

DÉTECTION NON SUPERVISÉE D'ANOMALIES DANS LES IMAGES SATELLITES POUR LE MONITORING DES SURFACES OCÉANIQUES À L'AIDE DE L'ACP ROBUSTE ET DU TEST DE GOODNESS OF FIT BASÉ SUR LA DISTANCE DE WASSERSTEIN ENTRE PROCESSUS PONCTUELS

Julien Bastian¹ & Stéphane Chrétien^{1,2} & Ben Gao^{1,3} & Rémi Vaucher^{1,3}

¹ *Laboratoire ERIC, France. E-mail : julien.bastian, stephane.chretien, ben.gao, remi.vaucher@univ-lyon2.fr*

² *The Alan Turing Institute, London*

³ *HALIAS Technologies, France. E-mail : ben.gao, remi.vaucher@halias.fr*

Résumé. L'utilisation de l'imagerie satellitaire pour la détection d'anomalies à la surface de la mer offre une solution prometteuse pour une surveillance systématique. Nous proposons une approche efficace et non supervisée pour la détection de certaines anomalies de la surface de la mer à l'aide d'une ACP robuste, combinée à un test d'adéquation utilisant la distance de Wasserstein pour caractériser statistiquement la différence entre la forme des anomalies et la loi uniforme.

Mots-clés. Robust PCA, Détection d'anomalie, Monitoring de la surface marine.

Abstract. The use of satellite imagery for sea surface anomaly detection offers a promising solution for systematic monitoring. We propose an efficient and unsupervised approach for the detection of certain sea surface anomalies using Robust PCA, combined with a goodness-of-fit test using Wasserstein distance to statistically characterize the difference between the shape of the anomalies and the uniform law.

Keywords. Robust PCA, Anomaly detection, Marine surface monitoring.

1 Présentation du problème de détection

La détection de fuites d'hydrocarbures dans les environnements marins est un défi majeur pour l'industrie pétrolière et les organismes de surveillance environnementale. L'utilisation d'images satellite pour la détection offre une alternative prometteuse en fournissant une couverture étendue et une actualisation régulière des données observées.

1.1 Décomposition des images par Robust PCA

L'analyse de ces images par des méthodes avancées de statistique en haute dimension, utilisant la sparsité et la sparsité spectrale, telle la Robust PCA, permet d'identifier les

signaux faibles associés aux fuites d'hydrocarbures, qui seraient autrement noyés dans le bruit de fond ou masqués par d'autres caractéristiques de la surface marine.

En effet la surface marine dispose d'une régularité due au mouvement naturel de l'eau, qui peut être représenté par une matrice de rang faible.

1.1.1 Rappels sur l'ACP robuste

L'ACP robuste est une technique puissante pour séparer une matrice d'observations en composantes de structure différente :

$$X = L + S,$$

où L est une matrice de rank faible et S une matrice parcimonieuse.

Appliquée à l'analyse d'images satellite pour la détection de fuites d'hydrocarbures, cette méthode consiste à décomposer la matrice image en la somme de deux matrices distinctes : une matrice de rang faible et une matrice sparse (ou parcimonieuse). La matrice de rang faible capture la partie stationnaire ou de fond de l'image, incluant les caractéristiques constantes de la surface de l'eau et les objets stables, tandis que la matrice sparse isole les anomalies ou les éléments transitoires, comme les fuites d'hydrocarbures.

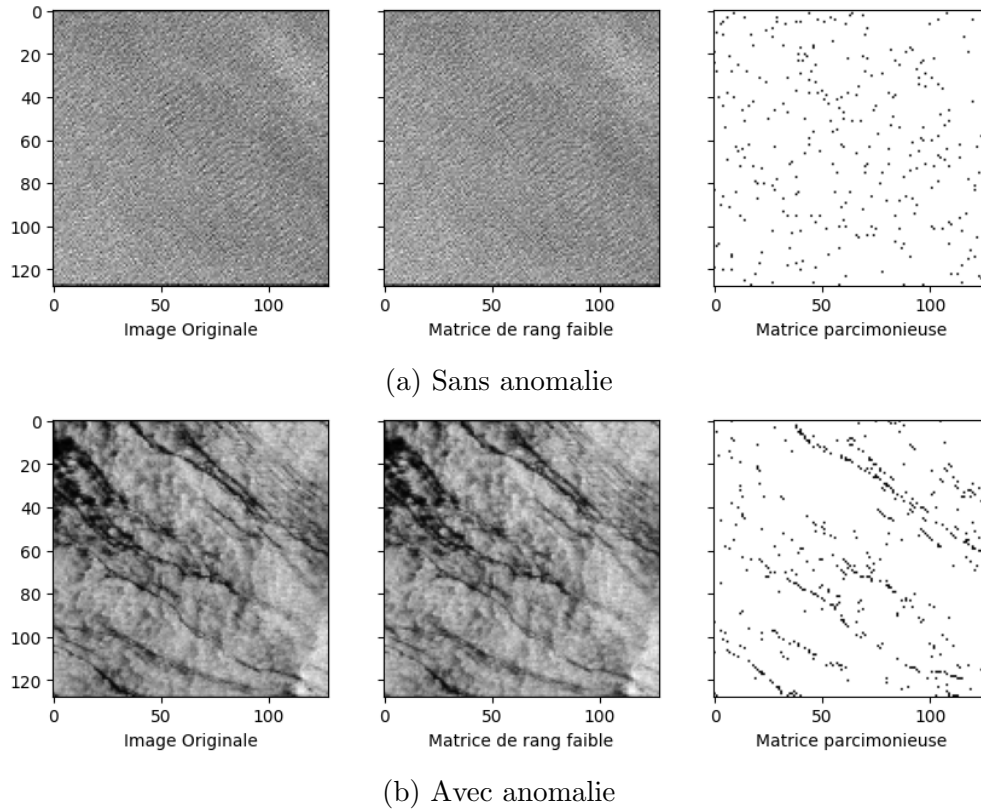


FIGURE 1 – Comparaison des décomposition Rang faible/Sparse pour des images sans et avec anomalies

1.1.2 Méthode numérique

Plusieurs approches numériques ont été développées pour l'identification des deux matrices L et S dans la décomposition de l'image.

Moindres carrés pénalisés par la norme nucléaire et la norme ℓ_1 . L'objectif est de trouver une matrice qui minimise la somme d'une fonction de perte (habituellement les moindres carrés) et d'une pénalité combinaison linéaire de la norme nucléaire de la matrice L et de la norme ℓ_1 de la matrice S . La norme nucléaire de la matrice L , notée $\|L\|_*$ est la somme des valeurs singulières de la matrice. Le problème d'estimation de L et S revient alors à résoudre le problème d'optimisation

$$\min_{L, S \in \mathbb{R}^{d_1 \times d_2}} \frac{1}{2} \|X - L - S\|_F^2 + \lambda \|L\|_* + \mu \|S\|_1, \quad (1)$$

pour lequel une sélection précise des paramètres de régularisation est requise et peut être fondée sur la Validation Croisée, par exemple. Cette approche étend celle proposée dans [Candes et al., 2009] au cas bruité.

Méthode de minimisation alternée. Une méthode de minimisation alternée rapide peut être mise en œuvre. Il est possible par exemple d'implémenter la méthode proposée dans [Cai et al., 2019]. Nous nous sommes restreint à la méthode Principal Component Pursuit de [Candes et al., 2009], dans les expériences proposées.

2 Détection par test de Wasserstein Goodness of Fit

Dans le contexte de la détection de fuites, les composantes non-nulles de la matrice Sparse correspondent aux pixels d'intérêt où les fuites sont susceptibles d'être présentes. Le problème principal est que les pixels sélectionnés ne permettent pas de circonscrire une zone bien définie mais se répartissent plutôt comme la réalisation d'un processus ponctuel dont le support spatial coïncide avec les zones contaminées par les hydrocarbures.

L'approche proposée pour la détection consiste à vérifier à l'aide d'un test statistique que le support du processus ponctuel révélé par la matrice S est différent de celui d'un processus de Poisson associé à une intensité uniforme sur l'image.

2.1 Mise en place du test

Pour valider l'efficacité de cette méthode dans l'identification des fuites d'hydrocarbures, un test de goodness of fit est souvent réalisé en utilisant la distance de Wasserstein. Cette distance mesure l'écart entre la distribution spatiale des composantes identifiées comme fuites d'hydrocarbures et une distribution de référence pour laquelle aucune hypothèse sur la forme

du support n'est faite. Un écart significatif indique que la matrice sparse S détecte efficacement les anomalies correspondant aux fuites d'hydrocarbures, validant ainsi la pertinence de la décomposition obtenue par Robust PCA.

3 Expériences

3.1 Données

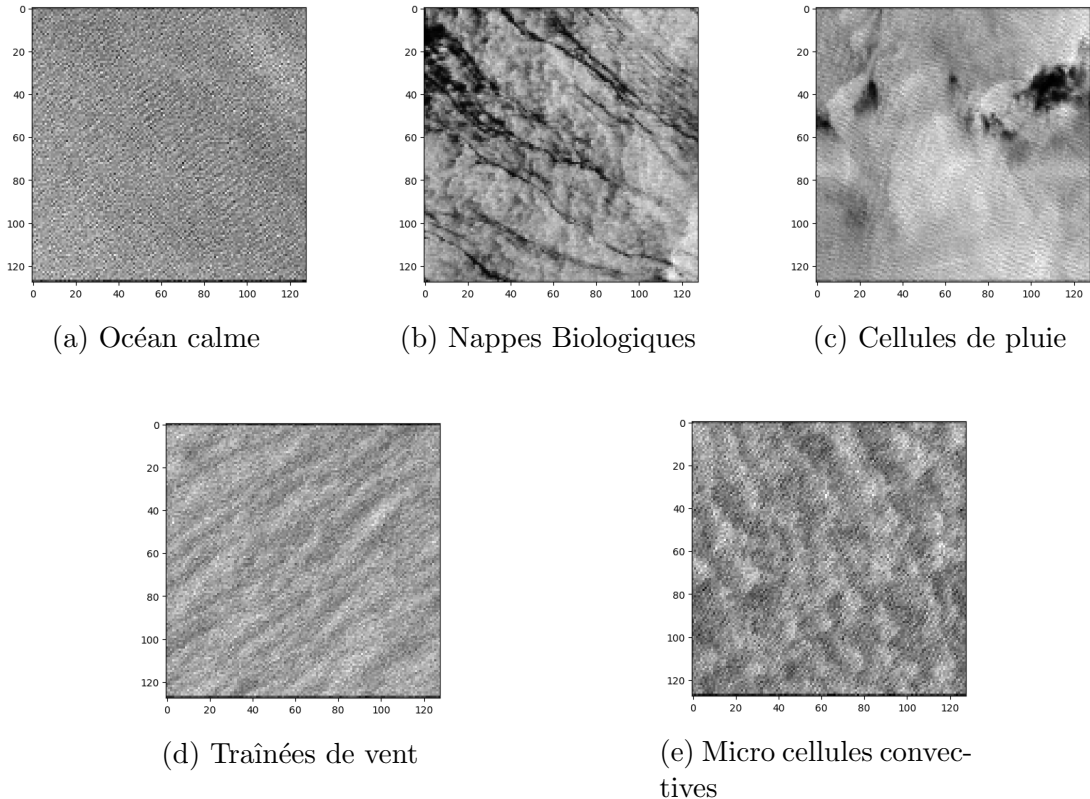


FIGURE 2 – Illustrations des différentes classes considérées.

Les données que nous avons utilisées pour les expériences numériques permettant d'évaluer les performances de l'approche sont celles proposées dans [Wang et al., 2019]. Ces données présentent une collection d'images satellites de type SAR, contenant divers types de configurations de surfaces marines labélisées suivant le type de phénomène présent. Les classes représentées sont "Biological Slicks", "MicroConvective Cells", "Pure Ocean Waves", "Rain Cells" et "Wind Streaks", illustrées en Figure 2.

3.2 Résultats

Nous avons choisi l'hyperparamètre $\lambda = \frac{2}{\sqrt{n_{max}}}$ pour la Robust PCA parmi les options suivantes : $\frac{1}{\sqrt{n_{max}}}$, $\frac{2}{\sqrt{n_{max}}}$ et $\frac{3}{\sqrt{n_{max}}}$. Seules les entrées négatives de la matrice parcimonieuse

Prédiction	Océan Pure	Nappes biologiques	Cellules de pluie	Trainées de vent	Micro cellules convectives
Rejet	0.060	0.895	0.730	0.115	0.125
Non rejet	0.940	0.105	0.270	0.885	0.875

TABLE 1 – Taux d’image pour lesquelles le test de Wasserstein Goodness of Fit rejette l’uniformité avec le seuil de significativité $\alpha = 0.05$.

obtenue sont conservées, les anomalies considérées ici correspondant à des zones de luminosité basses. Parce que celles-ci représentent les nappes que nous souhaitons détecter. De plus en supposant que les entrées non nulles proviennent d’une réalisation d’un processus ponctuel, nous préservons la structure du support de la matrice S plutôt que l’amplitude de ses entrées. En ce qui concerne le test de Wasserstein Goodness of Fit, nous avons généré, pour chaque image testé, 90 échantillons suivant une loi uniforme bi-dimensionnelle et qui ont la même taille que le nombre de points dans S , afin d’obtenir la distribution empirique de $T_n : W_2^2(\hat{P}_n, P_{uniform})$.

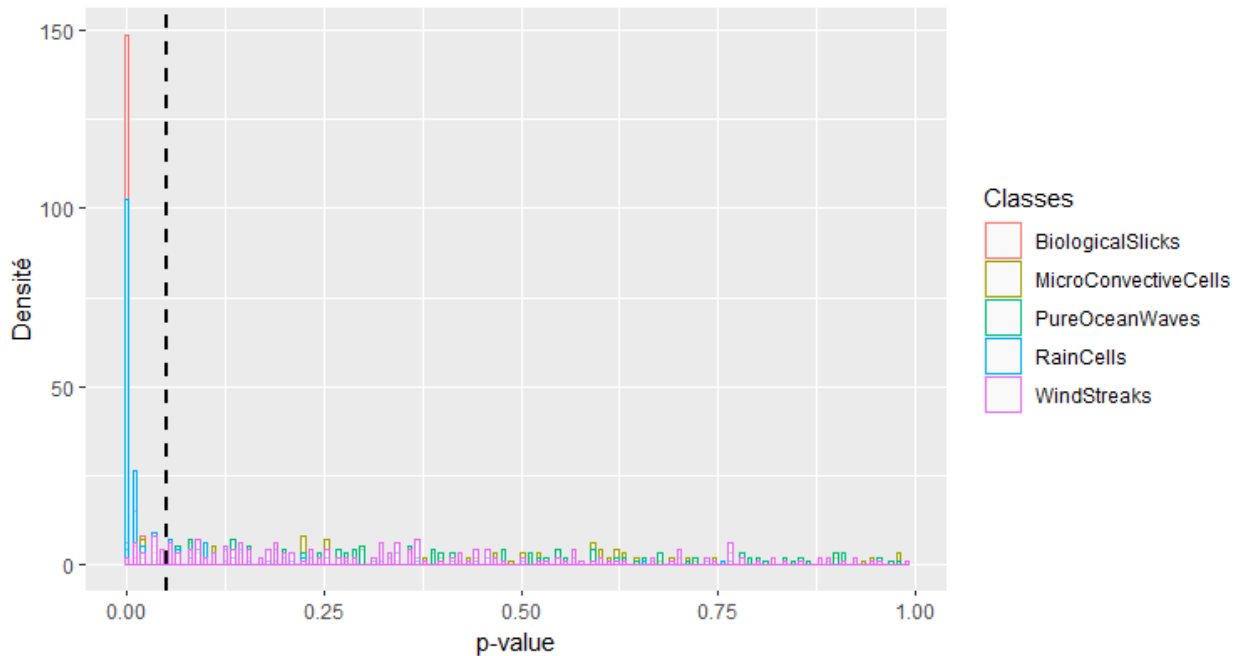


FIGURE 3 – Distribution des p-values du test de Wasserstein Goodness of Fit, la ligne verticale représente le seuil de significativité $\alpha = 0.05$

Les Figures 3 et 4 montrent que les deux classes principalement détectées par notre méthode sont les classes Biological Slicks et Rain Cells. Les matrices Sparses de la décomposition des images de type Rain Cells ont une forme caractéristique de type réalisation d’une gaussienne bi-variée et sont facilement détectables en comparant la distribution à une gaussienne plutôt qu’à un processus ponctuel uniforme. Les classes ”MicroConvective Cells”, ”Pure Ocean Waves” et ”Wind Streaks” ont une distribution beaucoup plus proche de la distribution uniforme. Cela démontre que ces anomalies sont d’un type pulvérisé et nous intéressent moins en terme de monitoring de la surface marine. Le tableau 1 illustre les performances de

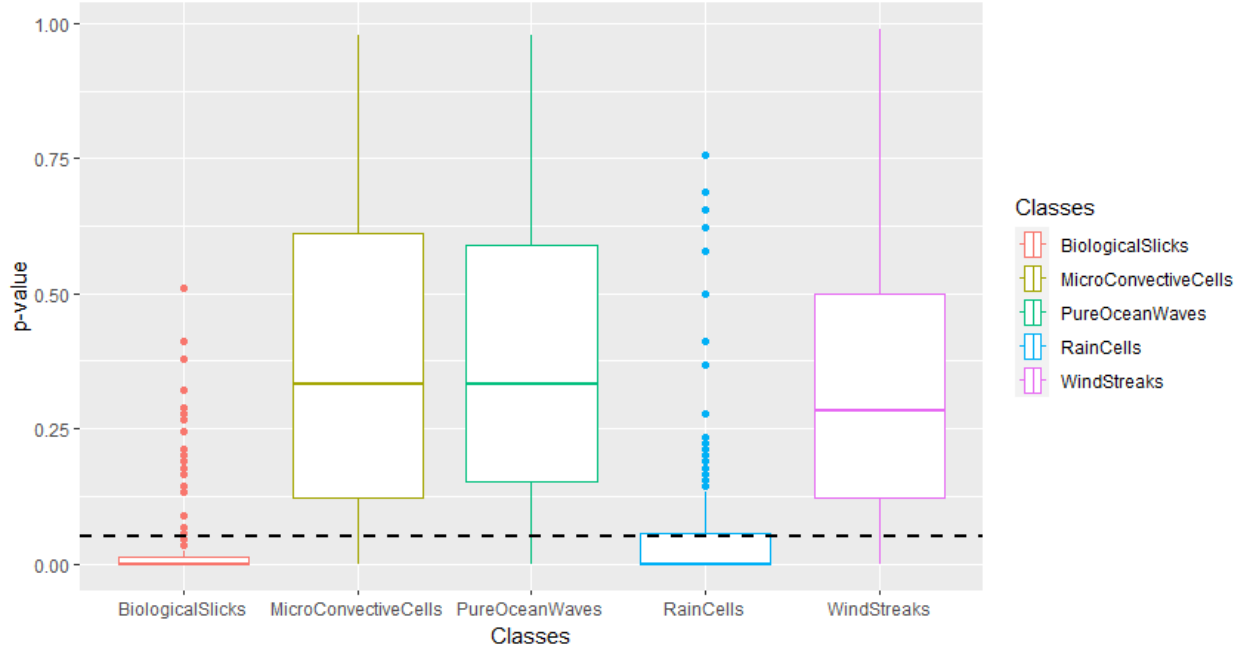


FIGURE 4 – Boxplot des p-values pour le test d’adequation fondé sur les distances de Wasserstein, la ligne horizontale représente le seuil de significativité $\alpha = 0.05$.

la méthode vis à vis de chaque classe considérée.

4 Conclusion

En conclusion, l’application de la Robust PCA à l’analyse d’images satellite pour la détection de fuites d’hydrocarbures représente une avancée significative dans la surveillance environnementale des installations pétrolières.

4.1 Avantages de l’approche proposée

La méthode Robust PCA combinée avec le test sur la distance de Wasserstein pour l’analyse d’images satellite en vue de la détection de fuites offre plusieurs avantages.

- Premièrement, elle permet une discrimination fine entre les caractéristiques de fond et les anomalies, rendant possible la détection de fuites même dans des conditions où les signaux sont faibles ou masqués par d’autres éléments de l’image.
- L’approche proposée est complètement non-supervisée et fournit un cadre automatisé qui peut être régulièrement mis à jour avec de nouvelles données d’observation.
- L’utilisation de la distance de Wasserstein comme critère de validation ajoute une couche supplémentaire de fiabilité à la détection, en offrant une mesure quantitative de la performance de la méthode.

4.2 Perspectives

Une amélioration de la puissance du test pourra être proposée dans une extension prochaine de ces travaux, en incorporant des *a priori* plus spécifiques sur la forme de la structure du support du processus ponctuel à détecter.

Bibliographie

- [Cai et al., 2019] Cai, H., Cai, J.-F., and Wei, K. (2019). Accelerated alternating projections for robust principal component analysis. *The Journal of Machine Learning Research*, 20(1) :685–717.
- [Candes et al., 2009] Candes, E. J., Li, X., Ma, Y., and Wright, J. (2009). Robust principal component analysis?
- [Wang et al., 2019] Wang, C., Mouche, A., Tandeo, P., Stopa, J. E., Longépé, N., Erhard, G., Foster, R. C., Vandemark, D., and Chapron, B. (2019). A labelled ocean sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode. *Geoscience Data Journal*, 6(2) :105–115.