

Simulation of Baryon Acoustic Oscillations and Detection Thresholds

September 30, 2026

EPFL - Laboratory of Astrophysics
Internship Report

Author

ROMIEUX Nathan
nathan.romieux@epfl.ch

Supervisor

VERDIER Aurélien
aurelien.verdier@epfl.ch

Abstract

This report presents research conducted within the astrophysics laboratory of EPFL (LAS-TRO), aimed at analyzing the detectability of Baryon Acoustic Oscillations (BAO). Initially focused on processing calibration data from the 4MOST spectroscopic survey, the study concentrated on validating the spectral analysis pipeline. Results demonstrate high reliability of the automation for redshift determination and object classification (LRG, ELG, QSO, stars), with a concordance rate exceeding 95% for measurements associated with an adequate confidence index. Faced with statistical limitations in the validation data (insufficient volume and complex geometry), the project was reoriented towards the numerical simulation of cosmological structures. By generating synthetic universes integrating an artificial BAO signature at 147 Mpc, the impact of the numerical galaxy density ($\bar{n}$) and the signal proportion on the signal-to-noise ratio (SNR) of the two-point correlation function $\xi(r)$ was quantified. The simulations highlight the existence of a critical density threshold below which Poisson noise masks the BAO signature, rendering its detection impossible. These findings illuminate the strategy of current major cosmological surveys (DESI, 4MOST) : having satisfied this density condition, their major challenge now lies in maximizing the surveyed volume to increase the number of accessible modes and measurement precision.

Contents

1	Introduction	3
1.1	History	3
1.2	Theory	3
1.3	Methodology	4
2	Pipeline Verification for Universe Mapping	4
2.1	Position	4
2.2	Type	5
2.3	Results	5
3	BAO Simulation	6
3.1	Motivation	6
3.2	Universe with BAO	6
3.3	Detectability	7
4	Conclusion	9

1 Introduction

1.1 History

A few seconds after the Big Bang, the universe was a hot plasma in which photons (light), electrons, and baryons (matter) were mixed. At these temperatures, photon-electron interactions trapped light within matter. Quantum fluctuations generated overdense regions, and gravity attracted the plasma to these points. This adiabatic compression increased the momentum of the fluid's photons, meaning the temperature increased locally. Highly energetic photons collided with free electrons, transferring momentum to them. The electrons then dragged protons along via the Coulomb force. Thus, gravity pulled baryons and photons inward into the overdensity, while compressed photons pushed electrons outward, dragging protons with them. This coupling generated pressure waves similar to sound waves in air. They are called Baryon Acoustic Oscillations (BAO). As the universe expanded, the plasma cooled. 380,000 years after the Big Bang, a critical temperature threshold was crossed : the average photon energy became lower than the ionization energy of hydrogen, and the first hydrogen atoms formed. These atoms interact very little with photons. Consequently, light decoupled from matter and traveled in straight lines over cosmological distances. This primordial light, observed in all directions, is called the Cosmic Microwave Background (CMB).

With this decoupling, the waves froze at a precise measurable radius. Matter density was higher on the wavefronts. Consequently, a greater number of galaxies formed in these locations. In measurements, only the first wavefront is observed, although several existed. This is because the BAO signature is mathematically described by a sinc-type function, which theoretically possesses an infinite number of oscillations (multiple peaks) but is multiplied by an exponential damping term that suppresses them all [2]. When observing the universe, spherical shells with a point at their center are not directly visible as in the accompanying representation (Fig.1), as they have been mixed with expansion and gravitational effects. The presence of these shells is validated by observing an overrepresentation of galaxies separated by the same distance [3] : the radius of the sphere. This distance stretches with global expansion, measuring this radius at different epochs allows crucial information about our universe to be derived.

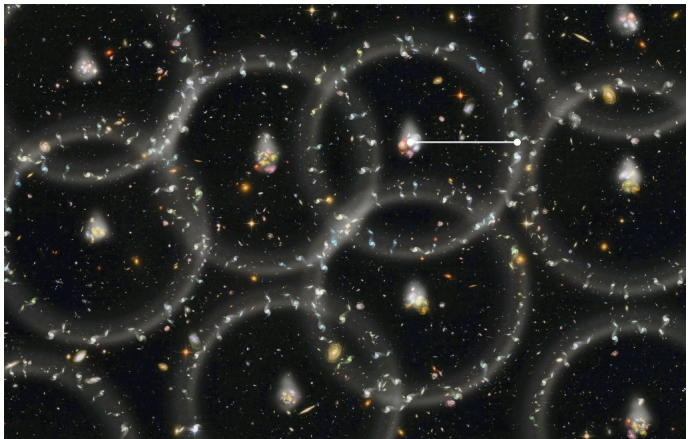


Figure 1: Artistic 2D representation of BAO shells (here circles) with their central overdensity.[1]

1.2 Theory

To statistically quantify this overrepresentation of galaxy pairs at a characteristic scale, cosmologists use the two-point correlation function, denoted $\xi(r)$. This measures the excess probability, compared to a purely random (Poisson) distribution, of finding two galaxies separated by a distance r . Formally, the differential probability dP of finding an object in a volume dV at a distance r from another object is written [4]:

$$dP = \bar{n} [1 + \xi(r)] dV, \quad (1)$$

where $\bar{n}$ is the mean numerical density of galaxies. In a universe without structure, $\xi(r) = 0$, a value $\xi(r) > 0$ indicates matter clustering. In practice, $\xi(r)$ is estimated by comparing observed data (D) to a simulated reference catalog (R), called *randoms*, which reproduces the geometry and selection effects of the observation but without any physical correlation. The standard estimator, known as Landy-Szalay [4], minimizes variance due to the edges of the observed field and is expressed as:

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}, \quad (2)$$

where DD , DR , and RR represent the normalized numbers of galaxy pairs separated by a distance r in the data-data, data-randoms, and randoms-randoms catalogs, respectively. The precision of this measurement depends crucially on the sampling density : as demonstrated later in this study, an insufficient density $\bar{n}$ introduces dominant Poisson noise ($\propto 1/\bar{n}$) that can completely mask the BAO signature, rendering the construction of random catalogs and the calculation of $\xi(r)$ ineffective below a critical threshold.

1.3 Methodology

Initially, the intention was to measure this signal in new data produced by the European project 4MOST (4-metre Multi-Object Spectroscopic Telescope) as part of a calibration phase prior to the launch of large-scale measurements. To achieve this, the pipeline used to determine the position and type of observed objects first had to be validated. Subsequently, random catalogs necessary for the calculation of the correlation function were to be generated. However, due to technical constraints related to the enormous size of the catalogs involved, the fact that these data were not designed for this type of measurement (low density in many zones instead of high density in a single zone), and the difficulty of the task, the decision was made to simulate the structure of the universe and the BAO phenomenon in order to study the density prerequisites for its detection.

2 Pipeline Verification for Universe Mapping

2.1 Position

As the angular position of observed objects was known, their radial distances remained to be determined. For this purpose, the redshift phenomenon was utilized. Waves emitted by a source in motion relative to the observer are stretched or compressed (Fig.2). For a source moving away from the observer, the wavelengths of the received light are stretched. An object does not emit at a single precise wavelength, it emits a continuous signal. This signal can be decomposed on a graph presenting light intensity as a function of wavelength, this graph is the object's spectrum. Due to the redshift phenomenon, the received wavelengths are longer, on an increasing graph, a shift to the right is therefore observed. By recognizing certain known patterns associated with the studied objects, for example, constant gaps between emission or absorption peaks (Fig.3), it is possible to estimate the extent of the signal shift and thus the speed at which the source is receding.

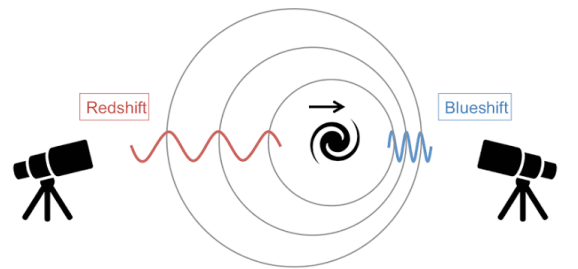


Figure 2: Explanatory diagram of the Doppler effect.

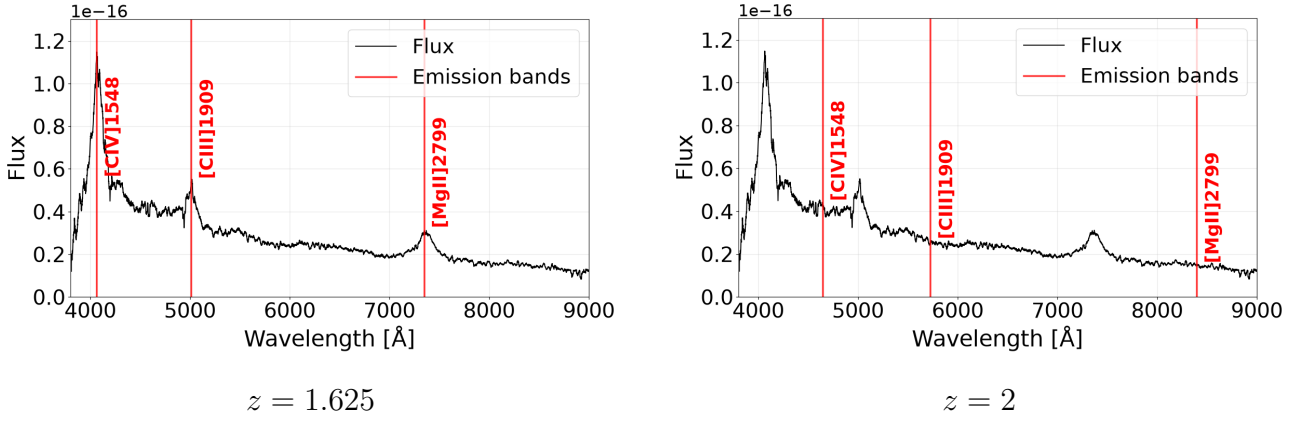


Figure 3: Spectrum of a QSO at $z = 1.1625$ and at $z = 0$. Emission bands shifted to the redshift value and to an arbitrary value to highlight the presence of characteristic bands with precise spacing used to determine the redshift.

As the universe is expanding homogeneously, the farther an object is, the faster it recedes. One can imagine a large balloon being inflated homogeneously : two close points recede less quickly than two distant points because less matter is expanding between them. For galaxies relatively close to us, the velocity related to expansion is given by $v = d \cdot H(z)$, where H is the Hubble constant. Given that the peculiar velocity of galaxies is negligible compared to the expansion velocity, determining the redshift yields the velocity and then the distance. In the general case, the distance is found via the redshift z according to the following equation [5] :

$$D_C = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')} \quad (3)$$

where c is the speed of light in a vacuum, H_0 is the Hubble-Lemaître constant, and $E(z)$ is an expansion factor related to the evolution of the universe's content.

2.2 Type

During cosmological analysis, observed objects must be classified to treat biases related to their observations separately. To validate the pipeline, the spectra of 4 types of objects had to be recognized : LRG, ELG, QSO, STAR. Respectively : Luminous Red Galaxy, Emission Line Galaxy, Quasi-Stellar Object. A confidence level ranging from 1 to 4 was associated with each measured redshift.

2.3 Results

Figure 4.a characterizes the correlation between the type and the confidence level of the found redshift. It is observed that if a spectrum is classified as a galaxy, the confidence level of its redshift is high (mean of 3.36). When the spectrum quality does not allow type detection, it generally does not allow redshift detection with a good confidence level (mean of 1.11). Stars are generally rated lower (mean < 3) than galaxies and QSOs. In the analyzed sample, two-thirds of the redshifts have a confidence level greater than or equal to three (Fig.4.b). Finally, Figure 5 illustrates the high reliability of the pipeline when its certainty percentage exceeds 40%, with more than 95% of the redshifts equal to the manual verification (z_{vi}) to within one hundredth.

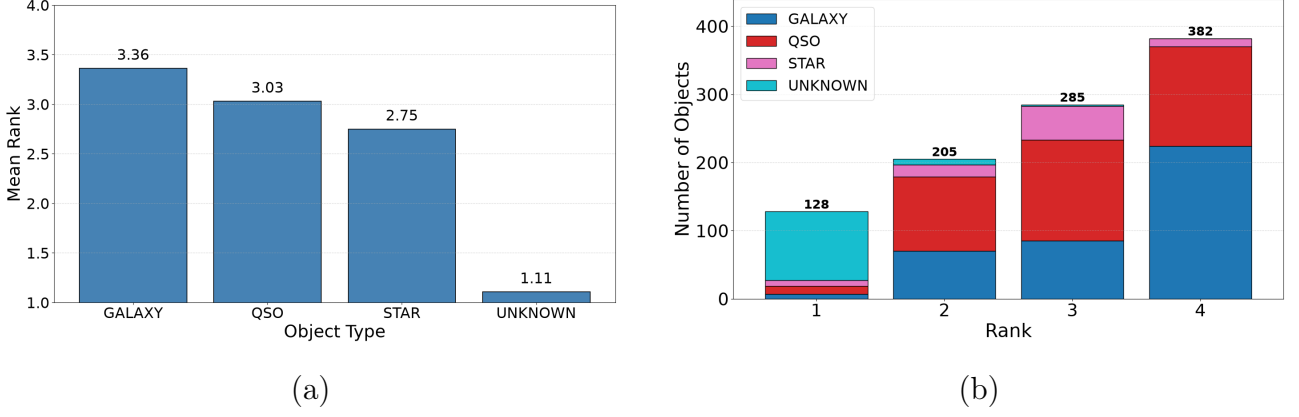


Figure 4: (a) Mean rank by type. (b) Distribution of object types by rank.

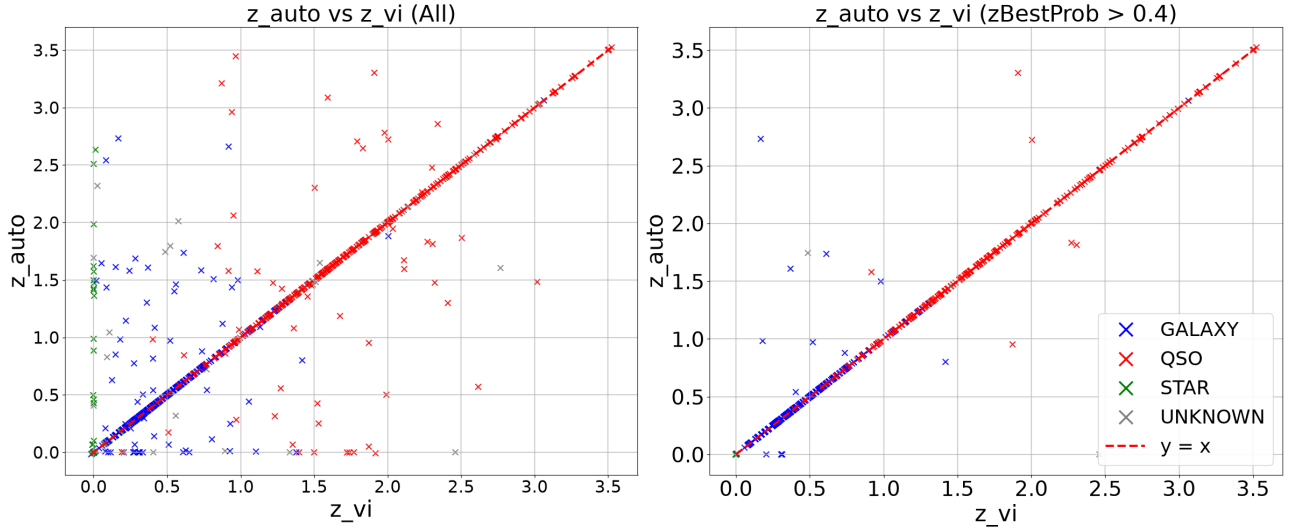


Figure 5: Graph of z_{auto} (pipeline) vs z_{vi} (manual), with the entire sample redshifts on the right and only redshifts with a pipeline certainty value greater than 0.4 on the left.

3 BAO Simulation

3.1 Motivation

Initially, large quantities of 4MOST spectra with a pipeline confidence level greater than 40% were used to attempt a measurement of the BAO signal. Random catalogs were generated to use the standard method [4], which corrects for complex geometry and selection effects. However, this catalog was a validation survey intended to verify target selection, spectrum quality, and measurement pipelines before the launch of the actual survey. Attempts to measure the BAO signal were unsuccessful because the data were not compatible with this type of analysis (scattered galaxies and insufficient quantity). Consequently, the project was reoriented towards signal simulation to characterize its detectability.

3.2 Universe with BAO

The fractal structure of the universe was simulated (see Fig.6) by defining three cluster sizes (small, medium, large). Their parameters (number and size) were set empirically so that the correlation function followed a power law as in reality [6]. Once the correct structure was

obtained, spherical shells with nuclei were injected to simulate the BAO. The radius was fixed at $r = 147$ Mpc, as observed in the literature [7]. The simulation cube has a density of $4 \cdot 10^{-6} \text{ h}^3 \cdot \text{Mpc}^{-3}$, which is 100 times lower than measured real densities [8]. This density choice was dictated by RAM and time constraints.

3.3 Detectability

Once the simulation was successful, detectability was studied by varying the proportion of points allocated to "shell + nucleus" systems as well as the total number of galaxies (thus the density). Figure 7 presents the correlation functions of repeated simulations, varying first the density, then the BAO proportion.

It is recalled that the correlation function $\xi(r)$ follows a power law $\xi(r) \propto r^{-\gamma}$ [6], manifesting as a decreasing straight line in a log-log diagram, upon which the characteristic bump of baryon acoustic oscillations (BAO) is superimposed. When the mean numerical density $\bar{n}$ decreases, a vertical upward shift of this line is observed, accompanied by an apparent attenuation of the BAO peak. This phenomenon is explained by the very definition of the correlation function, $\xi(r) = \frac{n(r)}{\bar{n}} - 1$ [9], where $n(r)$ is the conditional local density dictated by the fractal geometry. In a regime where $\xi(r) \gg 1$, a reduction in the sampling density $\bar{n}$ mechanically increases the measured amplitude of the correlation according to the relation:

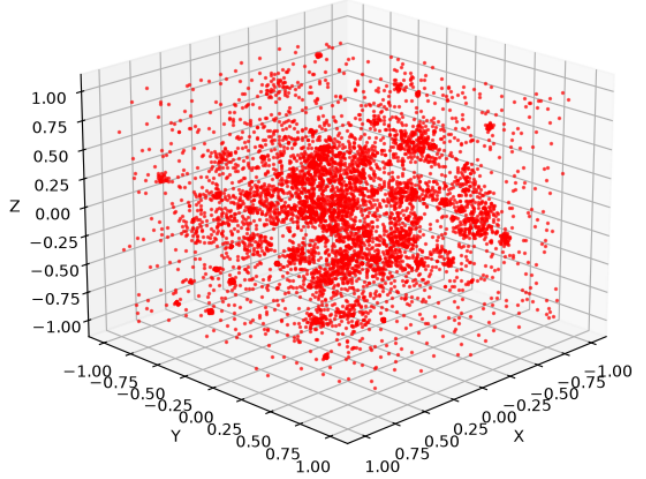


Figure 6: Simulation of the universe in a cube of $\approx 8.13 \text{ Gpc}^3$ with a density of $4 \cdot 10^{-6} \text{ h}^3 \cdot \text{Mpc}^{-3}$. Each point represents a galaxy. 10% of points were allocated to shells and nuclei (BAO).

$$\xi_{\text{measured}}(r) \approx \frac{n(r)}{\bar{n}} \propto \bar{n}^{-1} r^{-\gamma}. \quad (4)$$

In logarithmic space, this translates to a vertical translation of the curve : $\log \xi(r) = -\gamma \log r - \log \bar{n} + C$. Simultaneously, the decrease in $\bar{n}$ exacerbates Poisson noise, the additive term of which degrades the signal-to-noise ratio. This increased background noise reduces the relative contrast of the BAO, $\Delta\xi/\xi$, giving the illusion of a decrease in peak amplitude while its physical position remains unchanged, drowned in growing statistical variance. For the variation of the proportion, the result is more intuitive. Indeed, the lines remain relatively superimposed, and the peak decreases as the proportion decreases. For a proportion of 5%, the peak is no longer visually visible. However, in real surveys where the peak is visible, the BAO proportion lies between 1 and 5%. This difference is explained by the density in the simulations being approximately 100 times lower than in reality, which drowns the signal in statistical noise as seen previously. The evolution of the signal-to-noise ratio (SNR) as a function of the number of galaxies N (Fig.8) highlights a clear transition : a low-efficiency regime for low densities, followed by a critical threshold. This behavior suggests the existence of a minimum required density for the discretization of the density field to allow the emergence of the BAO signal above background fluctuations. Below this threshold, the rarity of galaxy pairs at the characteristic scale likely prevents a robust estimation of the correlation function, regardless of the underlying model. The heatmap (Fig.9) of the signal-to-noise ratio (SNR) as a function of the BAO

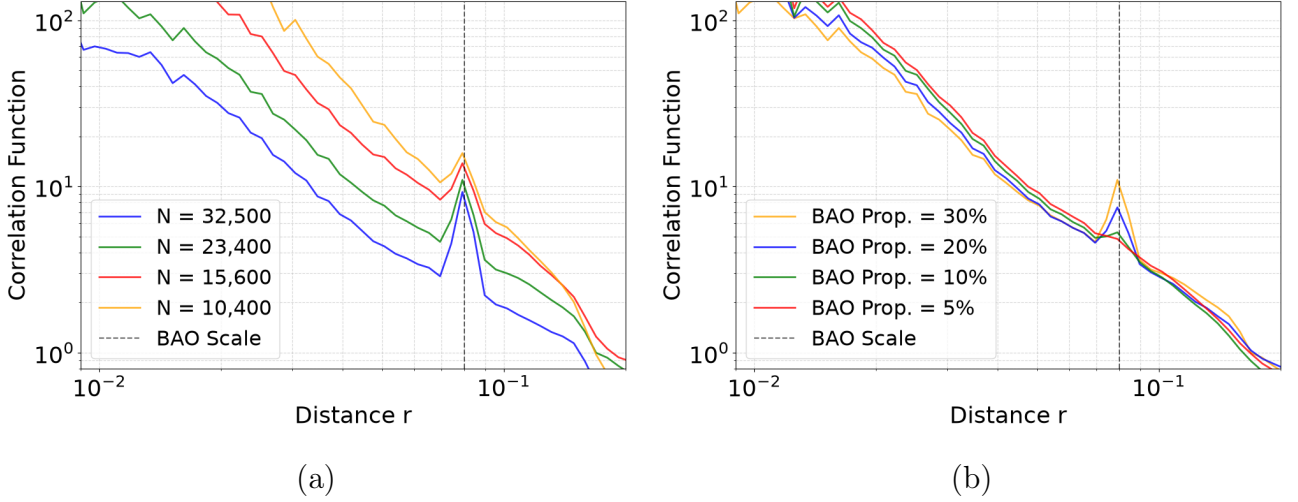


Figure 7: (a) Correlation function for different sample sizes N on a log scale (30% BAO). (b) Correlation function for different BAO proportions on a log scale ($N = 18000$).

proportion and $\sqrt{N}$ presents an apparent diagonal color gradient, suggesting a linear relationship between the two parameters. Although the fundamental theoretical relationship follows a hyperbolic law ($A_{BAO} \propto \text{SNR}/\sqrt{N}$), the observed visual linearity is explained by the restriction of the numerical exploration domain. Indeed, the simulations cover a restricted interval of low densities, imposed by computational cost constraints. On this limited segment of the parameter space, the curvature of the theoretical hyperbola is negligible, and the function can be approximated with high precision by its local tangent, thus appearing as a diagonal straight line. Furthermore, the discretization of simulation points (low sampling in N) accentuates this "staircase" effect which visually smooths the remaining curvature.

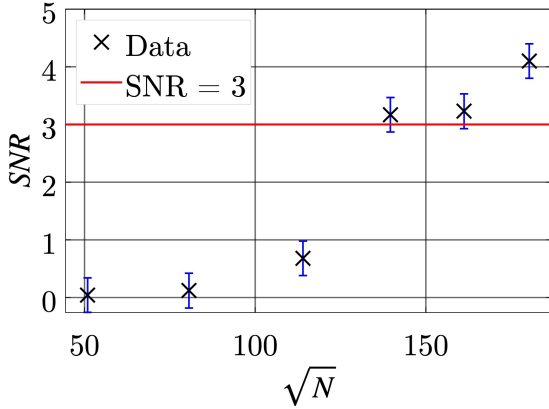


Figure 8: Signal to noise ratio vs $\sqrt{N}$ (30% BAO).

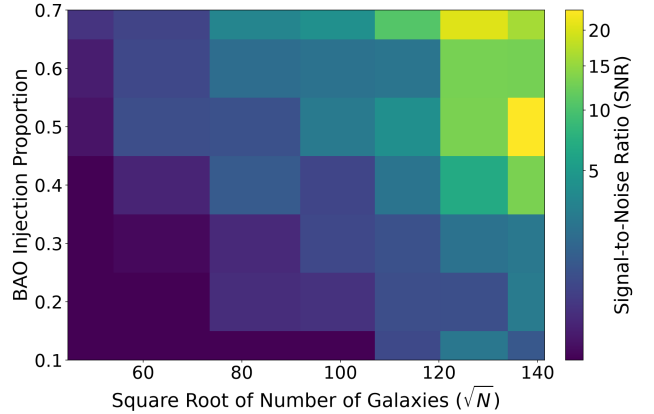


Figure 9: BAO Detection Map : Sensitivity to Sample Size and Signal Strength.

This result, although influenced by the numerical limits of the study, remains physically instructive. It demonstrates that in the explored density regime, the trade-off between signal amplitude and sampling density behaves in a quasi-linear manner. This implies that, within this observation window, every marginal gain in density directly and proportionally compensates for a loss in signal amplitude, an empirical rule useful for optimizing observation strategies with limited computational cost. The simulations highlight the existence of a minimum density threshold below which the BAO signal remains undetectable, validating the necessity of prior statistical resolution. This constraint illuminates the strategy of contemporary major surveys

(DESI, 4MOST) : having now crossed this critical density threshold, the community no longer seeks to increase it, but maximizes the surveyed volume to increase the total number of modes N . Volume expansion allows precision limits to be pushed back without modifying the local density, transforming the BAO into a robust standard ruler for precision cosmology.

4 Conclusion

This internship allowed methodological challenges linked to modern observational cosmology to be addressed, ranging from instrumental validation to theoretical modeling. The first phase, dedicated to the analysis of 4MOST project data, confirmed the robustness of redshift determination and spectral classification algorithms. The close correlation between object type and redshift confidence level, as well as the near-perfect concordance with manual verifications for high-quality spectra, represents the first steps towards validating the processing pipeline for future large-scale surveys. However, the impossibility of detecting the BAO signal in the calibration data, due to an insufficient quantity of galaxies and a fragmented observation geometry, led to the second phase of the project : simulation. By artificially recreating the fractal structure of the universe and injecting a characteristic BAO scale, this study allowed critical detectability parameters to be isolated. The analysis of simulated correlation functions demonstrates that the sampling density $\bar{n}$ is a determining parameter : below a critical threshold, Poisson noise dominates pair statistics, erasing the acoustic oscillation signature despite the physical presence of the signal. These works recall that robust BAO detection relies on a fundamental balance between sampling density and surveyed volume. If current major spectroscopic surveys (such as DESI or 4MOST) have now reached the necessary critical densities, making signal extraction possible, their major challenge today lies in extending the explored cosmological volume to reduce statistical errors. In this context, this simulation study confirms that density remains a *sine qua non* condition : it validates the minimum threshold from which increasing the volume becomes effectively relevant. Finally, this internship provided an opportunity to develop practical skills in massive data processing, numerical simulation, and statistical analysis applied to astrophysics.

Acknowledgements

I would like to particularly thank my supervisor, Aurélien Verdier, for accepting me into their lab and for the guidance they provided. I also extend my gratitude to the Laidlaw Foundation for this amazing opportunity and enriching experience, as well as to Ms. Dan-Thi for her attentive support. I would also like to thank all the other EPFL Laidlaw Scholars I spent the summer with. I am so happy to have met you, I am finishing this internship with a head full of wonderful memories and a big smile on my face !

References

- [1] Zosia Rostomian and SDSS-III Collaboration. Boss measures the universe to one-percent accuracy, jan 2014. Artistic illustration by Zosia Rostomian (LBNL) accompanying the announcement of the results of SDSS-III BOSS.
- [2] Daniel J. Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496:605, 1997.

- [3] D. J. Eisenstein and I. Zehavi. Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *The Astrophysical Journal*, 633(2):560–574, nov 2005.
- [4] Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, July 1993.
- [5] David W. Hogg. Distance measures in cosmology. *arXiv preprint astro-ph/9905116*, 1999. Version 4, December 2000.
- [6] M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267:465–482, April 1983.
- [7] Planck Collaboration, N. Aghanim, et al. Planck 2018 results. VI. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, sep 2020.
- [8] A. G. Adame et al. DESI 2024 III: baryon acoustic oscillations from galaxies and quasars. *Journal of Cosmology and Astroparticle Physics*, 2025(04):012, 2025.
- [9] Peter H Coleman and Luciano Pietronero. The fractal structure of the universe: Correlations, averages, and the sample size. *Physics Reports*, 213(6):311–326, 1992.