

## Inference cosmologique du champ de densité de galaxies dans le relevé spectroscopique DESI

**Spécialité** Astrophysique

**Niveau d'étude** Bac+5

**Formation** Master 2

**Unité d'accueil** [DPhP](#)

**Candidature avant le** 20/04/2022

**Durée** 4 mois

**Poursuite possible en thèse** oui

**Contact** [de MATTIA Arnaud / François Lanusse](#)  
+33 1 69 08 62 34  
[arnaud.de-mattia@cea.fr](mailto:arnaud.de-mattia@cea.fr) / [francois.lanusse@cea.fr](mailto:francois.lanusse@cea.fr)

### Résumé

L'objectif est d'initier le développement d'une nouvelle méthode d'analyse des données de DESI, reposant sur des simulations numériques du relevé et des nouvelles techniques de machine learning et d'inférence statistique, plus performante que les analyses standards.

### Sujet détaillé

DESI (Dark Energy Spectroscopic Instrument) est un spectrographe multi-objets monté sur le télescope Mayall à Kitt Peak, en Arizona, qui a récemment commencé à prendre des données. Il va permettre la mesure de 35 millions de redshifts de galaxies et de quasars entre 0.05

L'étude des propriétés statistiques du champ de densité de galaxies (déterminé à partir des positions mesurées des galaxies), permet de contraindre les modèles cosmologiques.

Les analyses standard reposent sur l'estimation de la variance du champ de densité en fonction de l'échelle (ce que l'on appelle spectre de puissance ou fonction de corrélation). Cette mesure est ensuite utilisée pour l'inférence des paramètres cosmologiques. Cependant, le champ de densité de galaxies est fortement non-gaussien, ce qui rend cette compression non optimale.

Dans ce stage, nous proposons au contraire d'explorer une approche théoriquement sans perte pour extraire des informations cosmologiques des relevés de galaxies. Cette technique nouvelle et en plein développement vise à reproduire la densité de galaxies observée à l'aide de simulations. En pratique, un champ de densité initial aléatoire de matière noire est généré dans une boîte cubique. Ce champ de densité est évolué dans le temps selon les équations de la gravité. Un modèle est ensuite utilisé pour relier ce champ de matière noire à un champ de galaxies, puis les effets de sélection du relevé sont appliqués. La vraisemblance du champ de densité des galaxies observé étant donné le champ simulé est calculée, et sa valeur est utilisée pour échantillonner les conditions initiales du champ de densité.

Nous nous appuyerons sur le solveur FlowPM, disponible publiquement, reposant sur TensorFlow (accéléré par GPU). Nous proposons d'y ajouter des modèles pour la connexion matière noire - galaxies et la fonction de sélection du relevé. Il en résultera une première version du modèle qui pourra être comparé à des simulations réalistes de DESI et

---

à des données réelles.

### **Mots clés**

cosmologie, relevés de galaxies, statistiques, machine learning, analyse de données

### **Compétences**

inférence en grande dimension, machine learning

### **Logiciels**

Python, TensorFlow

---

# Forward modeling of the galaxy density field in the DESI spectroscopic survey

## Summary

The goal is to initiate the development of a novel analysis pipeline for DESI galaxy redshift data, based on numerical simulations of the survey and state-of-the-art machine learning and statistical inference techniques to overcome limitations of standard analyses.

## Full description

The Dark Energy Spectroscopic Instrument (DESI) is a multi-object spectrograph mounted on the Mayall telescope at Kitt Peak, Arizona, which has recently started taking data. It will enable redshift measurements of 35 millions of galaxies and quasars between 0.05

The study of the statistical properties of the galaxy density field, inferred from observed galaxy positions, enables stringent constraints to be put on cosmological models. Standard analyses consist in measuring the variance of the galaxy density field as a function of scale (known as the power spectrum or correlation function) and use the obtained summary statistics as a data vector for further inference of cosmological parameters. Yet, the galaxy density field is highly non-Gaussian, such that the above compression is not optimal.

The Dark Energy Spectroscopic Instrument (DESI) is a multi-object spectrograph mounted on the Mayall telescope at Kitt Peak, Arizona, which has recently started taking data. It will enable redshift measurements of 35 millions of galaxies and quasars between 0.05

The study of the statistical properties of the galaxy density field, inferred from observed galaxy positions, enables stringent constraints to be put on cosmological models. Standard analyses consist in measuring the variance of the galaxy density field as a function of scale (known as the power spectrum or correlation function) and use the obtained summary statistics as a data vector for further inference of cosmological parameters. Yet, the galaxy density field is highly non-Gaussian, such that the above compression is not optimal.

In this internship we propose instead to explore a theoretically lossless approach to extracting cosmological information from galaxy surveys. This novel and rapidly developing forward modeling technique aims at reproducing the observed galaxy density with simulations. Namely, an initial random dark matter density field is generated in a cubic box. This field is evolved forward in time following the equations of gravity. The galaxy density field is then modelled on top of the simulated dark matter field and survey selection effects are applied. The likelihood of the observed galaxy density field given the simulated one is computed, and its value is used to iterate over, or to sample, initial conditions of the density field.

We will start from the publicly available GPU-accelerated gravity solver FlowPM in the TensorFlow framework. We propose to integrate in the same framework models for the dark matter - galaxy connection and survey selection function. This will result in a first version of the forward modeling pipeline that may be compared to realistic DESI simulations and real data.

## Keywords

cosmology, galaxy surveys, statistics, machine learning, data analysis

## Skills

high-dimensional inference, machine learning

## Softwares

---

Python, TensorFlow