



COSMOSTAT

PhD Project, 2025-2028

Generative AI for Robust Uncertainty Quantification in Astrophysical Inverse Problems

Application Deadline: January 15th 2025

Keywords: Inverse Problems - Deep Generative Models - Astrophysics

Context: Inverse problems, i.e. estimating underlying signals from corrupted observations, are ubiquitous in astrophysics, and our ability to solve them accurately is critical to the scientific interpretation of the data. Examples of such problems include inferring the distribution of dark matter in the Universe from gravitational lensing effects [1], or component separation in radio interferometric imaging [2].

Thanks to recent deep learning advances, and in particular **deep generative modeling techniques** (e.g. diffusion models), it now becomes not only possible to get an estimate of the solution of these inverse problems, but **to perform Uncertainty Quantification** by estimating the full Bayesian posterior of the problem, i.e. **having access to all possible solutions** that would be allowed by the data, but also plausible under prior knowledge.

Our team has in particular been pioneering such Bayesian methods to combine our knowledge of the physics of the problem, in the form of an explicit likelihood term, with data-driven priors implemented as generative models. This physics-constrained approach ensures that solutions remain compatible with the data and prevents “hallucinations” that typically plague most generative AI applications.

However, despite remarkable progress over the last years, **several challenges still remain** in the aforementioned framework, and most notably:

- **[Imperfect or distributionally shifted prior data]** Building data-driven priors typically requires having access to examples of non corrupted data, which in many cases do not exist (e.g. all astronomical images are observed with noise and some amount of blurring), or might exist but may have distribution shifts compared to the problems we would like to apply this prior to. This mismatch can bias estimations and lead to incorrect scientific conclusions. Therefore, the adaptation, or calibration, of data-driven priors from incomplete and noisy



CosmoStat

observations becomes crucial for working with real data in astrophysical applications.

- **[Efficient sampling of high dimensional posteriors]** Even if the likelihood and the data-driven prior are available, correctly sampling from non-convex multimodal probability distributions in such high-dimensions in an efficient way remains a challenging problem. The most effective methods to date rely on diffusion models, but rely on approximations and can be expensive at inference time to reach accurate estimates of the desired posteriors.

The **stringent requirements of scientific applications are a powerful driver for improved methodologies**, but beyond the astrophysical scientific context motivating this research, **these tools also find broad applicability in many other domains**, including medical images [3].

PhD project: The candidate will aim to address these limitations of current methodologies, with the overall aim to make uncertainty quantification for large scale inverse problems faster and more accurate.

As a first direction of research, we will extend recent methodology concurrently developed by our team and our Ciela collaborators [4,5], based on Expectation-Maximization, to iteratively learn (or adapt) diffusion-based priors to data observed under some amount of corruption. This strategy has been shown to be effective at correcting for distribution shifts in the prior (and therefore leading to well calibrated posteriors). However, this approach is still expensive as it requires iteratively solving inverse problems and retraining the diffusion models, and is critically dependent on the quality of the inverse problem solver. We will explore several strategies including variational inference and improved inverse problem sampling strategies to address these issues.

As a second (but connected) direction we will focus on the development of general methodologies for sampling complex posteriors (multimodal/complex geometries) of non-linear inverse problems. Specifically we will investigate strategies based on posterior annealing, inspired from diffusion model sampling, applicable in situations with explicit likelihoods and priors.

Finally, we will apply these methodologies to some challenging and high impact inverse problems in astrophysics, in particular in collaboration with our colleagues from the Ciela institute, we will aim to improve source and lens reconstruction of strong gravitational lensing systems.

Publications in top machine learning conferences are expected (NeurIPS, ICML), as well as publications of the applications of these methodologies in astrophysical journals.

Scientific environment: The PhD will be carried out in the [CosmoStat laboratory](#) at the Departement d'Astrophysique at CEA Paris-Saclay under the supervision of Dr. [Francois Lanusse](#) and Dr. [Tobías I. Liaudat](#). The CosmoStat laboratory has a long tradition of developing cutting-edge statistical tools for the analysis of astronomical and cosmological data and is heavily involved in several projects including the ESA Euclid space telescope.

This project also falls in the context of a **collaboration with the [Ciela Institute](#) at the**



COSMOSTAT

Université de Montréal, which has strong connections to the Montreal-based [Mila](#) machine learning institute. Multiple visits to Montreal to work on these topics with the Ciela team are expected.

Computational resources: The successful candidate will have access to the [Jean Zay supercomputer](#), largest GPU cluster for research in France (which has recently been upgraded with 1400 H100 GPUs), as well as the IRFU's CPU cluster. Most of the development will rely on GPUs, through the JAX and XLA libraries.

Profile and skills required: To be able to make progress on the core scientific challenges of the project, the successful candidate should already be comfortable with software development and machine learning frameworks (in particular JAX). A background in signal/image processing, deep learning, statistical inference, would be greatly beneficial.

Contact:

- Dr. Francois Lanusse (francois.lanusse@cea.fr)
- Dr. Tobias I. Liaudat (tobias.liaudat@cea.fr)

References:

[1] Benjamin Remy, Francois Lanusse, Niall Jeffrey, Jia Liu, Jean-Luc Starck, Ken Osato, Tim Schrabback, *Probabilistic Mass Mapping with Neural Score Estimation*, <https://www.aanda.org/articles/aa/abs/2023/04/aa43054-22/aa43054-22.html>

[2] Tobias I Liaudat, Matthijs Mars, Matthew A Price, Marcelo Pereyra, Marta M Betcke, Jason D McEwen, *Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging*, RAS Techniques and Instruments, Volume 3, Issue 1, January 2024, Pages 505–534, <https://doi.org/10.1093/rasti/rzae030>

[3] Zaccharie Ramzi, Benjamin Remy, Francois Lanusse, Jean-Luc Starck, Philippe Ciuciu, *Denosing Score-Matching for Uncertainty Quantification in Inverse Problems*, <https://arxiv.org/abs/2011.08698>

[4] François Rozet, G r me Andry, Fran ois Lanusse, Gilles Louppe, *Learning Diffusion Priors from Observations by Expectation Maximization*, NeurIPS 2024, <https://arxiv.org/abs/2405.13712>

[5] Gabriel Missael Barco, Alexandre Adam, Connor Stone, Yashar Hezaveh, Laurence Perreault-Levasseur, *Tackling the Problem of Distributional Shifts: Correcting Misspecified, High-Dimensional Data-Driven Priors for Inverse Problems*, <https://arxiv.org/abs/2407.17667>