



## These 2020

**Title** Machine Learning for Euclid Mass Mapping and Cosmological Parameter Estimation

**Laboratory:** IRFU/DAP/CosmoStat, CEA Saclay

**Supervisors:** Jean-Luc Starck and Francois Lanusse

**Contact:** [jstarck@cea.fr](mailto:jstarck@cea.fr) ☎ : 01 69 08 57 64 <http://jstarck.cosmostat.org>

**Keywords:** deep learning, graph, Euclid, Cosmology, cosmology, weak lensing, spherical data

### Subject:

The Euclid satellite, to be launched in 2022, will observe the sky in the optical and infrared, and will be able to map large scale structures and weak lensing distortions out to high redshifts. Weak gravitational lensing is thought to be one of the most promising tools of cosmology to constrain models. Weak lensing probes the evolution of dark-matter structures and can help distinguish between dark energy and models of modified gravity. Thanks to the shear measurements, we will be able to reconstruct a dark matter mass map of 15000 square degrees. Mass mapping entails the construction of two-dimensional maps using galaxy shape measurements, which represent the integrated total matter density along the line of sight. Small- field mass maps have been frequently used to study the structure and mass distribution of galaxy clusters, whereas wide-field maps have only more recently become possible given the broad observing strategies of surveys like CFHTLenS, HSC, DES, and KiDS. Mass maps contain significant non-Gaussian cosmological information and can be used to identify massive clusters as well as to cross-correlate the lensing signal with foreground structures.

A standard method to derive mass maps from weak-lensing observations is an inversion technique formulated by Kaiser & Squires [2]. It has many limitations, however, including the need to smooth the data before (and often after) inversion, thereby losing small-scale information. An alternative method called GLIMPSE has been developed in the CosmoStat laboratory based on sparse reconstruction that avoids this problem and improves the recovery of non-Gaussian features [3, 4]. The algorithm has been tested on simulations and was also recently used to study the A520 merging galaxy cluster with Hubble Space Telescope data [5]. More recently, machine learning has emerged as a promising technique for mass map recovery [6].

The goal of this thesis is to i) compare this technique to the state of the art and investigate if it can be used in practice, ii) extend the method for spherical data, and iii) develop a new machine learning approach to estimate the cosmological parameters. At the core of this new statistical framework will be the development of fast and differentiable cosmological simulations capable of emulating the Euclid survey under various cosmologies. This simulation tool will be based on the FastPM N-body simulation code [7] and implemented directly in the TensorFlow machine learning framework, yielding a differentiable physical forward simulation pipeline which can be directly interfaced with deep learning components or with inference techniques relying on having access to the derivatives of the simulation.

As part of the CosmoStat Laboratory, located at CEA Saclay, the successful candidate will be embedded in a leading French research group, heavily involved in the preparation of the Euclid space mission, and with a long tradition of developing cutting-edge statistical tools for the analysis of astronomical and cosmological data.

## **Résumé:**

### **Le contexte**

Le satellite Euclid, qui sera lancé en 2022, observera le ciel dans les domaines optique et infrarouge, et mesurera les distorsions gravitationnelles jusqu'à des redshifts très élevés. L'effet de lentille gravitationnelle faible est considérée comme l'un des outils les plus prometteurs de la cosmologie pour contraindre les modèles. Les lentilles faibles sondent l'évolution des structures de la matière noire et peuvent aider à distinguer l'énergie noire des modèles de gravité modifiée. Grâce aux mesures de cisaillement, nous pourrions reconstruire une carte de masse de matière noire de 15 000 degrés carrés. La cartographie de masse implique la construction de cartes bidimensionnelles utilisant des mesures de forme de galaxie, représentant la densité de matière totale intégrée le long de la ligne de visée. Les cartes de masse sur des petits champs ont souvent été utilisées pour étudier la structure et la distribution en masse des amas de galaxies, alors que les cartes à grand champ ne sont possibles que depuis peu, en raison des stratégies d'observation de relevés de galaxies tels que CFHTLenS, HSC, DES et KiDS. Les cartes de masse contiennent des informations cosmologiques non gaussiennes significatives et peuvent être utilisées pour identifier des amas massifs ainsi que pour effectuer une corrélation croisée entre le signal de lentille et les structures d'avant plan.

Une méthode standard pour obtenir des cartes de masse à partir d'observations à lentille faible est une technique d'inversion élaborée par Kaiser & Squires [2]. Cependant, elle comporte de nombreuses limitations, notamment la nécessité de lisser les données avant (et souvent après) l'inversion, perdant ainsi des informations à petite échelle. Une méthode alternative appelée GLIMPSE a été développée dans le laboratoire CosmoStat, basée sur une reconstruction parcimonieuse qui améliore la restaurations des structures non gaussiennes [3, 4]. L'algorithme a été testé sur des simulations et a récemment été utilisé pour étudier le cluster de galaxies en fusion A520 avec les données du télescope spatial Hubble [5]. Plus récemment, l'apprentissage automatique est apparu comme une technique prometteuse pour la récupération de cartes de masse [6].

Le but de cette thèse est i) de comparer cette technique à l'état de l'art et de déterminer si elle peut être utilisée dans la pratique, ii) d'élargir la méthode à des données sphériques, et iii) de développer une nouvelle approche d'apprentissage automatique pour estimer la valeur cosmologique. paramètres. Le développement de simulations cosmologiques rapides, capables de simuler le relevé Euclid sous diverses cosmologies, sera au cœur de ce nouveau cadre statistique. Cet outil de simulation sera basé sur le code de simulation FastPM N-body [7] et il sera implémenté directement dans le cadre d'apprentissage automatique avec TensorFlow, ce qui permettra de générer un pipeline de simulation physique différentiable pouvant être directement interfacé avec des composants d'apprentissage approfondi ou avec des techniques d'inférence reposant sur l'avantage d'avoir accès aux dérivées de la simulation.

Dans le cadre du laboratoire CosmoStat, situé au CEA Saclay, le candidat retenu sera intégré dans un groupe de recherche français de premier plan, fortement impliqué dans la préparation de la mission spatiale Euclid et ayant une longue tradition de développement d'outils statistiques de pointe pour le analyse des données astronomiques et cosmologiques.

### **L'environnement scientifique:**

La thèse se déroulera au sein du groupe de recherche pluridisciplinaire, CosmoStat, au Département d'Astrophysique du CEA sous la direction de Jean-Luc Starck et de Francois Lanusse. L'équipe est très impliquée dans le projet Euclid et le weak lensing, avec de nombreuses responsabilités (Jean-Luc Starck est leader de l'unité OULE3, en charge de définir les algorithmes utilisés pour dériver les produits finaux et également du Work Package OULE3-Weak Lensing. L'équipe a une très forte expertise dans différents domaines (problèmes inverses, weak lensing, machine learning etc.). L'étudiant thésard disposera donc d'un environnement idéal pour mener à bien ses travaux. A l'issue de sa thèse, l'étudiant maîtrisera des outils sophistiqués, le savoir faire pour estimer les paramètres cosmologiques à partir de données weak lensing, et aura une très bonne connaissance du projet Euclid. Il pourra alors soit continuer dans le domaine académique, soit exploiter ses compétences dans le domaine industrie.

## **References**



1. Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291.
2. Kaiser, N. & Squires, G. 1993, ApJ, 404, 441.
3. Leonard, A., Lanusse, F., & Starck, J.-L. 2014, MNRAS, 440, 1281.
4. Lanusse, F., Starck, J.-L., Leonard, A., & Pires, S. 2016, A&A, 591, A2.
5. Peel, A., Lanusse, F., & Starck, J.-L. 2017, ApJ, 847, 23.
6. Niall Jeffrey et al, submitted. <https://arxiv.org/abs/1908.00543>
7. Y. Feng, M. Yat Chu, U. Seljak, and P. McDonald. MNRAS, 463(3):2273–2286, 2016.