



# Détection automatique d'anomalies dans les données satellitaires

Aymeric Arnould

Thales Alenia Space, Cannes, France

## Résumé

Dans le cadre de la surveillance de l'activité d'un satellite, l'objectif du stage était d'implémenter des méthodes pour faciliter et cibler l'analyse de quantités importantes de données.



## Séparation de comportements pour post-traitement

Afin de réduire les volumes de données, une transformation issue de la méthode SAX[1] a été utilisée pour déterminer une représentation des séries temporelles (voir Figure 1). L'implémentation d'un *clustering* hiérarchique[2] permet ensuite de regrouper les séries par comportements (voir Figure 2).



Fig. 1 : Méthode de transformation d'une série temporelle : quantification et représentation matricielle (exemple d'une mesure de vitesse d'une roue à inertie)

## Clustering interactif

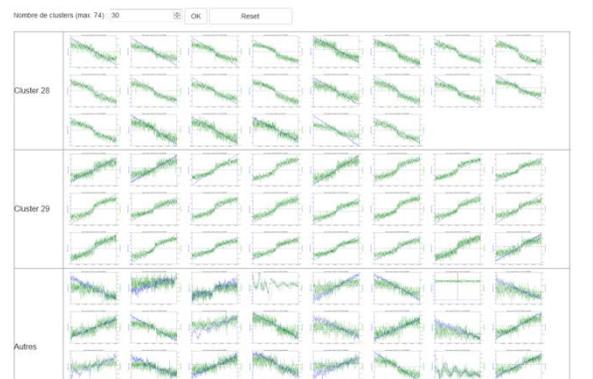


Fig. 2 : Aperçu de l'outil de classification interactive. Les 2 premières lignes du tableau sont des clusters, la dernière correspond aux points non regroupés.

## Détection d'anomalies sur flux de données

L'algorithme de détection d'anomalies mis en place permet d'analyser des flux de données (obtenus ici depuis un système de stockage, l'objectif étant une implémentation temps-réel).

L'architecture utilisée permet de gérer plusieurs critères (par exemple un critère dédié à une anomalie donnée, voir Figure 3) ainsi que différents paramètres.

L'algorithme attire l'attention de l'opérateur sur les données détectées anormales (nouveautés ou peu probables).

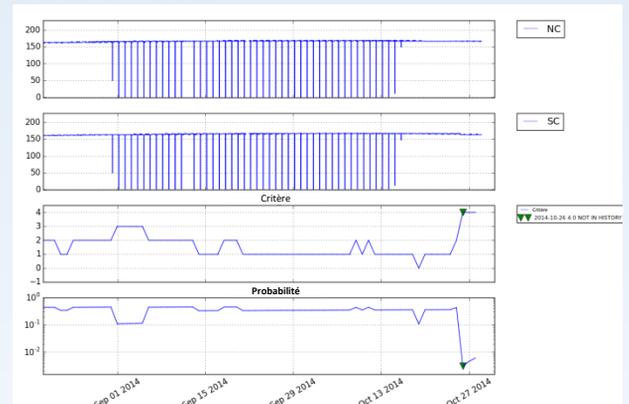


Fig. 3 : Affichage issu de l'algorithme de détection d'anomalies sur flux de données (mesures de courants de générateurs solaires, critère et probabilité)

## Conclusion

Dans ce type d'applications, il est difficile de réaliser un compromis entre l'application de critères spécifiques et une détection plus souple et robuste.

Les algorithmes mis en place permettent un traitement automatique de certaines données et pourront servir pour l'implémentation de programmes évolutifs et plus autonomes.

## Références

- [1] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. *A Symbolic Representation of Time Series, with Implications for Streaming Algorithms*. Data Mining and Knowledge Discovery, Springer Science + Business Media, 2003.
- [2] S. C. Johnson. *Hierarchical Clustering Schemes*. Psychometrika, 32(3), September 1967.