CRIStAL Centre de Recherche en Informatique, Signal et Automatique de Lille

PhD proposal
2025

Bayesian Estimate of Galaxy Cluster Masses for Improved Cosmological Parameter Inference

Estimation Bayésienne de la Masse des Amas de Galaxies pour une Meilleure Inférence des Paramètres Cosmologiques

Supervisors:

P. Chainais	`pierre.chainais@centralelille.fr`	`http://pierrechainais.ec-lille.fr/`
J. Sorce	`jenny.sorce@univ-lille.fr`	`https://sorcej.github.io/Jenny.G.Sorce/`
P.-A. Thouvenin	`pierre-antoine.thouvenin@centralelille.fr`	`https://pthouvenin.github.io/`

Laboratory: CRIStAL (`http://www.cristal.univ-lille.fr`), Centrale Lille - University of Lille

Teams: SIGMA team (*Signal, Modèles et Applications*)

Keywords: Inverse Problem, Bayesian inference, Galaxy Cluster Masses

Abstract. This PhD thesis proposal focuses on improving galaxy cluster mass estimation to enhance cosmological parameter inference, in order to address critical challenges such as the S_8 cosmological parameter discrepancy. By leveraging multi-wavelength observations, advanced statistical techniques including Bayesian inference and machine learning to solve this inverse problem, the project aims to refine relations between the cluster mass and their observable properties, reduce data dimensionality, and efficiently process large datasets. Anticipated outcomes include innovative Bayesian hierarchical models, optimized data reduction algorithms, and a robust inference pipeline for generating a high-fidelity galaxy cluster mass catalog towards testing cosmological models.

Scientific context. The standard cosmological model successfully describes the Universe's large-scale structure and evolution. Observational evidence, including the cosmic microwave background (CMB) and the hierarchical distribution of galaxies in the cosmic web, provides strong support for this model. These data suggest that dark matter and dark energy dominate the Universe, together constituting approximately 95% of its total content.

Despite its success, this model faces challenges. For instance, the value derived for the S_8 parameter, which characterizes the matter distribution on certain scales, differs by three standard deviations when using galaxy cluster counts instead of CMB observations. Resolving this tension is critical to determine whether new physics theories are required, or systematic errors in data analysis pipelines need to be accounted for.

Galaxy clusters are a key tool to address these challenges. As the largest gravitationally bound structures in the Universe, their number and masses are closely related to the underlying cosmology. Accurately estimating their masses and number across diverse environments and redshifts is essential to test the standard model and refine our understanding of the Universe [1].

This PhD project focuses on developing Bayesian hierarchical models to improve the estimation of galaxy cluster masses. By leveraging multi-wavelength observations, advanced data reduction, and machine learning techniques, this research will provide more precise mass estimates, refine scaling relations, and facilitate the inference of cosmological parameters from large datasets.

Objectives. This project will introduce new approaches to estimate galaxy cluster masses towards improving cosmological parameter inference. The research will achieve this by:

- developing techniques to extract accurate cluster observables (e.g., pressure, temperature) from multi-wavelength synthetic and observational images.
- refining scaling relations, such as those linking mass with those observables to reduce intrinsic variance and account for effects of redshift and environment.
- creating efficient inference techniques that combine data reduction and machine learning to handle large, multi-dimensional datasets while retaining critical information.
- applying the developed methods to observational data from surveys like Planck, SDSS, eROSITA, and Euclid to produce a refined catalog of galaxy cluster masses.

This work aims to enhance the accuracy and efficiency of cluster mass estimation, enabling robust tests of the standard cosmological model and potentially resolving discrepancies as those reported for the S_8 parameter.

Research Methodology. The research will follow a systematic approach, starting with synthetic data and progressing to real observational datasets. It begins with the development of techniques to estimate galaxy cluster observables using multi-wavelength data. These observables will serve as inputs for scaling relations, which link directly to fundamental cluster masses [1].

A significant part of the research will involve improving scaling relations by minimizing their intrinsic variance. To achieve this, statistical techniques such as kernel principal component analysis (PCA) and projection pursuit regression will be employed [2]. These methods will incorporate redshift and environmental dependencies, making the scaling relations more universally applicable across different types of clusters.

To manage the vast datasets produced by modern surveys, the project will also focus on data reduction techniques to images. Inspired by determinantal point processes [3], the approach will reduce the dimensionality of cluster data while retaining critical information for mass inference. This will enable faster processing and analysis without compromising accuracy.

Finally, inspired by [4], machine learning methods, such as convolutional neural networks (CNNs), will be used to infer cluster masses directly from raw and compressed data. The models will be trained on synthetic datasets with known masses to ensure accuracy before being applied to real observations. A key challenge will be accounting for observational systematics, such as noise and telescope artifacts, which will be modeled explicitly in the inference framework.

The final phase of the research involves applying the developed Bayesian techniques to observational data from major surveys like Planck, SDSS, and Euclid. The goal is to produce a comprehensive catalog of galaxy cluster masses to advance cosmological model testing.

Anticipated Outcomes. This project will make several significant contributions to data science and observational cosmology. It will produce new scaling relations with a 20–30% reduction in intrinsic variance compared to existing models, accounting for redshift and environmental effects. These relations will be applicable to a wide range of galaxy cluster types across different redshifts.

The data compression technique developed in this project is expected to achieve a compression ratio of at least 10,000:1 while preserving over 95% of the relevant information needed for property inference. This will make it feasible to handle the large-scale datasets generated by current and upcoming surveys.

The Bayesian inference pipeline based on its sophisticated modeling will enable rapid and accurate estimation of cluster masses, capable of processing multi-wavelength images for over 100,000 clusters in less than a day. These advancements will provide improved estimates of cluster masses complete with robust uncertainty quantification.

By applying these methods to observational data, the project will deliver a refined catalog of galaxy cluster masses. This catalog will be instrumental in testing the standard cosmological model and potentially resolving tensions such as the S_8 discrepancy.

Timeline. Within the first 18 months, this PhD project will develop techniques to estimate observables and refine scaling relations. The next phase will focus on data compression and mass inference, with the final year dedicated to applying the methods to observational data and producing the final catalog. Along the way, research findings will be published in peer-reviewed journals, culminating in a PhD thesis defense.

Conclusion. By refining the estimation of galaxy cluster masses, this research will address critical challenges in modern cosmology, including the S_8 tension. Leveraging state-of-the-art computational techniques, it will deliver tools and insights to test the standard cosmological model more rigorously and deepen our understanding of the Universe's fundamental components.

References

[1]. Lorenzo Lovisari, Ben J. Maughan, *Scaling relations of clusters and groups, and their evolution*, 2022, Chapter of the Section "Galaxy Clusters" (Section Editors: E. Pointecouteau, E. Rasia, A. Simionescu) of the "Handbook of X-ray and Gamma-ray Astrophysics" (Editors in chief: C. Bambi and A. Santangelo) (`https://arxiv.org/pdf/2202.07673`)

[2] Chad Schafer, Sukhdeep Singh, Yesukhei Jagvaral, *Reinterpreting fundamental plane correlations with machine learning*, 2024, MNRAS, 530, 3752 (`https://arxiv.org/pdf/2310.18177`)

[3] Claire Launay, Agnès Desolneux, Bruno Galerne, *Determinantal point processes for image processing*, 2020, SIAM Journal on Imaging Sciences, 14 (1), pp.304-348. (`https://hal.science/hal-02611259v2/file/DPP_for_image_processing.pdf`)

[4] Nicolas Cerardi, *Cosmologie - Amas de Galaxies - Intelligence artificielle*, 2024, (`https://theses.fr/s301274?domaine=theses`)