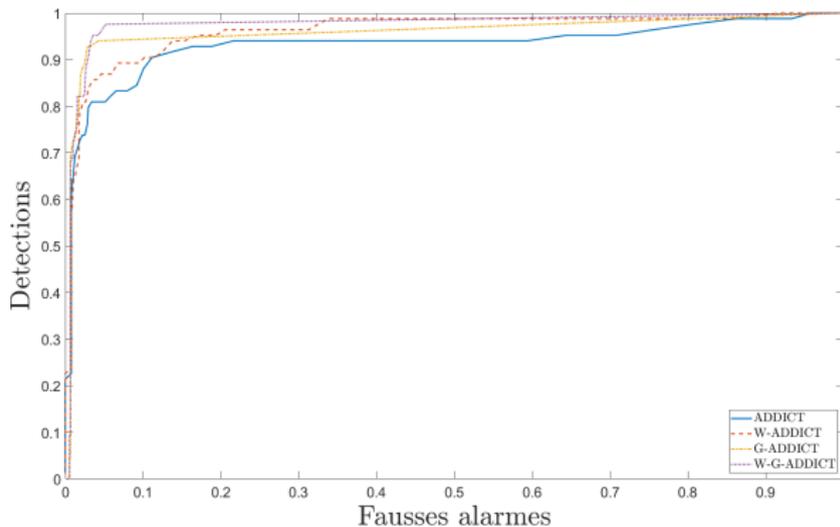
G-ADDICT

$$\min_{\mathbf{x}, \mathbf{e}} \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x} - \mathbf{e}\|_2^2 + a \|\mathbf{x}\|_1 + \sum_{k=1}^K b_k \|\mathbf{e}_k\|_2 \quad (9)$$

- Hyperparamètres univariés fixés.
- Généralisation du modèle.
- Estimation automatique des b_k (ex : OC-SDT).

7. B. Pilastre, S. D'Escrivan, L. Boussouf and J-Y. Tournet. "Spacecraft Health Monitoring using a Weighted Sparse Decomposition", World Congress on Condition Monitoring (WCCM'2019), Singapore, December 2019.

Résultats Expérimentaux



Method	Threshold	P_D	P_{FA}
ADDICT	4.2	80.85%	3%
W-ADDICT	4.5	85%	2.7%
G-ADDICT	ϵ	93.6%	3.3%
W-G-ADDICT	ϵ	97.6%	6%

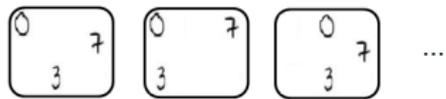
Contributions

- Détection d'anomalies multivariées à l'aide de l'algorithme ADDICT.
- Traitement conjoint de données mixtes.
- Des extensions disponibles.
- Un réglage automatique des hyperparamètres.

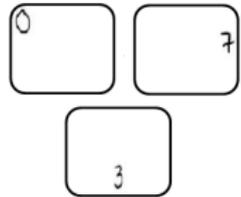
Estimation Parcimonieuse Convolutive

Estimation Parcimonieuse Convolutive : Introduction

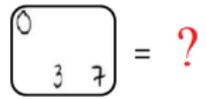
Données d'Apprentissage



Atomes appris

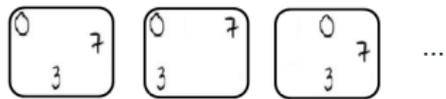


Estimation Parcimonieuse

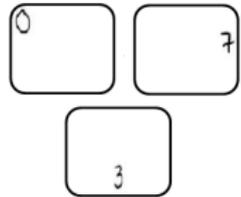


Estimation Parcimonieuse Convolutive : Introduction

Données d'Apprentissage



Atomes appris

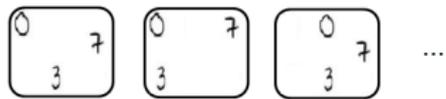


Estimation Parcimonieuse

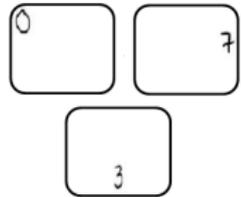
$$\begin{array}{|c|} \hline 0 \\ \hline 3 \quad 7 \\ \hline \end{array} = 1 \begin{array}{|c|} \hline 0 \\ \hline \quad \quad \\ \hline \end{array} + 1 \begin{array}{|c|} \hline \quad \quad \\ \hline 3 \\ \hline \end{array} + ? \begin{array}{|c|} \hline \quad \quad \\ \hline \quad \quad 7 \\ \hline \end{array}$$

Estimation Parcimonieuse Convulsive : Introduction

Données d'Apprentissage



Atomes appris

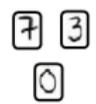


Estimation Parcimonieuse

$$\begin{array}{|c|} \hline 0 \\ \hline 3 \quad 7 \\ \hline \end{array} = 1 \begin{array}{|c|} \hline 0 \\ \hline \quad \quad \\ \hline \end{array} + 1 \begin{array}{|c|} \hline \quad \quad \\ \hline 3 \\ \hline \end{array} + ? \begin{array}{|c|} \hline \quad \quad \\ \hline \quad \quad 7 \\ \hline \end{array}$$

Estimation Parcimonieuse Convulsive

Filtres appris

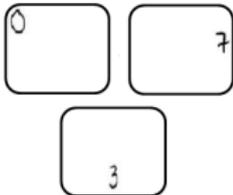


Estimation Parcimonieuse Convulsive : Introduction

Données d'Apprentissage



Atomes appris



Estimation Parcimonieuse

$$\begin{matrix} 0 \\ 3 \end{matrix} \begin{matrix} 7 \\ \end{matrix} = 1 \begin{matrix} 0 \\ \end{matrix} + 1 \begin{matrix} \end{matrix} \begin{matrix} 3 \\ \end{matrix} + ? \begin{matrix} \end{matrix} \begin{matrix} 7 \\ \end{matrix}$$

Estimation Parcimonieuse Convulsive

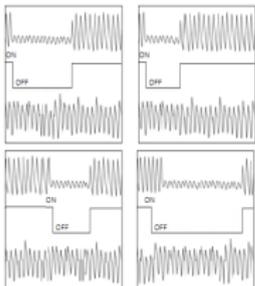
$$\begin{matrix} 0 \\ 3 \end{matrix} \begin{matrix} 7 \\ \end{matrix} = \begin{matrix} 0 \\ \end{matrix} * \begin{matrix} 1 \\ \end{matrix} + \begin{matrix} 7 \\ \end{matrix} * \begin{matrix} \end{matrix} \begin{matrix} 1 \\ \end{matrix} + \begin{matrix} 3 \\ \end{matrix} * \begin{matrix} \end{matrix} \begin{matrix} 1 \\ \end{matrix}$$

Filtres appris

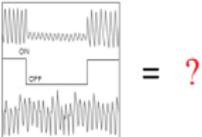


Estimation Parcimonieuse Convolutive : Introduction

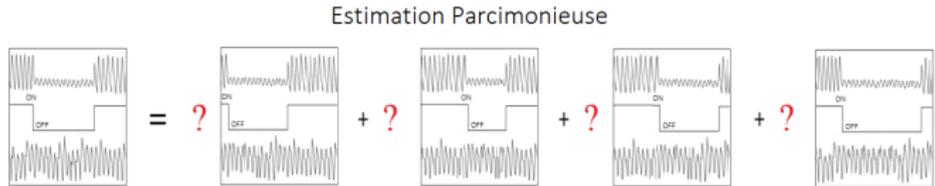
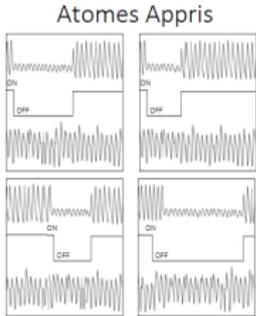
Atomes Appris



Estimation Parcimonieuse



Estimation Parcimonieuse Convolutive : Introduction



L'algorithme C-ADDICT⁸ : Formulation du Problème

$$\mathbf{Y} = \sum \mathbf{\Phi} * \mathbf{X} + \mathbf{B} + \mathbf{E}$$



- $\mathbf{Y} \in \mathbb{R}^{K \times W}$ est une matrice de signaux tests.
- $\mathbf{\Phi} \in \mathbb{R}^{K \times M \times L}$ est un dictionnaire convolutif appris à partir de données saines.
- $\mathbf{X} \in \mathbb{R}^{K \times W \times L}$ est une matrice parcimonieuse de coefficients.
- $\mathbf{E} \in \mathbb{R}^{K \times W}$ est une matrice d'anomalies.
- $\mathbf{B} \in \mathbb{R}^{K \times W}$ est une matrice de bruit.

Problème de minimisation

$$\min_{\mathbf{x}_l, \mathbf{E}} \frac{1}{2} \left\| \sum_l \mathbf{\Phi}_l * \mathbf{x}_l + \mathbf{E} - \mathbf{Y} \right\|_2^2 + \lambda \sum_l \|\mathbf{x}_l\|_1 + \beta \sum_{k,n} \|\mathbf{E}_{k,n}\|_2 \quad (10)$$

⁸ B. Pilastre, G. Silva, S. D'Escrivan, L. Boussouf, P. Rodriguez et J-Y. Tourneret. "Anomaly Detection in Mixed Time-Series using a Convolutional Sparse Representation with Application to Spacecraft Health Monitoring, Proc. of IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP'2020), Barcelona, Spain, May 2020.

L'algorithme C-ADDICT⁸ : Formulation du Problème

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} \mathbf{x} \xrightarrow{\text{équivalent à}} \begin{bmatrix} \mathbf{y}_1 = \Phi_1 \mathbf{x}_1 \\ \mathbf{y}_2 = \Phi_2 \mathbf{x}_2 \end{bmatrix} \text{ s.c. } \mathbf{x}_1 = \mathbf{x}_2. \quad (11)$$

Modèle Découplé

$$\min_{\mathbf{x}_{k,l}, \mathbf{e}_k} \frac{1}{2} \sum_k \left\| \sum_l \phi_{k,l} * \mathbf{x}_{k,l} + \mathbf{e}_k - \mathbf{y}_k \right\|_2^2 + \lambda \sum_{k,l} \|\mathbf{x}_{k,l}\|_1 + \beta \sum_{k,n} \|\mathbf{e}_{k,n}\|_2 \quad (12)$$

$$\text{s.c. } \mathbf{x}_{1,l} = \mathbf{x}_{2,l} = \dots = \mathbf{x}_{K,l}, \forall l \quad (13)$$

⁸ B. Pilastre, G. Silva, S. D'Escrivan, L. Boussouf, P. Rodriguez et J-Y. Tourneret. "Anomaly Detection in Mixed Time-Series using a Convolutional Sparse Representation with Application to Spacecraft Health Monitoring, Proc. of IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP'2020), Barcelona, Spain, May 2020.

L'algorithme C-ADDICT : Résolution du Problème

$$\begin{aligned}
 \begin{pmatrix} y_k \\ \vdots \end{pmatrix} &= \sum \left(\begin{pmatrix} \Phi_k \\ \vdots \end{pmatrix} * \begin{pmatrix} x_k \\ \vdots \end{pmatrix} \right) + \begin{pmatrix} b_k \\ \vdots \end{pmatrix} &\xrightarrow{\text{Equivalent à}} & \begin{pmatrix} y_k \\ \vdots \end{pmatrix} &= & \begin{pmatrix} \Phi_k \\ \vdots \end{pmatrix} \begin{pmatrix} x_k \\ \vdots \end{pmatrix} + \begin{pmatrix} b_k \\ \vdots \end{pmatrix}
 \end{aligned}$$

The diagram illustrates the transformation of a sparse convolutional model into a matrix-vector form. On the left, the vector y_k is shown as a sum of convolutions between filter vectors Φ_k and input vectors x_k , plus a bias vector b_k . The convolution operation is represented by a small grid with colored elements. An arrow labeled "Equivalent à" points to the right, where the same model is represented as a matrix multiplication. The matrix Φ_k is a large grid with a sparse, banded structure, where non-zero elements are highlighted in blue, yellow, and red. The input vector x_k and bias vector b_k are shown as tall vertical vectors.

L'algorithme C-ADDICT : Résolution du Problème

$$\begin{pmatrix} \mathbf{y}_k \end{pmatrix} = \sum \left(\begin{pmatrix} \Phi_k \end{pmatrix} * \begin{pmatrix} \mathbf{x}_k \end{pmatrix} \right) + \begin{pmatrix} \mathbf{b}_k \end{pmatrix} \xrightarrow{\text{Equivalent à}} \begin{pmatrix} \mathbf{y}_k \end{pmatrix} = \begin{pmatrix} \Phi_k \end{pmatrix} \begin{pmatrix} \mathbf{x}_k \end{pmatrix} + \begin{pmatrix} \mathbf{b}_k \end{pmatrix}$$

Le problème devient

$$\min_{\mathbf{x}_k, \mathbf{e}_k} \frac{1}{2} \sum_k \|\phi_k \mathbf{x}_k + \mathbf{e}_k - \mathbf{y}_k\|_2^2 + \lambda \sum_k \|\mathbf{x}_k\|_1 + \beta \sum_{k,n} \|\mathbf{e}_{k,n}\|_2 \quad (14)$$

$$\text{s.c. } \mathbf{x}_1 = \mathbf{x}_2 = \dots = \mathbf{x}_K, \quad (15)$$

L'Algorithme C-ADDICT : Détection d'Anomalies

$$\mathbf{Y} = \sum_{\Phi} \mathbf{\Phi} * \mathbf{X} + \mathbf{B} + \mathbf{E}$$

■ Élément nul
 ■ Élément non nul

Score d'anomalie

$$s(\mathbf{Y}_n) = \|\hat{\mathbf{E}}_n\|_F^2.$$

L'Algorithme C-ADDICT : Détection d'Anomalies

$$\mathbf{Y} = \sum \Phi * \mathbf{X} + \mathbf{B} + \mathbf{E}$$

Élément nul
 Élément non nul

Score d'anomalie

$$s(\mathbf{Y}_n) = \|\hat{\mathbf{E}}_n\|_F^2.$$

Règle de décision

Anomalie détectée si $s(\mathbf{Y}_n) > S_{\text{PFA}}$

L'Algorithme C-ADDICT : Apprentissage de Dictionnaire

Problème de minimisation

$$\hat{\mathbf{x}}_{k,l}, \hat{\phi}_{k,l} = \arg \min_{\mathbf{x}_{k,l}, \phi_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_{t,k} \left\| \sum_l \phi_{k,l} * \mathbf{x}_{t,k,l} - \mathbf{y}_{t,k} \right\|_2^2 + \lambda \sum_{t,k,l} \|\mathbf{x}_{t,k,l}\|_1 \quad (18)$$

$$\text{s.c. } \mathbf{x}_{t,1,l} = \mathbf{x}_{t,2,l} = \dots = \mathbf{x}_{t,K,l}, \forall t, l \quad (19)$$

L'Algorithme C-ADDICT : Apprentissage de Dictionnaire

Problème de minimisation

$$\hat{\mathbf{x}}_{k,l}, \hat{\phi}_{k,l} = \arg \min_{\mathbf{x}_{k,l}, \phi_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_{t,k} \left\| \sum_l \phi_{k,l} * \mathbf{x}_{t,k,l} - \mathbf{y}_{t,k} \right\|_2^2 + \lambda \sum_{t,k,l} \|\mathbf{x}_{t,k,l}\|_1 \quad (18)$$

$$\text{s.c. } \mathbf{x}_{t,1,l} = \mathbf{x}_{t,2,l} = \dots = \mathbf{x}_{t,K,l}, \forall t, l \quad (19)$$

Résolution Alternée

ADMM

$$\hat{\mathbf{x}}_{t,k,l} = \arg \min_{\mathbf{x}_{t,k,l}} \frac{1}{2} \sum_{t,k} \left\| \sum_l \phi_{k,l} * \mathbf{x}_{t,k,l} - \mathbf{y}_{t,k} \right\|_2^2 + \lambda \sum_{t,k} \|\mathbf{x}_{t,k}\|_1 \quad (20)$$

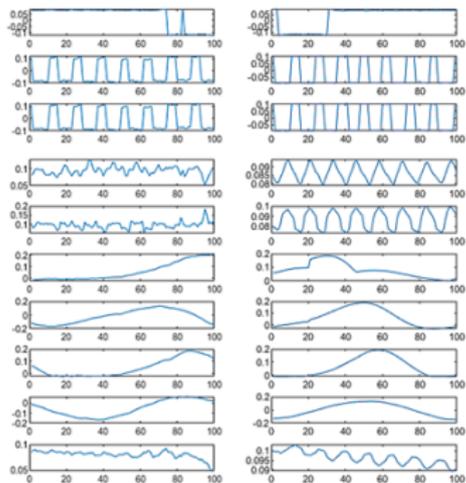
$$\text{s.c. } \mathbf{x}_{t,1,l} = \mathbf{x}_{t,2,l} = \dots = \mathbf{x}_{t,K,l}, \forall t, l \quad (21)$$

Descente de
Gradient

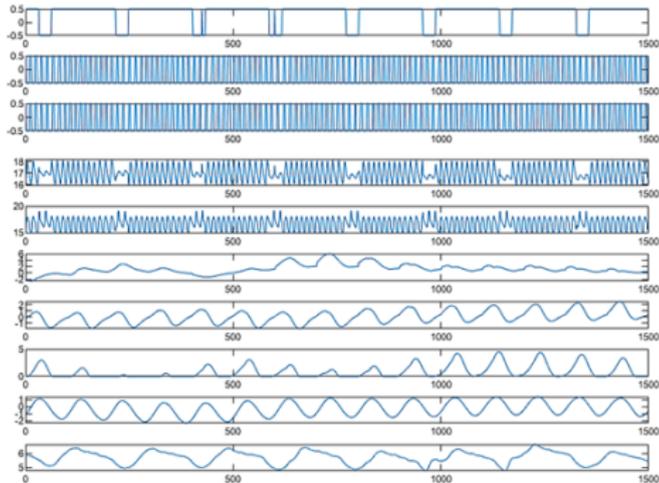
$$\hat{\phi}_{k,l} = \arg \min_{\phi_{k,l} \in \mathcal{C}} \frac{1}{2} \sum_{t,k} \left\| \sum_l \phi_{k,l} * \mathbf{x}_{t,k,l} - \mathbf{y}_{t,k} \right\|_2^2 \quad (22)$$

Résultats Expérimentaux : Dictionnaire

Filtres appris

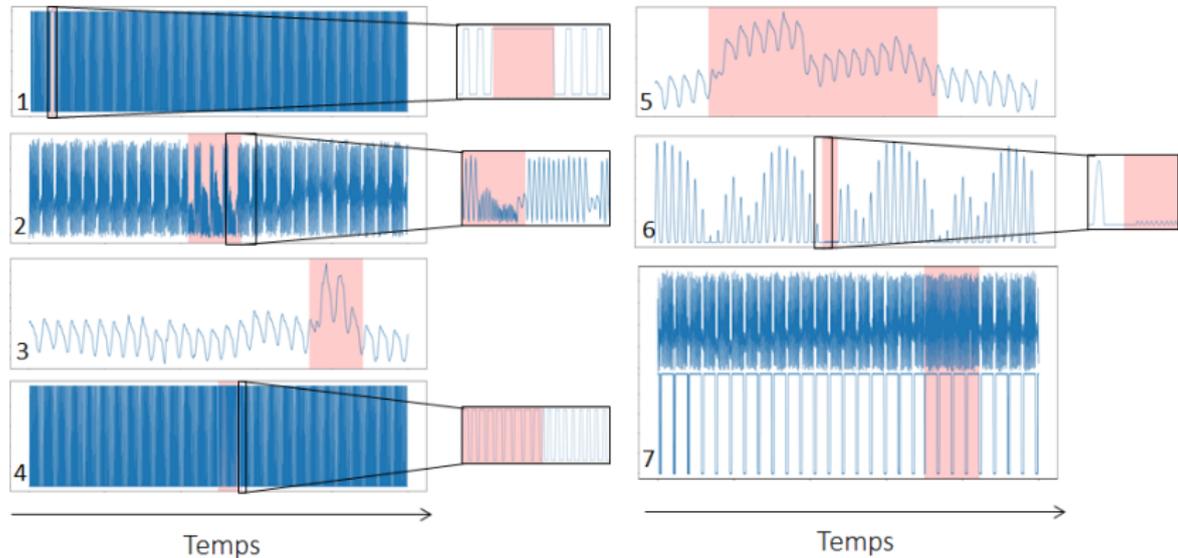


Signaux d'apprentissage



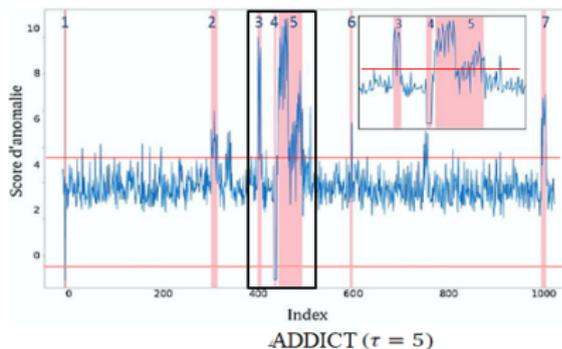
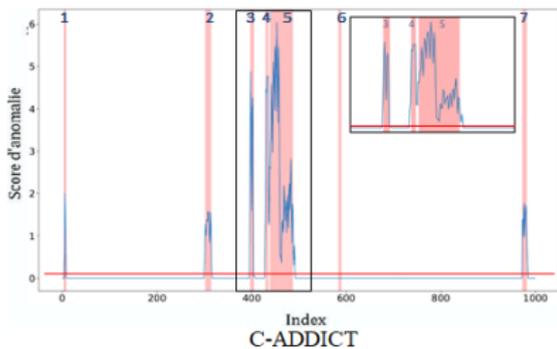
- $L = 100$ filtres appris.
- Des filtres de taille $M = 50$ (30 minutes) appris à partir de 2 mois de données saines.

Résultats Expérimentaux : Les Données de Test

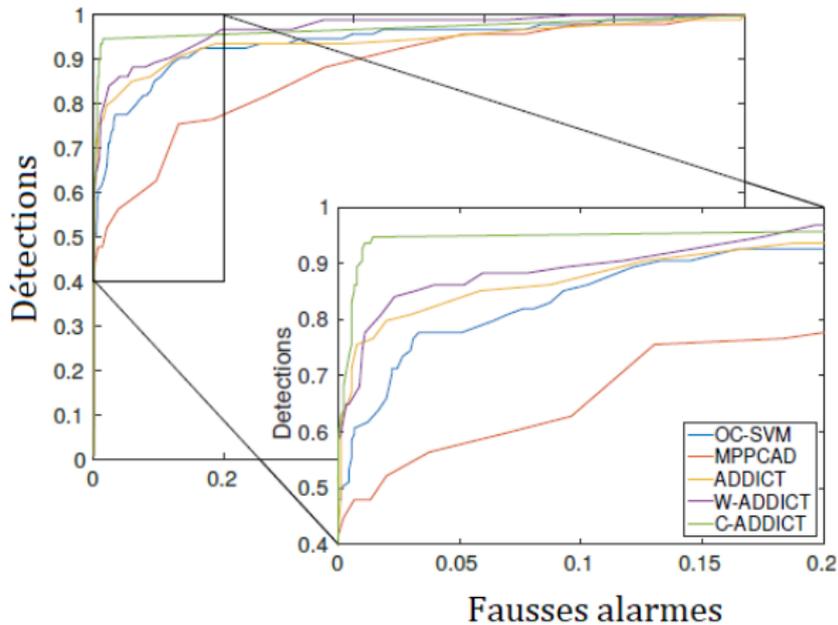


- 18 jours de télémessure.
- 7 périodes d'anomalies de durées variables.

Résultats Expérimentaux : Étude Qualitative



Résultats Expérimentaux : Étude Quantitative



Résultats Expérimentaux : Étude Quantitative

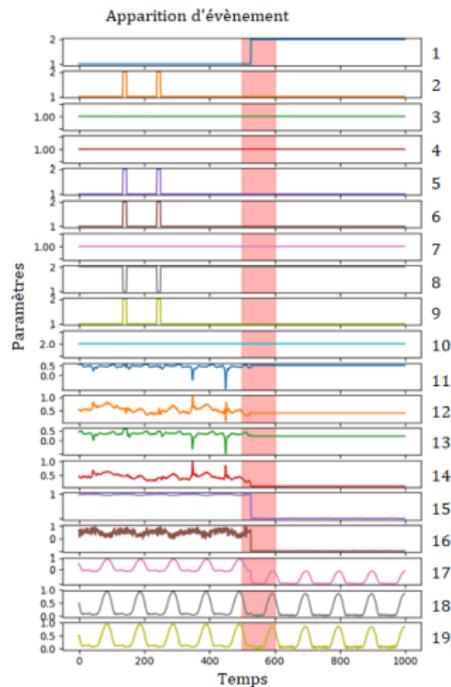
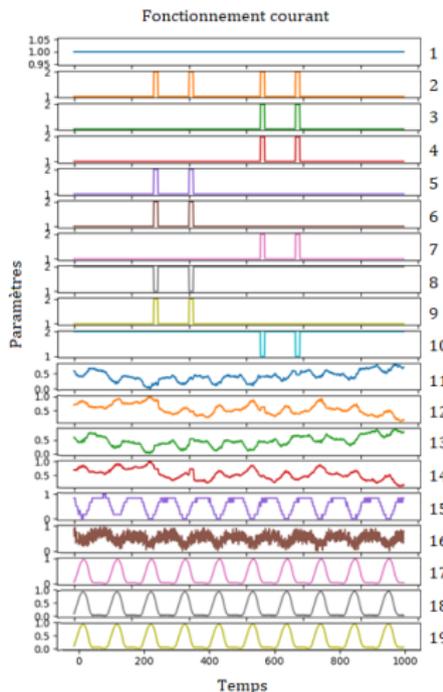
Méthode	Seuil	P_D	P_{FA}	AUC
OC-SVM	0.019	80.9%	7%	0.9413
MPPCAD	76	81.9%	25.9%	0.8779
ADDICT	4.1	81%	3%	0.937
W-ADDICT	4.5	85.1%	2.7%	0.9703
C-ADDICT	$\epsilon > 0$	94.7%	1.7%	0.9706

Contributions

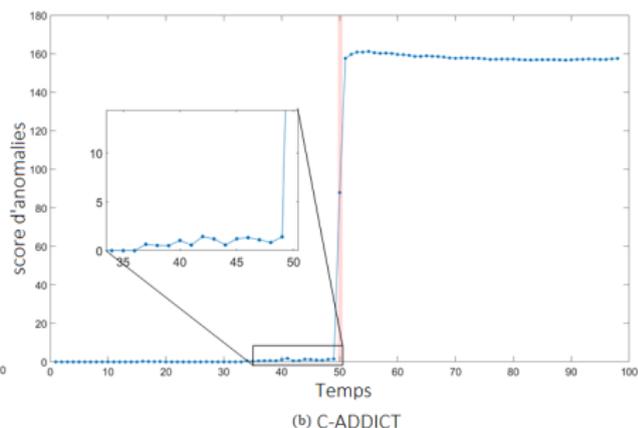
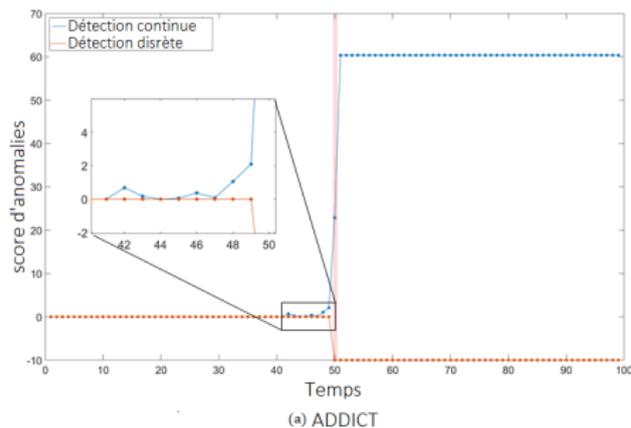
- Détection d'anomalies multivariées invariante par translation à l'aide de l'algorithme C-ADDICT.
- Traitement conjoint de données mixtes.
- Apprentissage automatique du dictionnaire convolutif.

Applications Industrielles

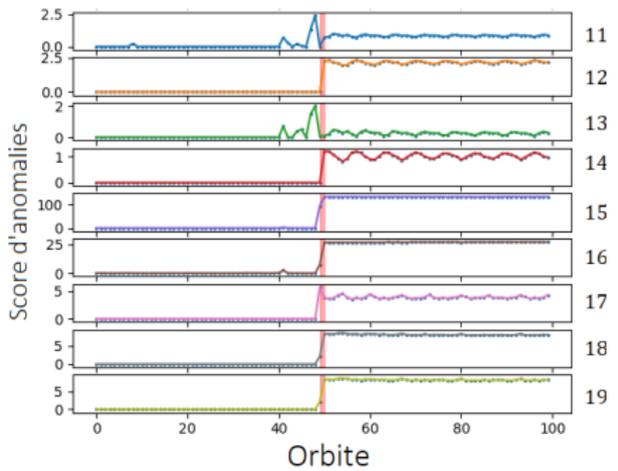
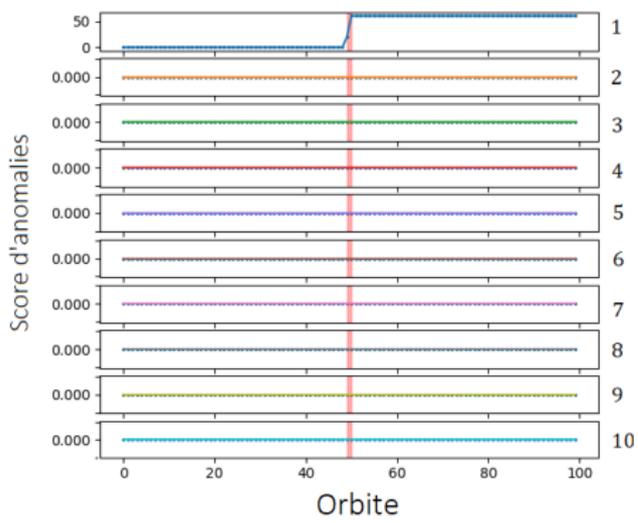
Premier Cas d'Usage : Détection d'un évènement connu



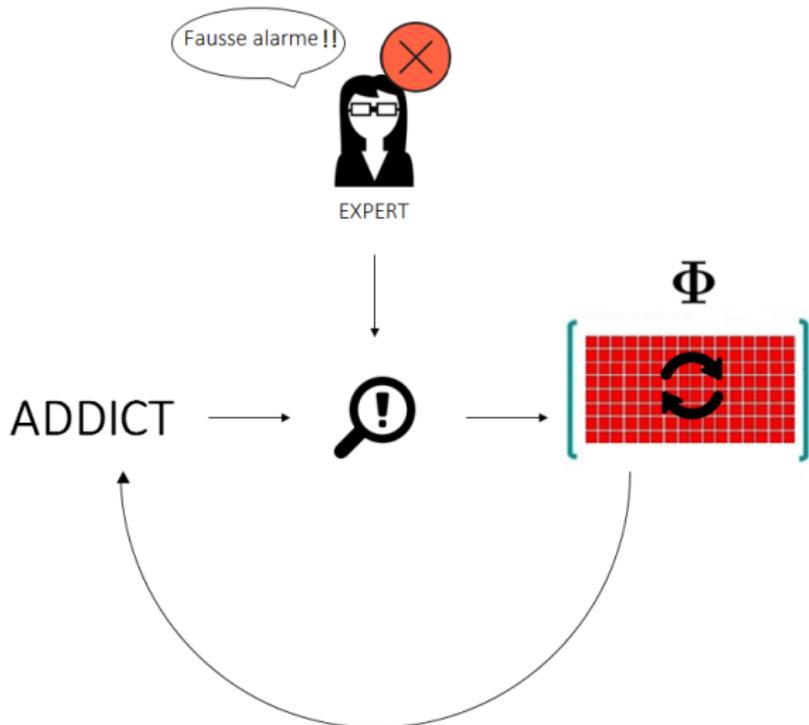
Premier Cas d'Usage : Détection ADDICT et C-ADDICT



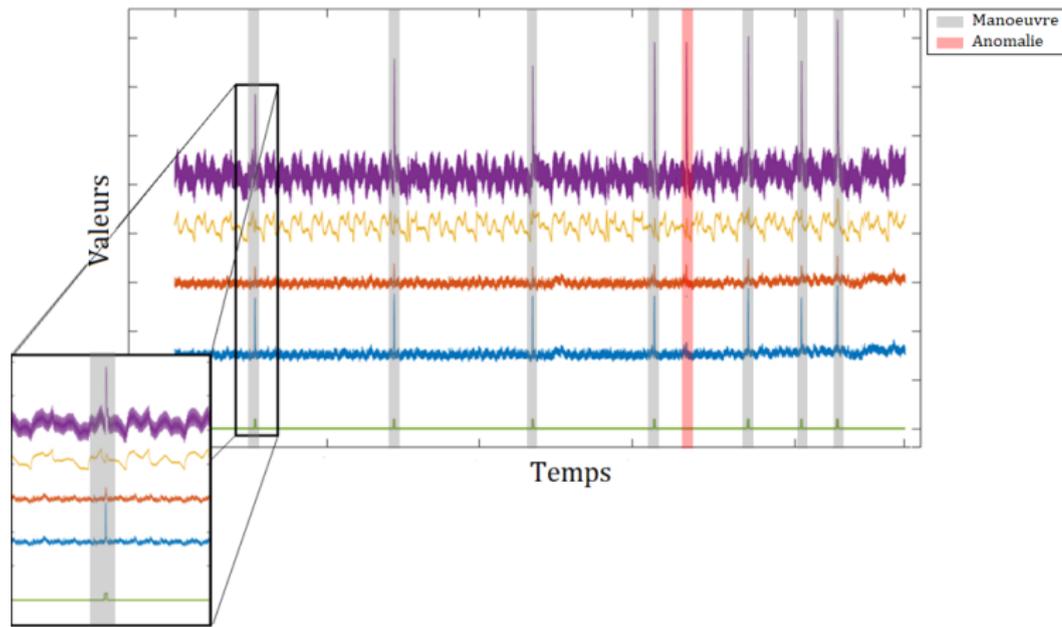
Premier Cas d'Usage : Détection ADDICT et C-ADDICT



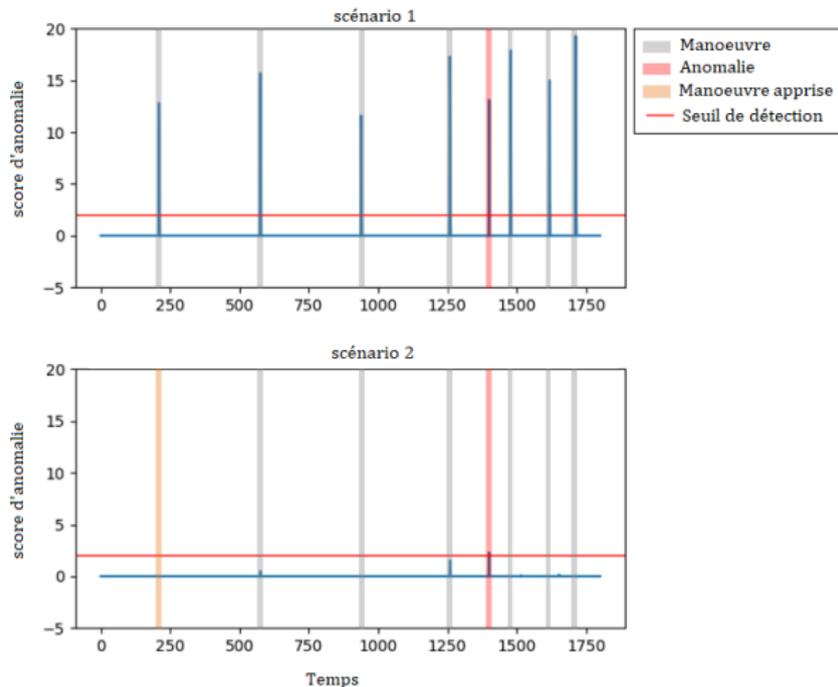
Deuxième Cas d'Usage : Détection Séquentielle



Deuxième Cas d'Usage : Manoeuvre de Correction d'Orbite



Deuxième Cas d'Usage : Détection C-ADDICT



Conclusion & Perspectives

Détection d'anomalies ADDICT

Avantages

- Traitement conjoint de données hétérogènes et mixtes.
- Détection d'anomalies univariées et multivariées.
- Extensions disponibles pour une meilleure détection.

Inconvénients

- Réglage d'hyperparamètres et d'un seuil de détection.
- Absence de méthode adaptée d'apprentissage de dictionnaires.

Perspectives

- Unification de la stratégie d'estimation parcimonieuse.
- Apprentissage de dictionnaire de données mixtes.
- Recherche de nouvelles pondérations.
- Réglage conjoint des hyperparamètres par une méthode automatique.
- Adaptation à la détection séquentielle.

Détection d'anomalies C-ADDICT

Avantages

- Traitement conjoint de données hétérogènes et mixtes.
- Détection d'anomalies univariées et multivariées.
- Apprentissage de dictionnaires convolutifs adaptés à la détection d'anomalies.
- Invariance par translation.

Inconvénients

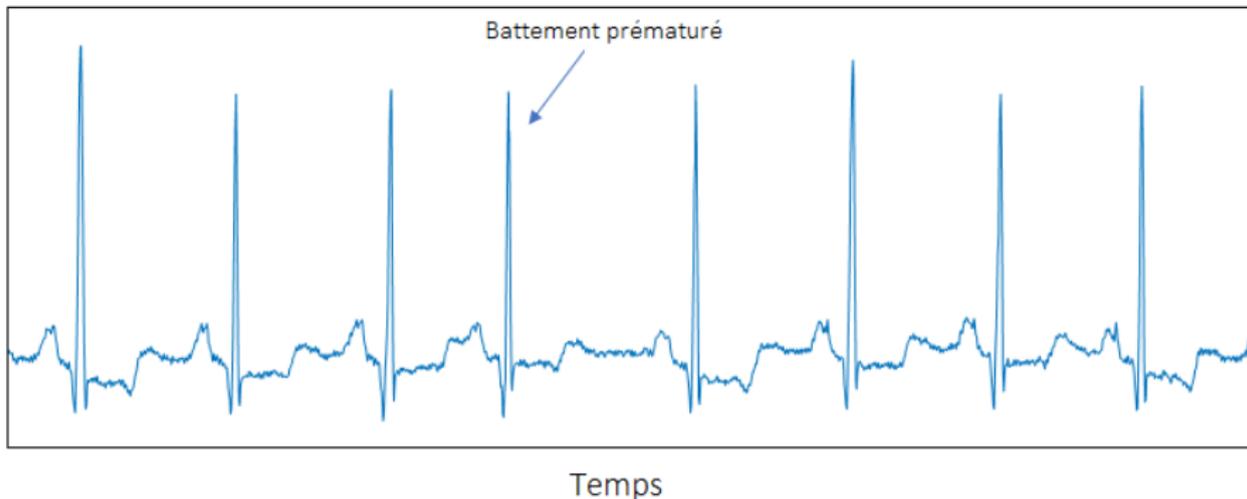
- Réglage d'hyperparamètres et d'un seuil de détection.
- Complexité calculatoire élevée.

Perspectives

- Adaptation à la détection séquentielle.
- Réglage conjoint des hyperparamètres par une méthode automatique.
- Limitation de la complexité calculatoire.
- Intégration d'une pondération.

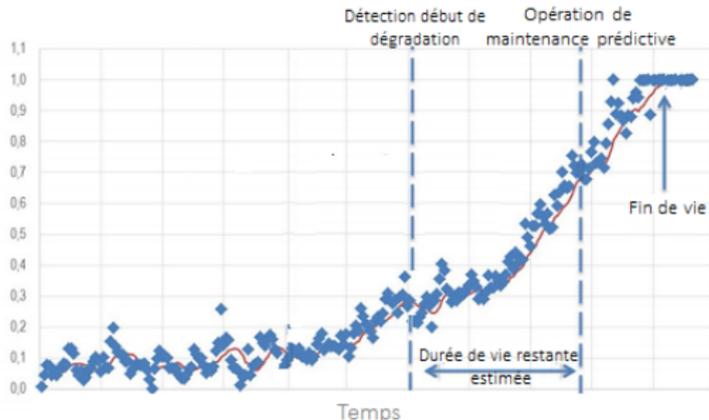
Détection d'Anomalies Temporelle⁸

Exemple d'une anomalie temporelle dans un signal ECG



⁸ E. Keogh, J. Lin and A. Fu. "HOT SAX : Efficiently Finding the Most Unusual Time Series Subsequence." In Proc. of the 5th IEEE Int Conf on Data Mining (ICDM'2005), pp. 226-233., Houston, Texas, Nov 27-30, 2005. ↻ 🔍

Maintenance Prédicte : vers l'Industrie 4.0^{9 10}



- Prévention/Prédiction des anomalies.
- Estimation de durée de vie utile restante

9. V. Atamuradov, K. Medjaher, P. Dersin and B. Lamoureux et N. Zerhouni. "Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation", International Journal of Prognostics and Health Management, vol. 8, pp. 31-63, 2017.

10. J. J. M. Jimenez, S. Schwartz, R. Vingerhoedsa, B. Grabot, M. Salaüna. "Towards multi-model approaches to predictive maintenance : A systematic literature survey on diagnostics and prognostics", Journal of Manufacturing Systems, vol. 56, pp. 539-555, 2020

Merci de votre attention !

- ⑥ Estimation Parcimonieuse par ADMM
- ⑦ Réglage Automatique des Hyperparamètres
- ⑧ C-ADDICT VS ADDICT & cie

Mise à jour ADMM

- Mise à jour de \mathbf{x}_C

$$\hat{\mathbf{x}}_C^{k+1} = (\Phi_M \Phi_M^T + \mu_C^k I)^{-1} (\Phi_M^T \mathbf{r}_C^k + \mathbf{m}_C^k + \mu_C^k \mathbf{z}^k) \quad (23)$$

où $\mathbf{r}_C^k = \mathbf{y}_C - \mathbf{e}_C^k$.

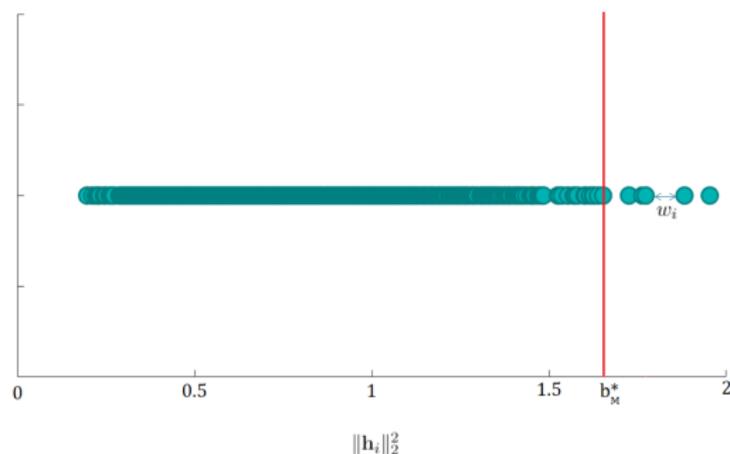
- Mise à jour de \mathbf{z}

$$\hat{\mathbf{z}}^{k+1} = S_{\gamma^k} \left[\mathbf{x}_C^{k+1} - \frac{1}{\mu_C^k} \mathbf{m}_C^k \right]$$

avec $\gamma^k = \frac{\alpha_C}{\mu_C^k}$, et S_{γ^k} est l'opérateur de seuillage doux.

- Mise à jour de \mathbf{e}_C $\hat{\mathbf{e}}_C = T_{b_C} [\mathbf{y}_C - \Phi_M \mathbf{x}_C]$ où T_{b_C} est l'opérateur de seuillage par groupe.
- Mise à jour de \mathbf{m} $\mathbf{m}^{k+1} = \mathbf{m}^k + \mu^k (\mathbf{z}^{k+1} - \mathbf{x}^{k+1})$

One-Class Support Data Thresholding (OC-SDT)



Problème Primal

$$\min_{b_M} b_M + C \sum_i w_i \xi_i \quad (24)$$

$$\text{s.c. } \|\mathbf{h}_i\|_2^2 < b_M + w_i \xi_i \quad (25)$$

$$\xi_i \geq 0 \quad (26)$$

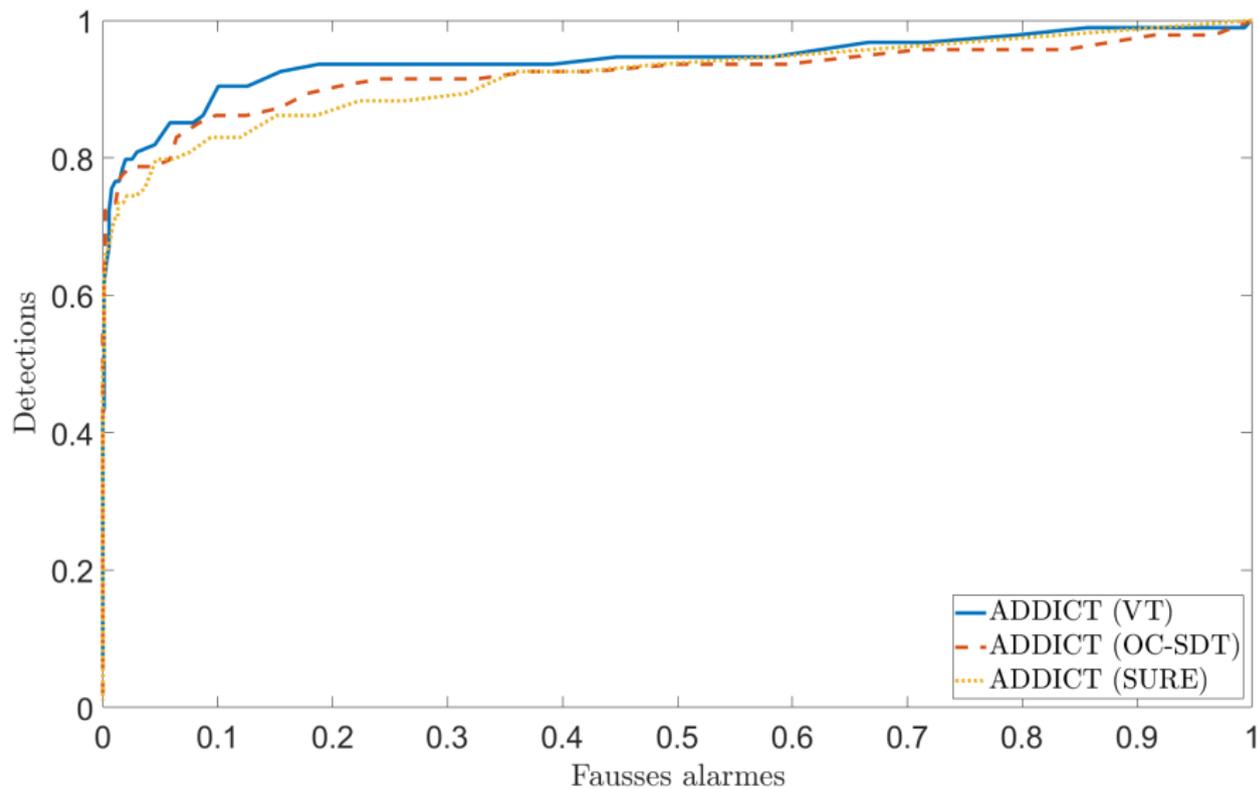
Problème Dual

$$\max_{\alpha_i} \sum_i \alpha_i \|\mathbf{h}_i\|_2^2 \quad (27)$$

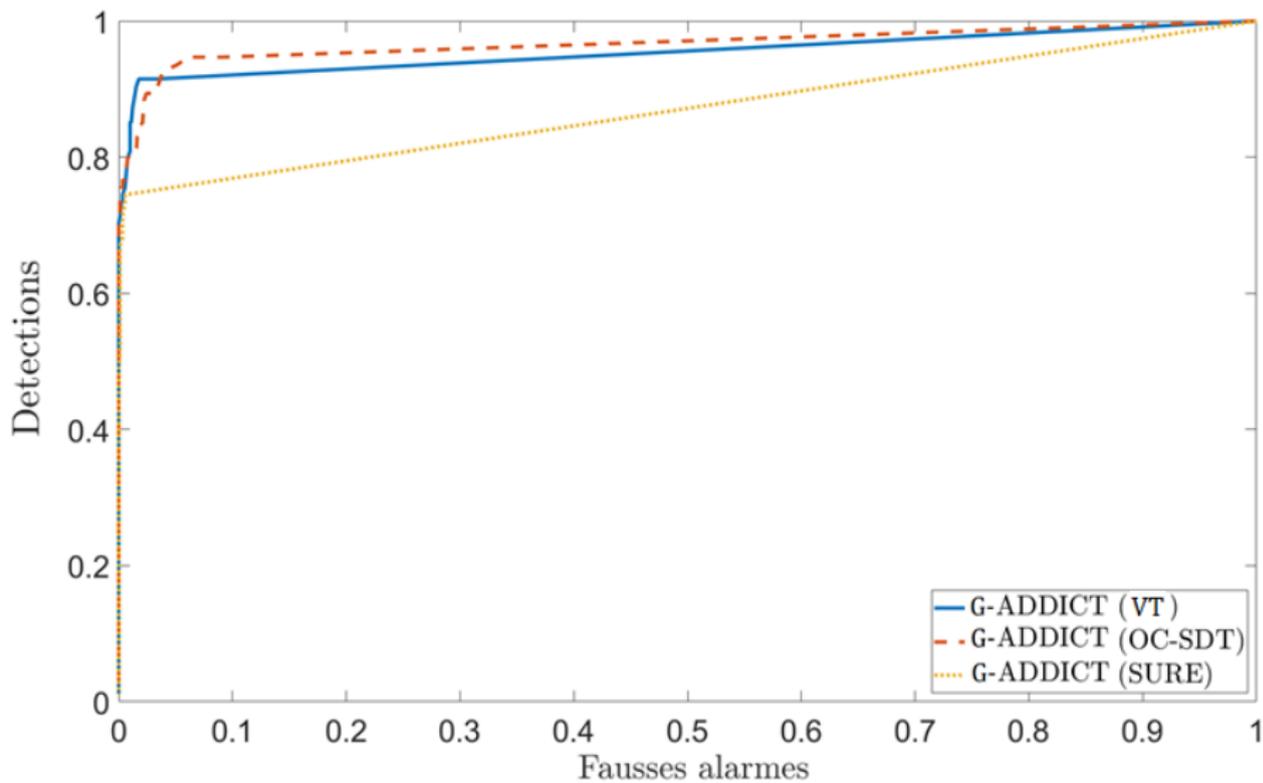
$$\text{s.c. } \sum_i \alpha_i = 1 \quad (28)$$

$$\alpha_i \in [0, Cw_i]. \quad (29)$$

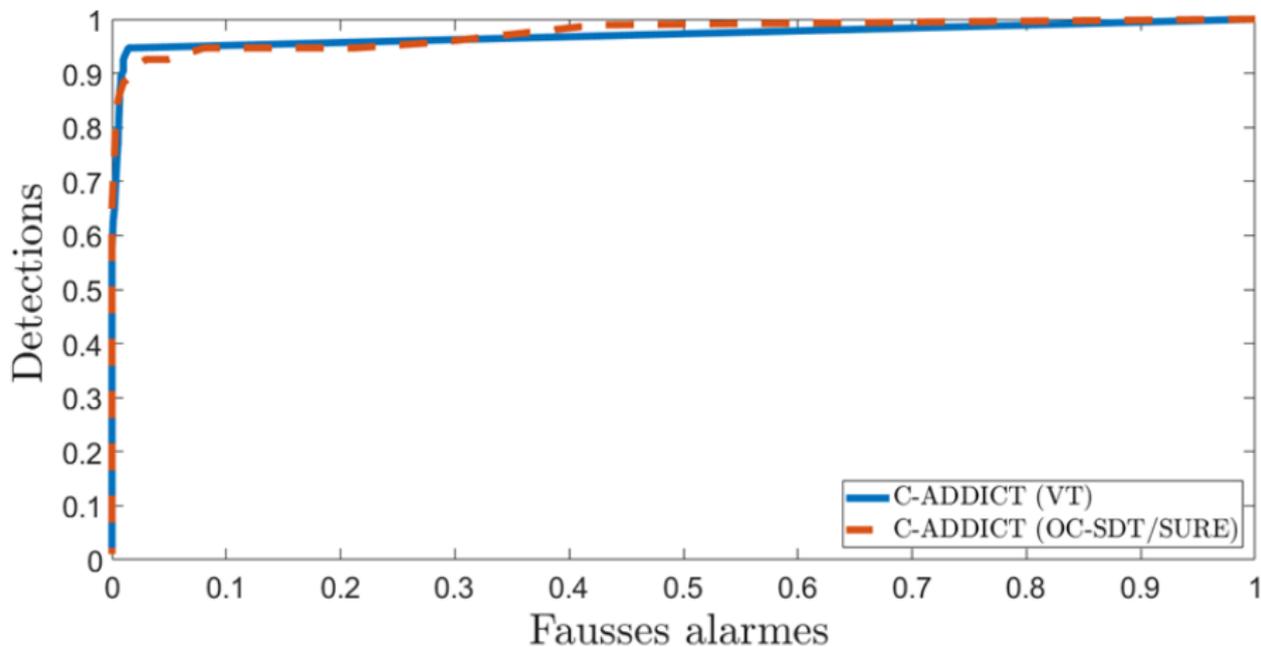
Résultats Expérimentaux



Résultats Expérimentaux



Résultats Expérimentaux



Résultats Expérimentaux

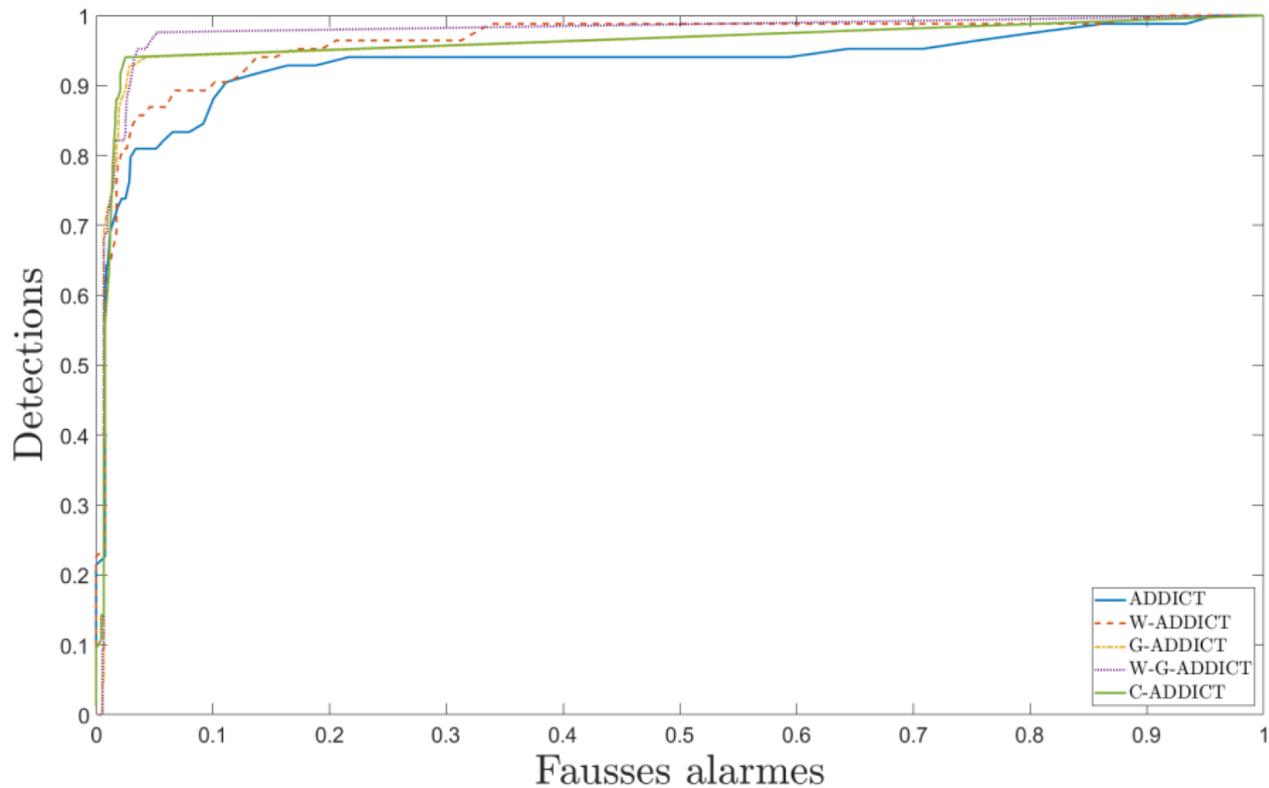
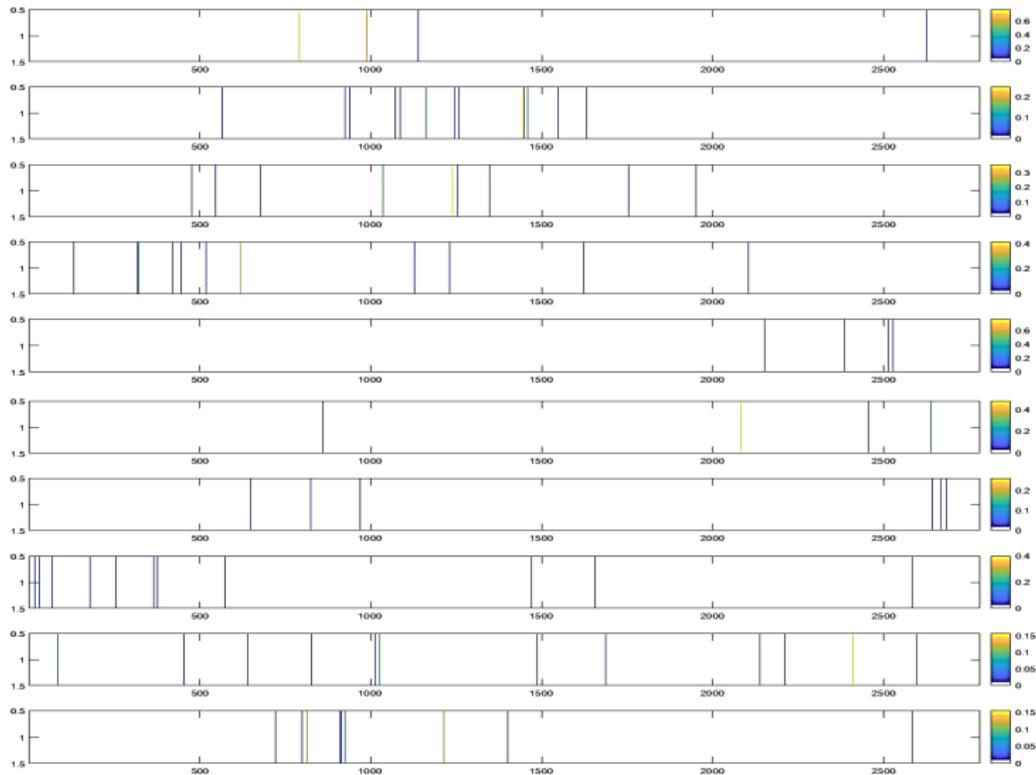


Table – Values of P_D and P_{FA} for ADDICT and GADDICT.

Method	Threshold	P_D	P_{FA}
ADDICT	4.2	80.85%	3%
C-ADDICT	ϵ	94.7%	1.7%
W-ADDICT	4.5	85%	2.7%
G-ADDICT	ϵ	93.6%	3.3%
W-G-ADDICT	ϵ	97.6%	6%

Exemple de carte de coefficients



Comparaison des algorithmes ADDICT et C-ADDICT

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