

# SP-519: Traitement des données spatiales

Projet : Prediction of 1 MeV Electron Fluxes in Low Earth Orbit Using NOAA-15 and Solar Wind Data

Müller Raphaël – Tournié Rémi–

Venot Tanguy

IPSA

Aéro 5 – ELSS

Deux objectifs principaux sont fixés :

- Prédire le flux d'électron P6 à 12h, 24h, 36h, 48h.
- Déetecter d'éventuel événement, de flux extrême supérieur au 95 ou 99 percentile du flux à 12h, 24h, 36h, 48h.

## 1 Source de données

Pour notre étude, nous aurons d'abord besoin des principaux paramètres solaires et interplanétaires, à savoir : la composante Bz du champ magnétique interplanétaire (IMF en coordonnées GSM), la vitesse du vent solaire V et le flux radio solaire F10.7. Nous récupérerons toutes ces valeurs grâce à la librairie SunPy. Celle-ci permet d'accéder aux données collectées par diverses agences. Ensuite, nous aurons besoin des données collectées par l'instrument SEM-2 du satellite NOAA-15, plus précisément des données du flux d'électrons (canal P6) et de la distance par rapport à la Terre L. Nous avons téléchargé ses données à partir du site internet.

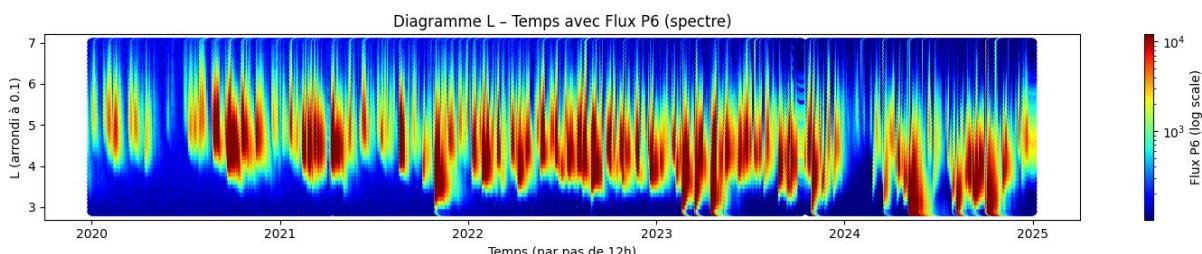


Figure 1 – diagramme L / Temps (2020-2025)

## 2 Préparation des Données et Analyse Statistique

### 2.1 Nettoyage des données

Pour nettoyer les données, nous avons utilisé plusieurs stratégies afin de repérer les outliers ou les valeurs manquantes (NaN), afin de les retirer et d'obtenir des données exploitables. En premier lieu, nous avons filtré toutes les valeurs de flux correspondant à des valeurs de L supérieures à 7 ou inférieures à 2,8. Ensuite, nous avons moyenné ces valeurs sur 12 h pour chaque intervalle de L, avec un pas de 0,1. Finalement, nous appliquons un score Z afin d'éliminer les points présentant des valeurs aberrantes.

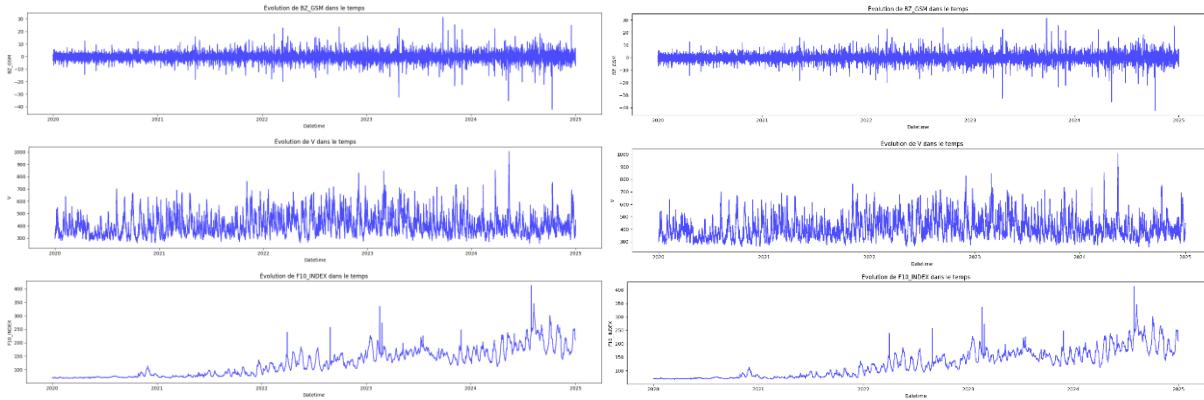


Figure 2 – Données avant après filtrage

## 2.2 Visualisation des Statistiques

Pour permettre la détection de flux intense lors de nos futur prédition nous avons définis le 95ieme percentile. Les valeurs prédites au-dessus de cette limite seront considérées comme des flux intenses qui nécessitent de notifier le responsable.

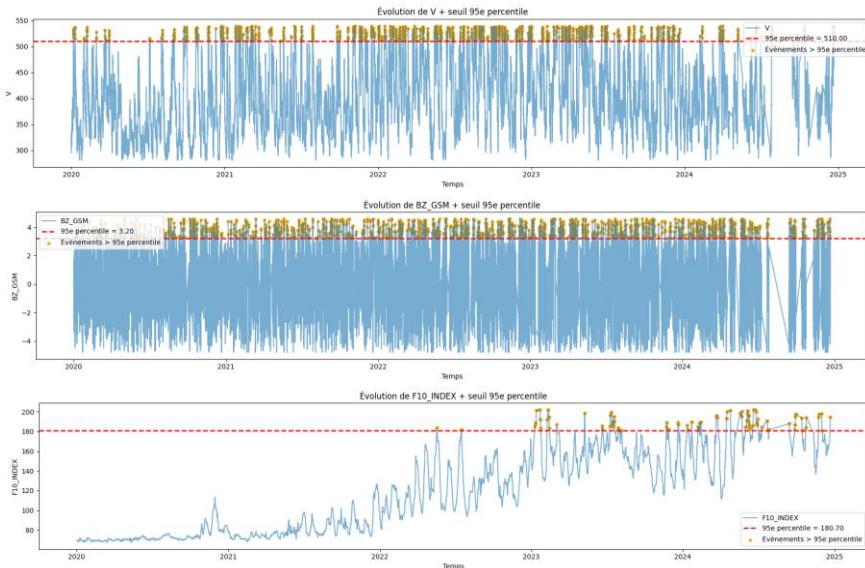


Figure 4 – Distribution du flux et seuil (95 percentiles)

## 3 Prédiction et détection

Nous avons utilisé et comparé plusieurs outils de prédition afin de sélectionner les plus pertinent.

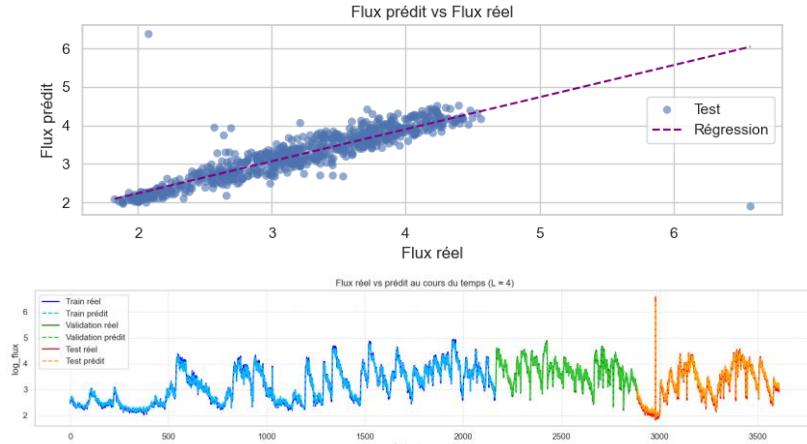


Figure 5 – Linear Regression

## 4 Conclusion

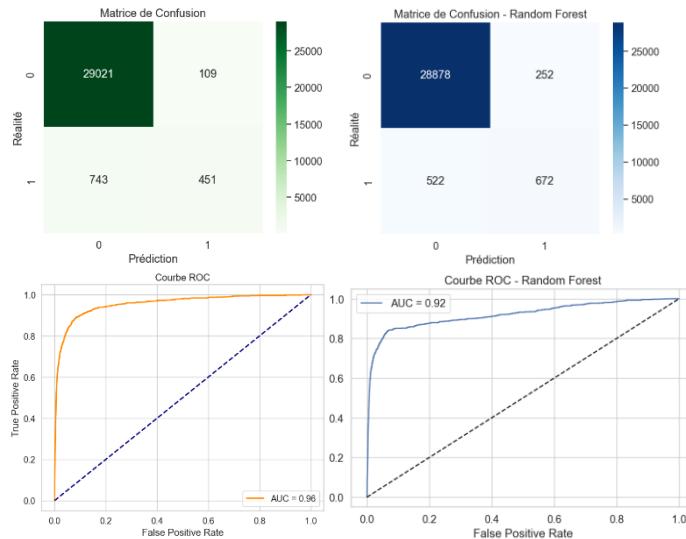


Figure 6 – Confusion matrix and ROC curve (regression , forest)

Le modèle proposé est léger, rapide et transparent : il ne repose que sur deux paramètres libres (le pas de discréttisation en  $L$  et la fenêtre de moyennage de 12 h), ce qui minimise le risque de overfitting et le rend accessible.

Cependant en ne s'appuyant que sur  $L$ , le modèle néglige d'autre paramètres qui peuvent être important. Le lissage 12 h / 0,1 L écrase les événements rapides, tandis que les incertitudes instrumentales contaminent les moyennes et biaissent le rejet des outliers.

Pour avancer : prendre en compte le MLT, Kp, jour de l'année et phase du cycle solaire, puis remplacer le lissage par un noyau plus fin.