Galaxy shape inference with machine learning methods

October 8, 2025

Keywords – Astrophysics - Weak lensing - Statistics - Machine learning

---

The CosmoStat team at CEA Paris-Saclay invites applications for a master student (M2) to work with Dr. Emma Ayçoberry and Dr. Samuel Farrens on a novel method to measure galaxy shape using machine learning methods.

CEA Paris-Saclay is located 20km south of Paris, France, in the vicinity of various universities and other research centres. The CosmoStat team is a diverse and multi-disciplinary team of researchers working on various topics in cosmology. Our team is committed to diversity and equality, and encourage applications from women and underrepresented minorities. We support a flexible and family-friendly work environment.

Context

Weak lensing is the distortion of the light from distant galaxies caused by massive objects located between these galaxies and the observer. It is therefore one of the most powerful probes of the large-scale structure of the Universe, in particular the distribution of dark matter. A key quantity in weak lensing is the shear, which corresponds to the distortion of the observed shape of galaxies. Measuring galaxy shapes is crucial because the shear signal is much smaller than the intrinsic shape of galaxies. Accurate shape measurements, therefore determine how well weak lensing can constrain cosmology. Standard methods rely on simplified models of galaxy shapes and can introduce biases in the shear estimation. While these methods are well adapted to the weak lensing regime, they may fail to capture the full complexity of the data.

An alternative approach to extract all the information from the data is to use a forward model, including a shearing operation, where the inference is performed using simulated high-resolution galaxies. The model predictions are then compared with actual observations, from which the shear can be inferred. The pipeline is composed of different components (parameter sampling, galaxy modelling, shear distortion, instrumental response (PSF), data comparison), and each step can be modified to include different effects and levels of complexity. In particular, modern machine learning techniques, such as deep generative models, can be incorporated to capture the full morphological diversity of galaxies and mitigate biases. The objective of this internship will be to extend the galaxy shape inference pipeline being developed by the CosmoStat team and to benchmark it against existing methods.

Outline of project objectives

The internship will essentially be comprised of the following tasks and objectives:

1. Become familiar with the sampling methods (HMC, MCMC, etc.) and the forward model pipeline.
2. Study how different inference tools impact the efficiency and robustness of the shear estimation.
3. Investigate the impact of different instrumental responses (PSF), or different telescopes.
4. Explore methods to disentangle shear from the intrinsic shapes of galaxies.
5. Explore the use of deep generative models for galaxy simulations.

Candidate

The candidate should be a Master 2 (or equivalent) student with background in either physics/astrophysics or applied maths/signal processing/data science. Knowledge of machine learning methods and usage of Git would be a plus. Experience with Python is not required, but would be advantageous.

Internship

The internship will take place in the CEA Paris-Saclay astrophysics department (AIM) in the CosmoStat team, under the supervision of Emma Ayçoberry and Samuel Farrens.

- *Deadline for applications:* November 28th, 2025.
  Please note that applications will be reviewed on a rolling basis, and the position may be filled before the deadline if a suitable candidate is found.
- *Contact:* Samuel Farrens (samuel.farrens@cea.fr), Emma Ayçoberry (emma.aycoberry@cea.fr).
- *Duration:* 4-6 months.
