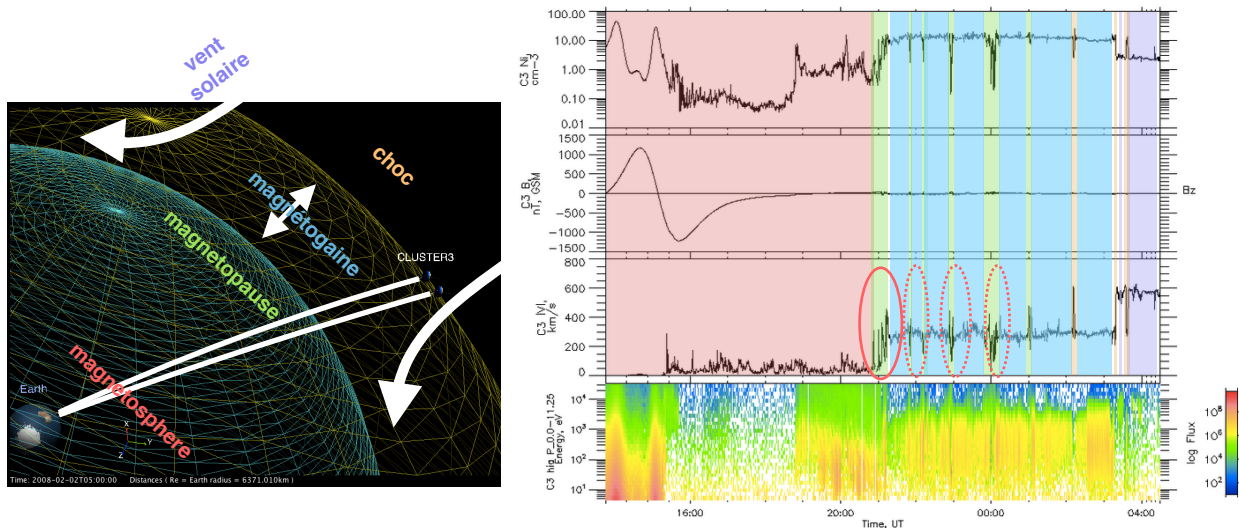


## Machine learning pour la détection automatique de phénomènes plasmas dans les données satellites in situ

Détecter les phénomènes plasmas spécifiques dans les données mesurées par les satellites dans le milieu interplanétaire nécessite aujourd'hui de reconnaître les signatures typiques qu'ils ont sur les propriétés du plasma (densité, vitesse etc.) et du champ électromagnétique.



Par exemple, dans la figure ci-dessus, on représente à droite les données mesurées par un satellite qui sort de la magnétosphère (région rouge), entre dans la magnétogaine (région bleue) puis enfin dans le vent solaire (région violette), en traversant successivement la magnétopause (frontière verte) et le choc (frontière orange). S'il est relativement aisé de repérer visuellement certaines de ces signatures dans les données, l'automatisation de la détection reste quasi impossible en raison de leur très grande variabilité.

En effet, les traversées de frontières ne se font jamais « franchement » mais quasi toujours de manière oscillante, à cause d'ondes de surfaces et autres phénomènes. Sur la figure ci-dessus par exemple, la magnétopause (frontière verte) est traversée plusieurs fois à cause de son mouvement propre oscillant. Deux traversées sont évidemment toujours différentes. On comprend donc pourquoi automatiser les détections sur la base de variations supposées « a priori » est très complexe. Le problème est encore plus compliqué si l'on veut identifier, au sein des frontières, des signatures associées à des phénomènes physiques bien spécifiques, comme la reconnexion magnétique. Celles-ci sont par exemple identifiables dans les intervalles de temps entourés en rouges.

Si tout ceci peut se faire « à l'oeil », c'est évidemment très lent, ce qui empêche tout espoir de collecter un ensemble statistiquement pertinent d'événements. Deuxièmement, elle se fait souvent sur des critères peu reproductibles d'une personne à l'autre, et repose entièrement sur la présence de signatures « typiques », ce qui empêche naturellement la détection d'événements dont les signatures seraient différentes, i.e. la découverte de nouvelle physique.

**Le but de ce stage** est de tenter une nouvelle méthode d'automatisation de ces détections en utilisant des algorithmes d'apprentissage automatique (machine learning). L'étudiant(e) devra se familiariser avec les données mesurées par les satellites, déterminer et extraire les propriétés essentielles permettant aux algorithmes d'apprentissage de bibliothèques existantes (e.g. [scikit-learn](#)) d'obtenir les meilleures performances de détection, à partir de listes pré-établies. Le travail se fera dans une petite équipe et dans le cadre du projet [SciQLOP](#) et en collaboration étroite avec des chercheurs et ingénieurs du laboratoire de physique des plasmas, et un mathématicien du laboratoire de mathématiques appliquées de l'X. Ce stage pourra se prolonger et s'étendre en thèse dans le cadre d'études statistiques des structures se propageant dans le vent solaire et leur interaction avec la magnétosphère terrestre.

**Mots clés :**

- analyse de données satellites
- machine learning

**Profil du candidat :**

- Gout pour l'analyse de données
- Connaissances en python
- Gout pour le travail d'équipe
- Autonomie