



## Apprentissage automatique et modèles de grains interstellaires

**Spécialité** Astrophysique

**Niveau d'étude** Bac+5

**Formation** Master 2

**Unité d'accueil** [DAP/LFEMI](#)

**Candidature avant le** 13/04/2022

**Durée** 3 mois

**Poursuite possible en thèse** oui

**Contact** [GALLIANO Frederic](#)  
+33 1 69 08 18 21  
[frederic.galliano@cea.fr](mailto:frederic.galliano@cea.fr)

### Résumé

Ce stage propose d'étudier l'efficacité des méthodes d'apprentissage automatique pour calculer des modèles complexes d'émission du milieu interstellaire. Nous concentrerons sur l'application aux modèles de poussières.

### Sujet détaillé

La poussière interstellaire est un ingrédient physique clef des galaxies, responsable de l'obscurcissement de la formation stellaire, de la régulation du chauffage et du refroidissement du gaz, et de la croissance de la complexité chimique. La modélisation de l'émission thermique des grains est un outil incontournable pour étudier aussi bien la physique du milieu interstellaire local que la formation d'étoiles enfouie des galaxies distantes.

Notre groupe au CEA a une expertise reconnue sur l'analyse des distributions spectrales d'énergie de poussière. Nous avons développé un code bayésien hiérarchique unique à cet effet (Galliano, 2018, MNRAS, 476, 1445). L'utilisation de ce code requiert de calculer un très grand nombre de modèles de grains (typiquement  $1.E11$ ). Pour le moment, cela se fait en interpolant une grille pré-calculée, ce qui limite le nombre de dimensions de l'espace des paramètres que l'on peut explorer. Nous souhaitons appliquer les techniques d'apprentissage automatique (machine learning) pour pallier à ce problème. Ces techniques ont démontré leur efficacité pour les modèles de photoionisation. L'idée est d'entraîner un algorithme pour reproduire les sorties du modèle. C'est le but du stage.

La première étape consistera à se familiariser avec le modèle physique. Celui-ci prend en compte l'émission d'un mélange de grains de tailles et compositions chimiques différentes. Le spectre émis est calculé à l'aide d'un code de chauffage stochastique, qui permet de décrire le comportement hors-équilibre des plus petits grains. Le spectre prédit est ensuite intégré dans les bandes passantes d'un grand nombre d'instruments (Spitzer, Herschel, Planck, etc.).

La deuxième étape consistera ensuite à déterminer quelle méthode d'apprentissage automatique est la plus adaptée et quels sont les paramètres du réseau de neurones dont nous avons besoin. Le but est de contrôler la précision du résultat. Cela se fera en étudiant la dispersion de plusieurs modèles concurrents.

---

S'il reste du temps, cette méthode pourra être étendue à des modèles plus complexes, prenant en compte le transfert radiatif, ou à des modèles d'évolution cosmique des grains. En fonction de l'investissement de l'étudiant, le travail de stage peut donner lieu à la publication d'un logiciel d'intérêt général qui pourra servir, entre autres, à l'interprétation des observations du JWST.

### **Mots clés**

Milieu interstellaire - physique des galaxies

### **Compétences**

Apprentissage automatique ; physique statistique.

### **Logiciels**

Fortran et Python.

---

## Machine learning and interstellar grain models

### Summary

This internship aims at studying the efficiency of machine learning methods to compute complex models of emission of the interstellar medium. We will focus on their application to dust models.

### Full description

Interstellar dust is a key physical ingredient of galaxies, obscuring star formation, regulating the heating and cooling of the gas, and building-up chemical complexity. Modeling thermal grain emission is a unavoidable tool to study both the physics of local interstellar medium and the embedded star formation in distant galaxies.

Our group at CEA has renown expertise on the analysis of dust spectral energy distributions. We have developed a hierarchical Bayesian code to that purpose (Galliano, 2018, MNRAS, 476, 1445). Using this code requires to compute a very large number of grain models (typically  $1.E11$ ). For now, we are interpolating a large pre-computed grid, limiting the number of dimensions of the parameter space that we can explore. We would like to apply machine learning technics to solve this problem. These technics have demonstrated their efficiency in the case of photoionisation models. The idea is to train an algorithm to reproduce the model outputs. This is the goal of this internship.

The first stage will consist in getting used with the physical model. It accounts for the emission of the mixture of grains with different sizes and chemical composition. The emitted spectrum is computed with stochastic heating code, which describe the out-of-equilibrium properties of the smallest grains. The predicted spectrum is then integrated into the filters of a large number of instruments (Spitzer, Herschel, Planck, etc.).

The second stage will consist in determining which machine learning method is the most appropriate and what are the neural network parameters we need. The goal is to control the accuracy of the result. This will be done studying the scatter of several competing models.

If there is more time, this method will be applied to more complex models, accounting for radiative transfer, or to chemical evolution models. Depending on the investment of the student, this internship work can result in the publication of a general interest software which could be useful, among others, to interpreting JWST data.

### Keywords

Interstellar medium - physics of galaxies

### Skills

Machine learning ; statistical physics.

### Softwares

Fortran et Python.