Advancing Cosmological Analysis: Data Transformation and Machine Learning to Simulate ARRAKIHS Observation

Wissarut Jarernsupapon[1], Krittapas Chanchaiworawit[2], Suwicha Wannawichian[1], Siramas Komonjinda[1]

[1] Department of Physics and Materials Science, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

[2] National Astronomical Research Institute of Thailand, Chiang Mai 50200, Thailand

Abstract

Galaxies' distribution and redshift behavior are important elements in cosmological research. ARRAKIHS (Analysis of Resolved Remnants of Accreted Galaxies as a Key Instrument for Halo Surveys) is an upcoming ESA (European Space Agency) space mission designed to further advance our understanding of the universe. ARRAKIHS aims to image faint galaxies located near our Milky Way galaxy. Thus, ARRAKIHS can aid us in the investigation of the nature of dark matter and galaxy formation by utilizing the data to compute differential number density (dN/dz). This study utilizes the dataset from COSMOS2020, which is a comprehensive catalog for approximately 1.7 million sources across a 2 square degree area of the sky with data in multi-wavelength photometry and photometric redshifts. The dataset is used to simulate the observations that ARRAKIHS would make by transforming it into the framework of ARRAKIHS. The framework transformation is proceeded by adapting the COSMOS2020 dataset to match the observational parameters of ARRAKIHS, such as its wavelength bands and transmission. By utilizing machine learning techniques to predict redshifts from the transformed data, ARRAKIHS' potential to provide accurate and reliable measurements of galaxy redshift distributions will be assessed. This is done by comparing the simulated distribution from the transformed data with the actual data from COSMOS2020, which will highlight ARRAKIHS' capability to measure and provide a constraint for galaxy redshift distributions. The advancement in the measurement of galaxy redshift distributions will significantly deepen our understanding of the universe, especially in testing the Λ-Cold Dark Matter (ΛCDM) cosmological model and the role of dark matter in galaxy formation.

Introduction

The ARRAKIHS mission (Analysis of Resolved Remnants of Accreted Galaxies as a Key Instrument for Halo Surveys) is a newly proposed ESA Fast (F-class) mission designed to explore the nature of dark matter and galaxy evolution on small scales. This mission aims to capture highly detailed images of faint galactic halos surrounding Milky Way-like galaxies, reaching exceptionally low surface brightness levels in visible and near-infrared bands. By surveying over 200 nearby galaxy halos (located 15–30 Mpc away), ARRAKIHS will assemble a statistical dataset of tidal stellar streams and satellite galaxies that serve as fossil records of past galaxy mergers. This mission provides a crucial window into galactic-scale cosmological structures, complementing large-scale studies of the distant universe.

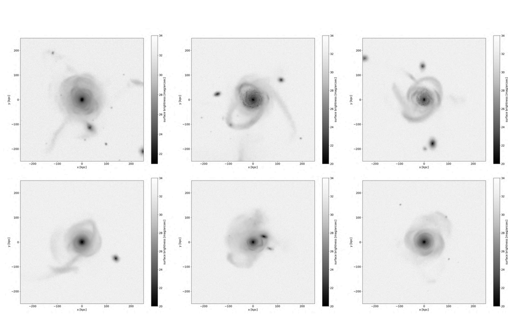

Fig 1: *Galaxies and their environment extracted from the FIREbox computer simulation of our Universe. Each panel shows examples of gravitational interactions between the central and satellite galaxies, which ARRAKIHS aims to observe in reality.*

Fig 2: *Binocular optical system developed by the Spanish micro satellites company Satlantis for the ARRAKIHS Instrument.*

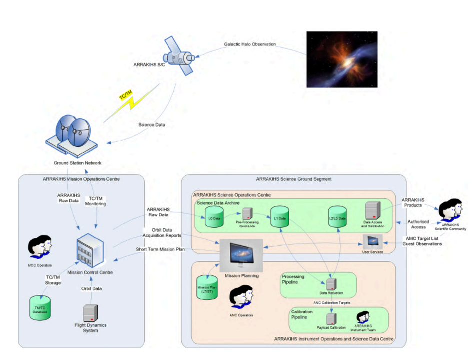

Fig 3: *Diagram of ARRAKIHS ground segment components.*

The COSMOS survey is a key resource in extragalactic astronomy, providing deep multi-band imaging over 2 deg^2. The latest COSMOS2020 release offers a wide-range view of the universe, cataloging ~1.7 million sources, and includes ultra-deep optical/IR coverage, with some bands reaching AB ~30. This dataset improves photometric redshift precision and identifies galaxies up to z ~ 10. COSMOS2020 achieves sub-percent redshift accuracy for bright galaxies and ~5% precision for the faintest ones. It maps the evolving number density of galaxies, informs galaxy evolution models, and serves as a benchmark for photometric redshift techniques and machine learning. With its rich multi-wavelength data, COSMOS2020 is crucial for simulating ARRAKIHS observations and testing the ΛCDM model.

Method

Our approach combines data from COSMOS2020, machine learning models, and instrument simulations to predict ARRAKIHS' outcomes. The methodology involves several steps: transforming COSMOS photometric data into ARRAKIHS's observational framework, applying machine learning to estimate redshifts, and simulating ARRAKIHS's constraints (throughput) to compute expected galaxy redshift distributions.

We first start by getting the information about all filters of COSMOS2020, then we keep only the filters that have overlapped band wavelengths.

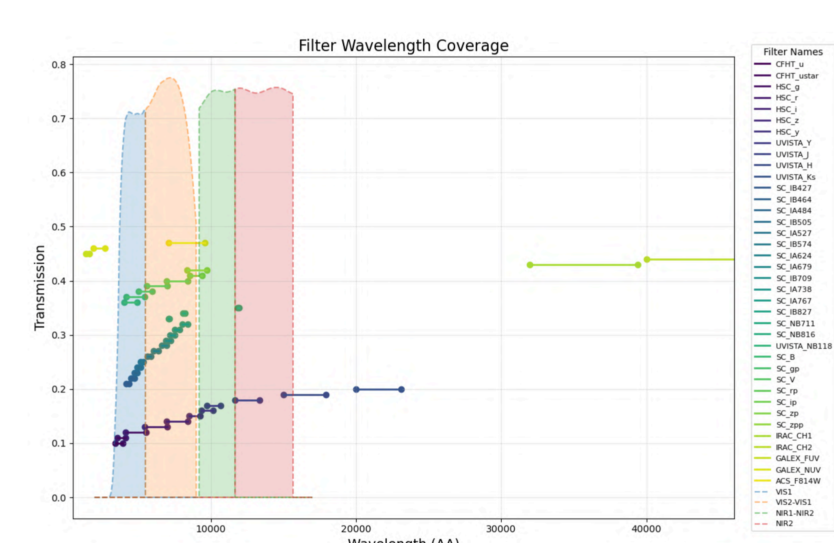

Fig 4: *Diagram of ARRAKIHS and COSMOS2020 filters coverage.*

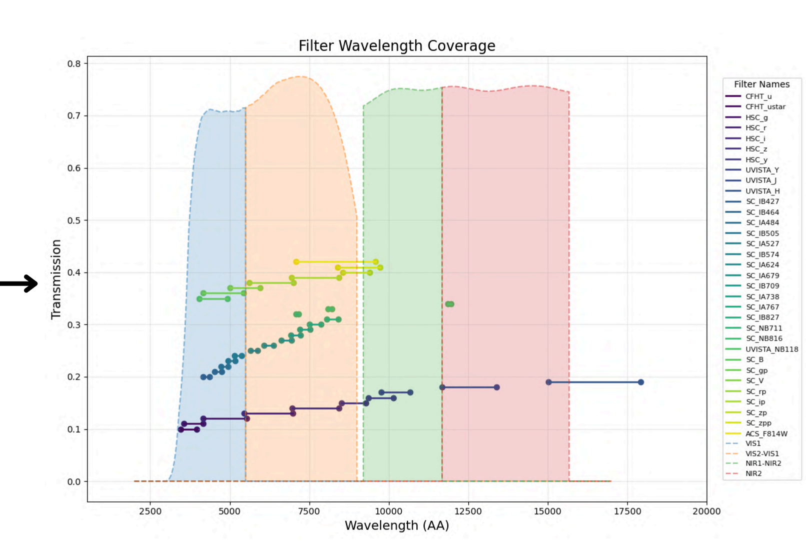

Fig 5: *Diagram of ARRAKIHS and COSMOS2020 filters coverage after considering only the overlapped ones.*

Then we start transforming flux frequency from COSMOS2020 into flux wavelength, generating step functions for each object where y-axis is flux wavelength (arbitrary unit) and x-axis is wavelength (AA), interpolating data in gaps, extrapolating data to fit the coverage of ARRAKIHS' filters, performing convolution with ARRAKIHS' throughput and integrating the resulting function.

(a)

$$\int_{\lambda_1}^{\lambda_2} F_\lambda(\lambda)d\lambda = \int_{\nu(\lambda_2)}^{\nu(\lambda_1)} F_\nu(\nu)d\nu$$

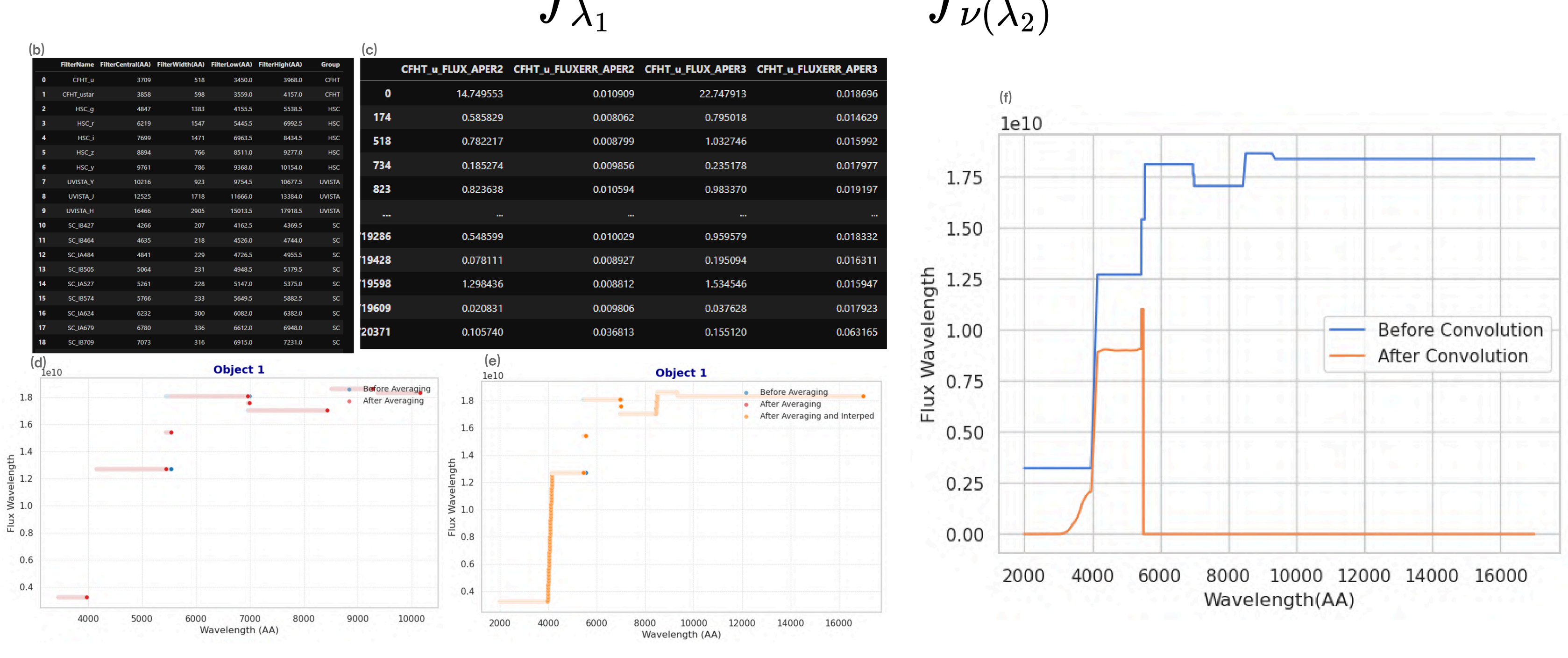

Fig 6: *(a) The relation between flux frequency and flux wavelength, (b) examples of COSMOS2020 filters information, (c) examples of COSMOS2020 data after being filtered, (d) example of the step function of COSMOS2020, (e) example of the step function after interpolating and extending, and (f) example of convolution between flux wavelength and ARRAKIHS' filters throughput*

We can then use the result with the machine learning techniques to estimate redshifts.

Results and Discussion

The preliminary results are promising. The trend and errors of a control set of data, which is a set of data before transforming into ARRAKIHS' framework, can predict redshift with RMSE = 0.4827 and the result of the transformed framework dataset has RMSE = 0.665. Note that the machine learning algorithms used is a basic random forest with default setting, meaning RMSE of the transformed framework can be further reduced in the future.

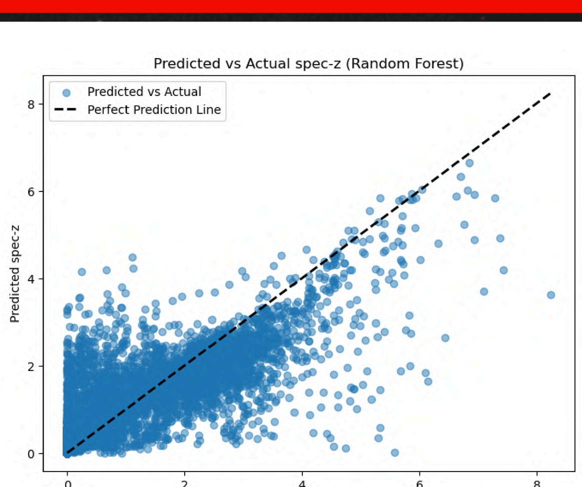

Fig 7: *Comparison between prediction and actual redshift*

Conclusion

The techniques for transforming dataset of a framework to another framework can be used to simulate a dataset that a future space mission can measure. And since the techniques are used by performing convolution of a set of filters and the data from another set of filters, it may be possible to perform such simulations for other missions not limited to space missions. Nevertheless, the simulation using COSMOS2020 proves that ARRAKIHS can be used to further study Cosmology by studying dN/dz using only 4 sets of filters. This opens a door to a whole new way to look at the cosmo.

Acknowledgement

We would like to thank Department of Physics and Materials Science, Faculty of Science, Chiang Mai University, the COSMOS team, and the ARRAKIHS team. We would like to also thank Dr. Krittapas Chanchaiworawit, Assistant Professor Dr. Suwicha Wannawichian, and Assistant Professor Dr. Siramas Komonjinda for having a vast contribution for the research.

References

1. European Space Agency. (2022). ARRAKIHS Mission. Retrieved from https://www.arrakihs-mission.eu/
2. Weaver, J. R. (2021). COSMOS2020 Galaxy Catalog. Retrieved from https://astroweaver.github.io/project/cosmos2020-galaxy-catalog/
3. Guzmán Llorente, R., et al. (2024). ARRAKIHS: ESA's New Fast-Implementation Science Mission. Proceedings of SPIE, 13092, 3020186. https://doi.org/10.1117/12.3020186