Studying inflation with quasars from the DESI spectroscopic survey

Edmond Chaussidon

université
PARIS-SACLAY

Studying inflation with quasars from the DESI spectroscopic survey

Etude de l'inflation avec les quasars du relevé spectroscopique de DESI

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 576, particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)
Spécialité de doctorat: Science des Astroparticules et Cosmologie
Graduate School : Physique
Référent: Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche
Département de Physique des Particules (Université Paris-Saclay, CEA),
sous la direction de **Christophe YECHE**, Directeur de Recherche.

Thèse soutenue à Paris-Saclay, le 22 septembre 2023, par

Edmond CHAUSSIDON

Composition du jury

Membres du jury avec voix délibérative

Sophie HENROT-VERSILLE Directrice de recherche, IJCLab, Université Paris-Saclay	Présidente
Johan RICHARD Astronome, CRAL, Observatoire de Lyon, Université Claude Bernard Lyon 1	Rapporteur & Examinateur
Benjamin D. WANDELT Professeur, IAP, CNRS, Sorbonne Université	Rapporteur & Examinateur
Emanuele CASTORINA Professeur assistant, University of Milan	Examinateur
Nathalie PALANQUE-DELABROUILLE Directrice de recherche, Lawrence Berkeley National Laboratory	Examinatrice

THESE DE DOCTORAT
NNT : 2023UPASP099

université PARIS-SACLAY

ÉCOLE DOCTORALE
Particules, hadrons, énergie et noyau : instrumentation, imagerie, cosmos et simulation (PHENIICS)

Titre: Etude de l'inflation avec les quasars du relevé spectroscopique de DESI

Mots clés: *Cosmologie, Inflation, Relevé spectroscopique, Structures à grande échelle, DESI*

Résumé: Depuis plusieurs décennies, l'inflation est considérée comme le paradigme principal pour décrire l'Univers primitif. Les théories décrivant l'inflation peuvent être testées et améliorées à partir de la mesure de différentes propriétés telles que l'inclinaison du spectre de puissance scalaire primordial, les ondes gravitationnelles primordiales ou la non-gaussianité primordiale (PNG). En effet, les modèles d'inflation les plus simples (single-field slow roll inflation) prédisent une distribution presque gaussienne des fluctuations primordiales, c'est-à-dire une quantité minimale de PNG alors que les modèles plus complexes comme par exemple l'inflation à champs multiples prédisent des quantités significatives de PNG. Pour contourner la limite de variance cosmique des observations du fond diffus cosmologique qui donne les meilleures contraintes actuelles sur les PNGs, une possibilité serait d'utiliser l'énorme puissance statistique des regroupements de galaxies en 3D, en sondant un grand volume de l'Univers. En particulier, une approche prometteuse consiste à utiliser l'empreinte minuscule laissée à grande échelle sur le spectre de puissance de la matière par la PNG locale, connue sous le nom de biais dépendant de l'échelle. C'est l'approche que nous suivrons dans cette thèse en analysant les 1,2 millions de quasars du relevé spectroscopique de la première année d'observation du Dark Energy Spectroscopic Instrument (DESI). Après avoir donné le cadre théorique permettant d'expliquer les phénomènes mis en jeu, je commencerai par décrire l'échantillon de quasars (QSO) de DESI. En particulier, je présenterai la sélection des cibles QSOs utilisées dans DESI, dont j'ai eu la charge pendant la Survey Validation, phase préliminaire au relevé spectroscopique, ainsi que le pipeline spectroscopique pour collecter le décalage vers le rouge. Cette sélection permet de collecter plus de 200 quasars par degré carré incluant 60 d'entre eux avec redshift (z) plus grand que 2,1, améliorant de plus de 20% les exigences du comité scientifique de DESI. Ensuite je présenterai le traitement des effets observationnels qui pourraient contaminer la mesure des PNGs et la façon de les traiter. Afin d'éviter tout biais de confirmation lors de cette étape de correction, je détaillerai une méthode pour masquer le signal caractéristique des PNGs. Finalement, je donnerai la mesure masquée des PNGs avec les 1,2M de quasars de la première année d'observation ainsi que les contraintes attendues. Avec cette première année d'observation, nous nous attendons d'ores et déjà à une meilleure sensibilité qu'avec le dernier relevé spectroscopique de quasars fourni par l'expérience eBOSS en 2022. Il faudra en revanche attendre le relevé complet de DESI et employer de nouvelles approches afin de pouvoir concurrencer la meilleure mesure actuelle effectuée par le satellite Planck.

Title: Studying inflation with quasars from the DESI spectroscopic survey

Keywords: *Cosmology, Inflation, Spectroscopic survey, Large Scale Structure, DESI*

Abstract: For several decades, inflation has been considered the main paradigm for describing the early Universe. The theory describing inflation can be tested and improved by measuring properties such as the tilt of the primordial scalar power spectrum, primordial gravitational waves or primordial non-Gaussianity (PNG). Indeed, the simplest inflation models (single-field slow-roll inflation) predict an almost Gaussian distribution of primordial fluctuations, i.e. a minimal amount of PNG, whereas more complex models such as multi-field inflation predict significant amounts of PNG. To circumvent the cosmic variance limit of cosmic microwave background observations, which give the best current constraints on PNGs, one possibility would be to use the enormous statistical power of 3D galaxy clustering, by probing a large volume of the Universe. In particular, one promising approach is to use the tiny imprint left on the large-scale matter power spectrum by local PNGs, known as the scale-dependent bias. This is the approach we will follow in this thesis, analyzing the 1.2 million quasars in the spectroscopic survey of the first year of observation by the Dark Energy Spectroscopic Instrument (DESI). After providing a theoretical framework to explain the phenomena at play, I will begin by describing the DESI quasar sample (QSO). In particular, I'll present the selection of QSO targets used in DESI, for which I was responsible during Survey Validation, the preliminary phase of the spectroscopic survey, as well as the spectroscopic pipeline for collecting the redshift. This selection enabled us to collect more than 200 quasars per square degree, including 60 with redshift (z) greater than 2.1, improving on the requirements of the DESI scientific committee by more than 20%. Next, I'll present how to deal with observational effects that could contaminate the measurement of PNGs. In order to avoid any confirmation bias during this correction stage, I will detail a method for masking the characteristic PNG signal. Finally, I'll give the masked measurement of PNGs with the 1.2M quasars from the first year of observation, along with the expected constraints. With this first year of observations, we already expect a better sensitivity than with the last spectroscopic survey of quasars provided by the eBOSS experiment in 2022. Nevertheless, we will have to wait for DESI's complete survey, and use a new approach to be competitive with the best measurement currently being made by the Planck satellite.

[†]IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

Studying inflation with quasars from the DESI spectroscopic survey

Edmond Chaussidon[†]

Paris,
October 9, 2023

Contents

Remerciements

Merci tout le monde, j'y suis arrivé ! Quoiqu'un peu simpliste, ce travail n'aurait pu exister sans bon nombre d'entre vous.

« Tout d'abord, je tenais à remercier Christophe et Nathalie de m'avoir proposé un sujet de thèse si passionnant et la bourse qui va avec, il y a presque 4 ans déjà. Je voulais aussi remercier personnellement Christophe pour son accueil et son encadrement, en particulier pendant la période du Covid. Ses fines connaissances sur la politique interne de la collaboration, ses nombreuses bonnes idées et son enthousiaste débordant m'ont poussé durant ces trois années. Merci de m'avoir encadré ! J'en profite pour remercier Julien de m'avoir accueilli au LBNL pour mon stage de M2 et qui m'a fait découvrir avec brio la cosmologie observationnelle et plus particulièrement DESI. Cet avant-goût californien y est certainement pour beaucoup dans la suite des évènements.

Je tenais également à remercier les membres de mon jury qui ont acceptés d'évaluer mon travail de thèse. En particulier à M.Richard et M.Wandelt d'avoir accepté de lire mon manuscrit en plein été.

Je salue également tous les membres du groupe de cosmo du département de physique des particules qui m'ont permis de passer trois riches années au CEA. En particulier, Etienne et Jean-Baptiste pour leurs nombreuses remarques et conseils lors de la préparation de ma soutenance mais également pour bon nombre de présentations. Une pensée pour ces slides que j'avais si bien préparées avec vous, Etienne et Christophe, à l'autre bout du monde et que je n'ai malheureusement pas eu la chance de présenter.

Je remercie bien évidemment tous les jeunes (Alexandre, Mathilde, Marie-Lynn) sans qui les pauses café ou les retours en bus auraient été bien plus long. A tous ceux avec qui j'ai pu partager le bureau 41B (Marie, Julien, Rafik, Henri) en ayant une pensée particulière à Corentin qui doit être en train de faire ses cartons pour partir à Clermont. J'espère que nos routes se recroiseront. Je souhaite également avoir une pensée pour Antoine avec qui j'ai effectué mes 3 années de thèse, je lui souhaite bon courage pour la semaine prochaine et tout le bonheur et la réussite dans sa vie future face au lac et aux montagnes. Je tenais aussi à remercier sincèrement Arnaud pour nos nombreuses discussions au cours de ces trois années et de toute l'aide que tu as pu me fournir, sans laquelle ce travail n'aurait pas été le même. Ça y est tu vas être considéré comme un vieux maintenant.

J'ai également une pensée pour tous ceux avec qui j'ai pu passer des superbes meetings de collaborations, mais aussi de riches écoles d'été pas toujours reposantes.

J'aimerais remercier tous mes amis, qui doivent travailler pour la plupart à l'heure actuelle, sans qui je n'aurais pas une vie aussi riche et si bien remplie. La liste est longue et je ne veux pas me risquer à en oublier, alors que vous soyez de ceux avec qui j'ai fait ma prépa, Supaéro, mon M2 ou alors provenant d'une rencontre fortuite verre à la main, merci du fond du cœur pour

tous ces moments passés à vos côtés ! Cette thèse aurait été bien différente sans les nombreux jours de télétravail passés chez Lena et Marine et les soirées improvisées. Ainsi que ceux passés chez Pierre et les pique-niques hebdomadaires sur le canal Saint-Martin à refaire le monde. Je tenais à saluer tous mes camarades avec qui j'ai joué, bataillé sur de bons nombres de terrains d'abord en jaune et noir, puis en bleu et blanc et finalement, par chance, une nouvelle fois en jaune et noir. C'est l'école de la vie qu'ils disent ! Finalement, j'ai une pensée émue à mes compagnons d'aventure de longue date (Bil, Pit, Bub, Ska, Jeannot, Touf) avec qui j'ai grandi.

Évidemment, je ne serais pas devant vous aujourd'hui sans le soutien indéfectible de ma famille. A mes deux grands mères avec qui j'ai passé de nombreux étés. François pour ta bienveillance en mon égard pour mes faibles connaissances sur l'histoire de France. Pour Armand et Anne, je suis très heureux et très fier de vous avoir pour frère et sœur ! Et enfin, merci Papa et Maman, je ne pouvais pas rêver meilleure jeunesse et éducation. Vous m'avez toujours permis d'être celui que je voulais devenir.

Finalement, je tenais à remercier Audrey pour ces presque deux récentes années que nous venons de passer ensemble. Merci de me supporter, ce n'est pas facile. Anne et Armand pourront le confirmer. Je nous souhaite le meilleur de l'autre côté de l'Atlantique.

Maintenant, Pot ! ≫

A Saclay, le 22 septembre 2023,
Edmond Chaussidon

Introduction

Remember to look up at the stars and not down at your feet. Try to make sense of what you see and wonder about what makes the Universe exist. Be curious.

Stephen Hawking

Cosmology, the scientific study of the Universe at its largest scales, seeks to unravel the mysteries of its origin, evolution and ultimate destiny. It encompasses a wide range of subjects, including the Big Bang theory, the structure and behavior of large-scale structures in the Universe, and the nature of dark matter and dark energy. Relevant dates are given in the timeline in Table 1.

The birth of "modern" cosmology can be traced back to the groundbreaking work of Albert Einstein and his theory of general relativity in 1915, which revolutionized our understanding of gravity by proposing a space-time dynamically influenced by matter and energy. The expansion of our Universe was predicted a few years later by Georges Lemaître and Alexander Friedmann, who independently solved Einstein's equations. Observational confirmation, known as the Hubble-Lemaitre law, was one of the most important turning points in modern physics.

Long questioned, the Big Bang theory, which describes an expanding Universe, was widely adopted after the discovery in 1964 by Arno Penzias and Robert Wilson of the cosmic microwave background radiation. Its discovery irrefutably confirmed that our Universe was hotter and denser in its early days. The cosmic microwave background has been an important object of study through the COBE, WMAP and Planck missions, and is still at the heart of major observational programs.

In 1998, observations of type Ia supernovae, a particular class of exploding white dwarf stars, showed that the expansion of the Universe was accelerating rather than decelerating under the influence of gravity alone. This led to the discovery of dark energy, whose existence and behavior are currently the biggest open question in cosmology.

To answer this burning issue, another set of data is crucial: observations of the large-scale structures of the Universe. Galaxies and cosmic voids form a vast cosmic web, revealing the distribution of matter across the Universe. The observation of galaxy redshifts, first carried out by the CfA Redshift Survey in the late 1970s, then extended by the Sloan Digital Sky Survey in the 2000s, and currently performed by the Dark Energy Spectroscopic Instrument, enables us to map the Universe with great precision, and to measure the expansion rate of the Universe at different epochs.

All these observations also seek to shed light on the first moments of our Universe. For several decades, inflation has been considered the leading paradigm for describing the early Universe. Inflation is a set of many models that can be tested and improved from the measurement of different properties such as the tilt of the primordial scalar power spectrum, the primordial gravitational waves, or the primordial non-gaussianity (PNG). The first is already

well constrained with the latest Planck CMB data, and the second is gaining a growing interest with future missions to observe the B-mode polarization of the cosmic microwave background. At variance, PNG remains to date still poorly constrained by current experiments, and it is the main subject of this dissertation.

The simplest inflation models predict a nearly Gaussian distribution of primordial fluctuations *i.e.* a minimal amount of PNG. In particular, the detection of local non-gaussianity: $\mathcal{O}\left(f_{NL}^{loc}\right) \sim 1$ will rule out single field inflation and will hint at more complex models *e.g.* multi-field inflation.

Currently, the best constraints on PNG are obtained from Planck data: $f_{NL}^{loc} = -0.9 \pm 5.1$. To circumvent the cosmic variance limit of CMB observations, one possibility would be to use the enormous statistical power in the 3D galaxy clustering, probing a large volume of the Universe. A promising approach is through the tiny imprint left at large scales on the matter power spectrum by local PNG, known as the scale-dependent bias. With this method, the Dark Energy Spectroscopic Instrument (DESI) is expected to constrain local PNG with similar accuracy to Planck. Although the measurement will strongly depend on the tracer linear bias value and its response to PNG, the quasar sample seems to be the target of choice, since it probes a very large volume.

This dissertation will first introduce the theoretical background required to study the large-scale structures of the Universe in Chapter 1. Then, it will briefly give an overview of the early Universe described by the inflation in Chapter 2. The Dark Energy Spectroscopic Instrument will be described in Chapter 3. Chapter 4 will explain the target selection for the quasars in DESI. Finally, Chapter 5 will show the measurement of primordial non-gaussinity with the DESI Y1 blind data.

This dissertation took full advantage of several useful references:

- Dodelson and Schmidt (2020)
- Lecture notes from Hannu Kurki-Suonio[1]
- Baumann (2022)
- en Français: Peter and Uzan (2005) et Bernardeau (2007)

My work was made much easier thanks to the use of open source packages (`CLASS`, `cosmoprimo`, `pycorr`, `pypower`, `mpytools`, `mockfactory`, `fastpm`, `desilike`), their associated software (in particular, all the scientific python environment), and all open-source community for their numerous responses in miscellaneous forums.

[1] https://www.mv.helsinki.fi/home/hkurkisu/

Table 1: Timeline of selected major events in Cosmology

1915	Alber Einstein publishes the theory of general relativity
1927 - 1929	Georges Lemaître and Edwin Hubble discover independently the expansion of the Universe
1933	Fritz Zwicky obtains the first signs of the existence of dark matter
1946	George Gamow proposes primordial nucleosynthesis, a mechanism for the formation of atomic nuclei in the early stages of an expanding Universe
1964	First observation of the Cosmic Microwave Background by Penzias & Wilson
1970	Vera Rubin measures the rotation of spiral galaxies, confirmation of existence of dark matter
1977	Start of CfA Redshift Survey
1978	**Nobel Prize** (A.Penzias & R.Wilson): "for their discovery of cosmic microwave background radiation"
1980s	Guth, Linde, Albrecht & Stein propose inflation to describe the early Universe
1989	Launch of COBE satellite
1990	Launch of Hubble Space Telescope from the Discovery Shuttle
1998	Discovery of the accelerated expansion of the Universe and introduction of dark energy
2000	Start of Sloan Digital Sky Survey
2001	Launch of WMAP satellite
2004	First detection of Baryonic Acoustic Peak with SDSS galaxy survey
2006	**Nobel Prize** (J.Mather & G.Smoot): "for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation"
2009	Launch of Planck satellite
2011	**Nobel Prize** (S.Permutter, B.Schmidt & A. Reiss): "for the discovery of the accelerating expansion of the Universe through observations of distant supernovae"
2016	First observation of gravitational waves by LIGO/Virgo
2017	**Nobel Prize** (R.Weiss, K.Thorne & B.Barish): "for decisive contributions to the LIGO detector and the observation of gravitational waves"
2019	**Nobel Prize** (J.Peebles): "for theoretical discoveries in physical cosmology"
2020	**Nobel Prize** (R.Penrose): "for the discovery that black hole formation is a robust prediction of the general theory of relativity", (R.Genzel, A.Ghez): "for the discovery of a supermassive compact object at the center of our galaxy"
2021	Start of DESI Main survey
2021	James Webb Space Telescope is launched from Ariane V
2023	Launch of Euclid satellite

Chapter 1

Cosmology with large-scale structures

Contents

Toutes les lois sont donc tirées de l'expérience ; mais pour les énoncer, il faut une langue spéciale ; le langage ordinaire est trop pauvre, il est d'ailleurs trop vague, pour exprimer des rapports si délicats, si riches et si précis. Voilà donc une première raison pour laquelle le physicien ne peut se passer des mathématiques ; elles lui fournissent la seule langue qu'il puisse parler.

Henri Poincaré, La valeur de la science (1905)

The advent of special relativity, developed by Albert Einstein in 1905, marked a significant transformation in classical physics by introducing the concept of space-time and linking energy and mass. While special relativity established the foundation, it was not until the invention of general relativity by Einstein in 1915 that a theory respecting special relativity was available to explain the gravitational force. Although it already explained the precession of Mercury's perihelion, general relativity was first validated in 1919 with the measurement of the deflection of light by the sun that it predicted (Eddington 1919, Dyson et al. 1920). This paved the way for the beginning of modern cosmology.

This chapter provides a succinct overview of the theory necessary to comprehend the origin and evolution of large-scale structures in the Universe. For in-depth information, please refer to the list of references provided in the introduction (p. viii).

1.1 A smooth and expanding Universe

1.1.1 General relativity

1.1.1.1 Equivalence principle & Metric

General relativity is a successful relativistic description of gravitation. It relies on the *equivalence principle* stating: at every point in space-time in an arbitrary gravitational field, it is possible to choose a *local* inertial coordinate system such that in a sufficiently small region around of this point, the law of nature takes the same form as in special relativity. In this local inertial coordinate system $X^\mu = (t, x, y, z)$, the linear element $\mathrm{d}s^2$ is given by the Minkowski metric:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = -c^2\mathrm{d}\tau^2, \tag{1.1}$$

where c is the speed of light and τ the *proper time*. We choose, the following convention for the Minkowski tensor: $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$.

Finding the local inertial coordinate system at each point of the space and working with it, is usually something complicated, and it may be better to work with a global coordinate system x^μ such that:

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu, \tag{1.2}$$

where $g_{\mu\nu}(x)$ is the metric tensor defined at each space-time position by

$$g_{\mu\nu}(x) = \eta_{\alpha\beta}\frac{\partial X^\alpha}{\partial x^\mu}\frac{\partial X^\beta}{\partial x^\nu}. \tag{1.3}$$

Following the Weinberg derivation, the equation of motion of a free particle can be derived directly from the equivalence principle. A free-falling particle does not accelerate in the local inertial freely falling coordinate system such that:

$$\frac{\mathrm{d}^2 X^\mu}{\mathrm{d}\tau^2} = 0. \tag{1.4}$$

The equation of motions, called the *geodesic equations*, can then be written in any coordinate system, x^μ, in applying a simple change of coordinates:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0, \tag{1.5}$$

where $\Gamma^\mu{}_{\alpha\beta}$ are the *Christoffel symbols*:

$$\Gamma^\mu{}_{\alpha\beta} = \frac{\partial x^\mu}{\partial X^\lambda}\frac{\partial^2 X^\lambda}{\partial x^\alpha \partial x^\beta} = \frac{1}{2}g^{\mu\lambda}\left(\partial_\alpha g_{\lambda\beta} + \partial_\beta g_{\lambda\alpha} - \partial_\lambda g_{\alpha\beta}\right), \tag{1.6}$$

using the useful notation for the partial derivate: $\partial_\mu = \partial/\partial x^\mu$.

In a flat space *i.e.* without any gravitational interactions, the geodesics are straight lines. In the presence of gravity, the space is not flat anymore, it is said to be *curved*, and the geodesics give the trajectory of a free particle undergoing gravity. All the information of how a particle moves in a gravitational field is then contained inside the metric $g_{\mu\nu}$.

1.1.1.2 Einstein equations

Motivated by Newtonian physics, the Einstein equations aim to describe the dynamical evolution of $g_{\mu\nu}$ *i.e.* describe the curvature of spacetime as a function of the distribution of matter and energy within it. Einstein proposed:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \tag{1.7}$$

where κ and Λ are free parameters, $G_{\mu\nu}$ is the Einstein tensor and $T_{\mu\nu}$ is stress–energy tensor.

The Einstein tensor is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R, \tag{1.8}$$

where $R = R^\mu{}_\mu$ is the Ricci scalar, $R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu}$ is the Ricci tensor and $R^\lambda{}_{\mu\kappa\nu}$ is the Riemann curvature tensor describing the local curvature of the spacetime. If $R^\lambda{}_{\mu\kappa\nu} = 0$ everywhere, the space is then completely flat and the coordinate system is equivalent to the Minkowski metric. The Riemann curvature tensor is linked to the metric $g_{\mu\nu}$ by

$$R^\lambda{}_{\mu\kappa\nu} = \partial_\kappa \Gamma^\lambda{}_{\nu\mu} - \partial_\nu \Gamma^\lambda{}_{\kappa\mu} + \Gamma^\lambda{}_{\kappa\omega}\Gamma^\omega{}_{\nu\mu} - \Gamma^\lambda{}_{\nu\omega}\Gamma^\omega{}_{\kappa\mu}. \tag{1.9}$$

The stress-energy tensor represents the distribution of matter and energy in spacetime. In particular, the stress-energy tensor for a perfect fluid is given by

$$T^{\mu\nu} = \left(\rho + \frac{P}{c^2}\right)u^\mu u^\nu + P g^{\mu\nu}, \tag{1.10}$$

where ρ is the energy density, u^μ is the four-velocity of the fluid, P is the pressure. Formally, in the fluid rest frame, T^{00} represents the energy density, T_{0i} the energy flux density which is equal to the momentum density T_{i0} and T_{ij} represents the spatial stress tensor. The conservation of energy in the fluid rest frame reads as

$$\partial_\mu T^{\mu\nu} = 0. \tag{1.11}$$

These equations give in the non-relativistic limit the well-known equation of continuity ($\partial\rho/\partial t + \boldsymbol{\nabla}\cdot(p\mathbf{v}) = 0$) and the Euler equations ($\rho\,\partial\mathbf{v}/\partial t + \rho\,(\mathbf{v}\cdot\boldsymbol{\nabla})\,\mathbf{v} + \boldsymbol{\nabla}P = 0$); see, for instance, Rieutord (2014). To conserve the equality in any general coordinate system, the partial derivate needs to be replaced by the covariant derivate $\boldsymbol{\nabla}_\mu$:

$$\boldsymbol{\nabla}_\mu T^{\mu\nu} \equiv \partial_\mu T^{\mu\nu} + \Gamma^\mu{}_{\mu\gamma} T^{\gamma\nu} + \Gamma^\nu{}_{\mu\gamma} T^{\mu\gamma} = 0, \tag{1.12}$$

Such an equation is called *covariant conserved*: thanks to the equivalence principle if it is true in an inertial rest frame, it is true under this transformation in any coordinate system.

The free parameter κ is chosen to recover the Poisson equation which gives the classical description of the gravity in the Newtonian limits[1]: $\kappa = \dfrac{8\pi G}{c^4}$.

The latest free parameter is the famous cosmological constant Λ, responsible for the late acceleration of the expansion of the Universe. The best current measurement gives $\Lambda = 1.1056 \times 10^{-52}\ \mathrm{m}^{-2}$ (Planck Collaboration et al. 2020) (see Section 1.3.2).

The energy conservation (1.12), is automatically verified in the Einstein equations (1.8), since by construction the Riemann curvature tensor follows the *Bianchi identity* such that

$$\boldsymbol{\nabla}_\mu G^{\mu\nu} = 0. \tag{1.13}$$

Energy conservation is a fundamental requirement for any theory of gravity. The *Lovelock theorem* (Lovelock 1972) ensures that the Einstein tensor is the only tensor, with the metric tensor itself, which contains only up to second derivatives of the four-dimensional spacetime metric, linear in the second derivatives of the metric and be covariant conserved *i.e.* divergent free along one index. Thus, the proposed Einstein equations are the only possible equations using up to second derivatives of the metric. Although artificially introduced by Einstein (1917) in order to obtain a static *i.e.* not expanding Universe[2], the presence of Λ is not just an artifact, it is mathematically promoted with this set of equations. Today, its presence explains the late accelerated expansion of the Universe (Riess et al. 1998, Perlmutter et al. 1999).

These equations are one possibility to describe gravity in the framework of special relativity. Several potential extensions, such as using high-order derivatives of the metric, are already being considered. However, the current Einstein equations have been extensively tested and validated up to solar system scales through numerous experiments and observations: the gravitational deflection of light by massive objects (Dyson et al. 1920), the perihelion of Mercury (Einstein 1916) or the detection of gravitational waves (Abbott et al. 2016). Hence, the Einstein equations will be assumed to be in the following.

Remark: In the following and as often in cosmology, we will set $c = 1$ and $\hbar = 1$.

[1] The Newtonian limit supposing that the gravitational fields are weak, particles move slowly compared to the speed of light and the gravitational fields vary slowly in time. Under these assumptions, geodesic equations (1.5) are reduced to $\mathrm{d}^2x^i/\mathrm{d}\tau^2 = 1/2\ \partial^i h_{00}$ where h is a small perturbation of the Minkowski metric such that $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. In classical mechanics, the Poisson equation $\nabla^2\Phi = 4\pi G\rho_m$ leads to $\mathrm{d}^2x^i/\mathrm{d}\tau^2 = -\partial^i\Phi$. Hence, $g_{00} = -(1+2\Phi)$. Comparing the 00 Einstein equation (1.8) to the Poisson equation gives $\kappa = \frac{8\pi G}{c^4}$.

[2] At that time, the only objects observed were stars with very small peculiar velocities.

1.1.2 Homogeneous description

1.1.2.1 A homogeneous and isotropic Universe: the cosmological principle

Describing gravity in the Universe requires finding a coordinate system in which we can easily describe the distribution of matter. First, consider the distribution of the matter at large scales.

Fig. 1.1a shows a fraction of the galaxy distribution observed by the Sloan Digital Sky Survey (SDSS)[3]. The different colors represent different types of galaxies but all the galaxies trace the matter across the Universe. Although an internal structure in the galaxy distribution is clearly visible on small scales ($< 100\ h^{-1}$ Mpc), it seems that the distribution is isotropic and homogeneous on large scales.

The isotropy is also observed in the temperature fluctuation around $T_{\rm CMB} = 2.7255\ K$ of the cosmic microwave background (CMB) as observed by ESA Planck mission[4] in Fig. 1.1b. CMB is the oldest light in the Universe that can be observed today. It was emitted when the Universe was just 380 000 years old. At first order, the temperature is isotropic, the fluctuations in Fig. 1.1b are very small, $\delta T/T \sim 10^{-5}$, and correspond to regions with slightly different densities which are the seed of the future large-scale structures.

The *Copernican principle*, which states the Earth is not a special place in the Universe, then implies the *cosmological principle*: the Universe is isotropic and homogeneous.

1.1.2.2 An expanding Universe: the Hubble diagram

A cornerstone of observational cosmology is the discovery of the expansion of the Universe. Lemaître and Hubble observed that galaxies are moving away from Earth at speeds proportional to their distance, described by the *Hubble-Lemaître law*

$$v = H_0 \times D, \tag{1.14}$$

where v is the speed of a galaxy moving away from the Earth, D is the distance between the galaxy and the Earth and H_0 is called the Hubble constant and represents the expansion rate of the Universe at the present time. The first observation and publication of this relationship were made by Lemaître (1927), as recorded in his French work. However, when he translates into English in Lemaître (1931), this aspect of his work was omitted. Meanwhile, Hubble made also the observation of (1.14) two years later in Hubble (1929).

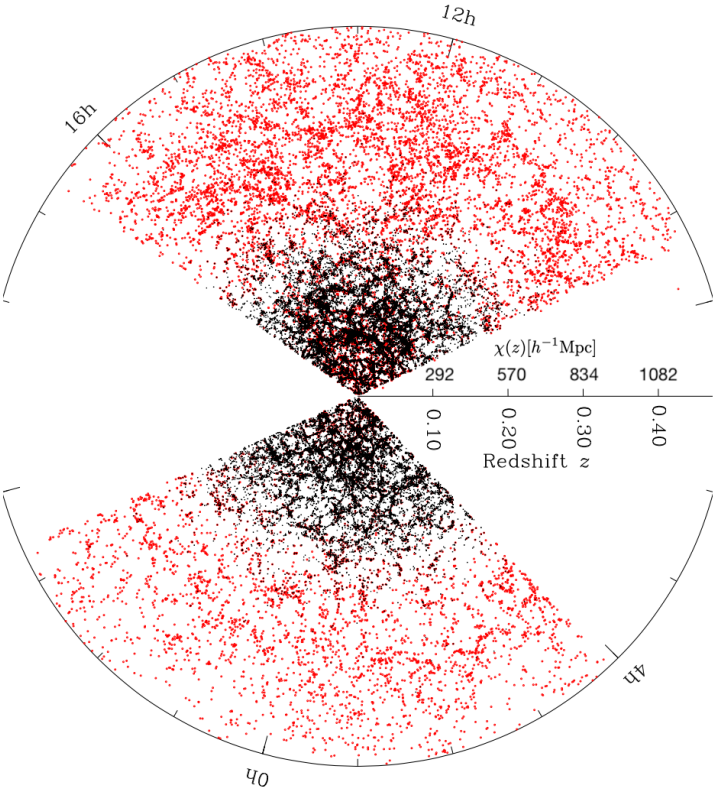
Fig. 1.2a shows the diagram of distance-velocity used by Lemaître in 1927 to obtain the first empirical value of the Hubble constant: $H_0 = 575$ km s^{-1} Mpc^{-1}, as reproduced by the solid black line (Block 2012). In 1929, Hubble published similar diagram, shown in Fig. 1.2b, where he found 530 km s^{-1} Mpc^{-1}.

These two measurements are the first observation of the expansion of the Universe which is today well observed. The best constraint on the Hubble constant is from Planck ESA mission: $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} (Planck Collaboration et al. 2020). As discussed in Section 1.3.2, the Hubble constant can be also derived thanks to the supernovae: $H_0 = 73.5 \pm 1.1$ km s^{-1} Mpc^{-1} (Brout et al. 2022). This is the world-famous H_0 tension, and it is one of the burning issues of the ΛCDM model.

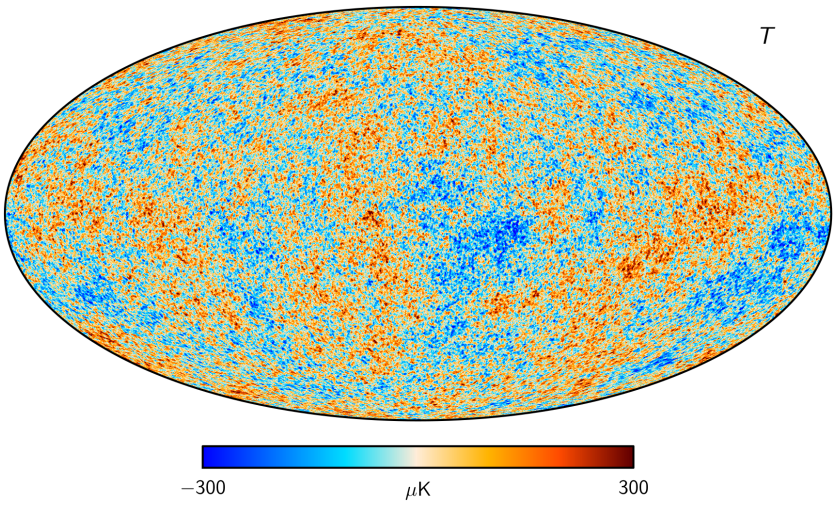
Remark: Distances in astronomy are generally given in megaparsec:

$$1\ \text{Mpc} = 3.26 \times 10^6\ \text{ly} = 3,086 \times 10^{22}\ \text{m}.$$

[3] https://www.sdss.org/
[4] https://www.esa.int/Science_Exploration/Space_Science/Planck

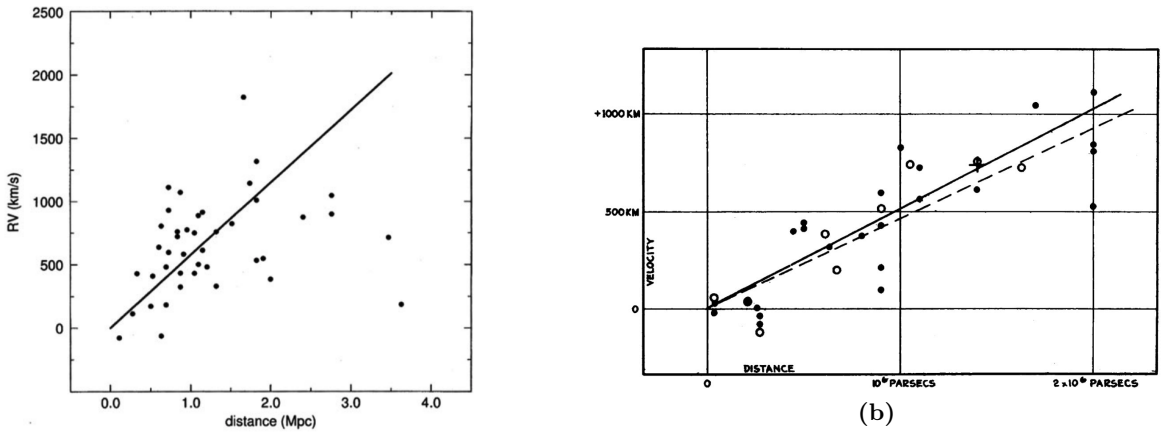


(a)



(b)

Figure 1.1: (a) Projected slice of the galaxy distribution observed by the Sloan Digital Sky Survey (SDSS). The black (*resp.* red) points are the main galaxy sample (*resp.* the luminous red galaxies). Coherent structures are visible at small scales ($< 100\ h^{-1}$Mpc), while the Universe seems quite isotropic at large scales ($> 100\ h^{-1}$Mpc) Credit: Michael Blanton and SDSS collaboration. (b) Distribution of the temperature fluctuation around $T_{\rm CMB} = 2.7255\ K$ of the cosmic microwave background (CMB) as observed by the ESA Planck mission. The relative fluctuation is about $\delta T/T \sim 10^{-5}$. Credit: Planck Collaboration.



(a) (b)

Figure 1.2: (a) Diagram of distance-velocity relation of galaxies used by Lemaître (1927; 1931) to yield the first empirical value of the expansion rate of the Universe: 625 km s^{-1} Mpc^{-1} and 575 km s^{-1} Mpc^{-1}. Solid black line has a slope of 575 km s^{-1} Mpc^{-1} (Block 2012). (b) Similar diagram published in Hubble (1929) founded the best slope of 530 km s^{-1} Mpc^{-1}.

In cosmology, they are generally expressed in h^{-1}Mpc

$$h = \frac{H_0}{100 \text{ km s}^{-1} \text{ Mpc}^{-1}},$$

where h is the reduced Hubble parameter.

1.1.2.3 FLRW metric

Under these three assumptions, our Universe should be described by a metric that represents a homogeneous, isotropic and expanding spacetime. Let's find the corresponding metric for the usual coordinate system $x^\mu = (t, \mathbf{r})$:

$$\mathrm{d}s^2 = \underbrace{g_{\mu,\nu}(t, \mathbf{x})}_{=g_{\mu\nu}(t)} \mathrm{d}x^\mu \mathrm{d}x^\nu \tag{1.15}$$

$$= g_{00}(t)\mathrm{d}t^2 + 2\underbrace{g_{0i}(t)}_{=0}\mathrm{d}t\mathrm{d}x^i + \underbrace{g_{ij}(t)}_{=a(t)^2 g_{ij}}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.16}$$

where the homogeneity is used in (1.15) and the isotropy in (1.16)[5]. The expansion is driven by the term $a(t)$. The final spatial component of the metric, $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$, represents a 3-dimensional space with maximal symmetry, due to the assumption of isotropy. This corresponds in a flat space, to the classic spherical metric; ergo leading to the famous *Friedmann-Lemaître-Robertson-Walker (FLRW)* metric with spherical coordinates (r, θ, ϕ):

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a(t)^2\left(\frac{\mathrm{d}r^2}{1 - Kr^2} + r^2\mathrm{d}\Omega^2\right), \tag{1.17}$$

where $a(t)$ is known as the *scale factor*, $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2$ is the solid angle element with $\theta, \phi \in [0, 2\pi], [-\pi/2, \pi/2]$ and K is the curvature term. If $K = 0$ (*resp.*, $K > 0$, $K < 0$), the Universe is said flat (*resp.*, closed, open). We choose K to be in units of length^{-2} and, then, r

[5]Note that g_{00} and also g_{0i} cannot depend on time to be able to synchronize clocks (Landau and Lifshitz 1975).

has to be in units in length and $a(t)$ is unitless. For convenience, $a(t)$ is chosen to be 1 today (t_0).

For computational reasons, it is generally convenient to write the FLRW metric using the conformal time η and the comoving coordinate χ:

$$\mathrm{d}s^2 = a(\eta)^2 \left[-c^2\mathrm{d}\eta^2 + \left(\mathrm{d}\chi^2 + S_K^2(\chi)\mathrm{d}\Omega^2\right)\right], \tag{1.18}$$

where the conformal time is defined as $\mathrm{d}t = a\mathrm{d}\eta$ and $S_K(\chi)$ depends on the value of the curvature term K:

$$S_K(\chi) = \begin{cases} K^{-1/2}\sin\left(\sqrt{K}\chi\right) & K > 0 \\ \chi & K = 0 \\ (-K)^{1/2}\sinh\left(\sqrt{-K}\chi\right) & K < 0. \end{cases} . \tag{1.19}$$

The position of an object can be written as $\mathbf{r} = a\mathbf{x}$ where $\mathbf{x}$ are the comoving coordinates of the object. In these coordinates, the object is not moving due to the expansion of the Universe. The velocity reads as

$$\mathbf{v} = \frac{\mathrm{d}a}{\mathrm{d}t}\mathbf{x} + a\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = H\mathbf{r} + a\mathbf{v}_{\mathrm{pec}}, \tag{1.20}$$

where H is the Hubble parameter

$$H = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}. \tag{1.21}$$

The velocity has two components: the Hubble flow from the expansion of the Universe and the peculiar velocity $\mathbf{v}_{\mathrm{pec}}$ due to the proper motion of the objects inside the Hubble flow. By construction, $\langle\mathbf{v}_{\mathrm{pec}}\rangle = 0$ and $\langle\mathbf{v}\rangle = H\mathbf{r}$. For nearby galaxies, typically those observed by Lemaître and Hubble, the Hubble parameter is constant, finding the Hubble-Lemaître law (1.14).

1.1.2.4 Redshift and distance measurements

Measuring the distance of extremely distant objects is a challenging task in astronomy. One of the most reliable methods is to measure the *redshift* from an object through its spectral analysis.

Suppose a galaxy emits a photon at time t_e and received it on Earth at time t_0. Photons respect, by definition, $\mathrm{d}s^2 = 0$. Then, from (1.18), the radial comoving distance of this galaxy is

$$\chi = \int_{t_e}^{t_0} \frac{c\mathrm{d}t}{a(t)}. \tag{1.22}$$

If a second photon is emitted shortly after the first one at time $t_e + \delta t_e$ and received at time $t_0 + \delta t_0$, the radial comoving distance of this galaxy will be still the same. Hence,

$$\frac{\delta t_e}{a(t_e)} = \frac{\delta t_0}{a(t_0)}, \tag{1.23}$$

since $a(t)$ can be supposed constant between t_e and $t_e+\delta t_e$ and also between t_0 and $t_0+\delta t_0$. Then, the wavelength of photons $\lambda = c\delta t$ propagating through the expanding Universe is stretched, creating the cosmological redshift:

$$\frac{\lambda_e}{\lambda_0} = \frac{1}{a(t_e)} \equiv 1 + z, \tag{1.24}$$

where $a(t_0) = 1$ by choice and z is the redshift of the observed galaxy. In addition to the expansion of the Universe, two other effects cause redshift

- Doppler redshift: similar to what happens on earth, the peculiar velocity $\mathbf{v}_{\rm pec}$ of galaxies along the line of sight creates a Doppler effect

$$1 + z_{\rm pec} = \sqrt{\frac{1 + v_{\rm pec}/c}{1 - v_{\rm pec}/c}} \approx 1 + \frac{v_{\rm pec}}{c}.$$

- Gravitational redshift: the energy of light decreases as it travels through a gravitational field. It has been observed firstly with the Pound-Rebka experiment. This contribution is generally negligible for cosmological scales except for the Sachs–Wolfe effect. This contribution is

$$1 + z_{\rm grav} = \left(1 - \frac{2GM}{rc^2}\right)^{-\frac{1}{2}},$$

where M is the mass of the object creating the gravitational field and r is the distance to the source.

These three redshift contributions can be combined as $1 + z_{\rm glob} = (1+z)(1+z_{\rm pec})(1+z_{\rm grav})$. In the following and for the considered scales, the gravitational and the Doppler redshift will be neglected.

Redshift is the sole method for measuring the distance of extremely distant objects and is crucial for any cosmological measurement. The goal of a galaxy survey that will be discussed in this dissertation, is to observe as accurately as possible the redshifts of millions of galaxies.

Several distances and geometrical elements are used in cosmology, some of them are displayed for the Planck 2018 parameters in Fig. 1.3. For a complete description refer to Hogg (1999).

The look-back time gives the time needed for a photon to travel a certain distance/redshift :

$$t_{\rm L} = \int_0^z \frac{dz'}{(1+z')\,H\,(z')} \tag{1.25}$$

In a flat Universe, the comoving distance is equal to the radial comoving distance expressed by (1.22). Noting that $\mathrm{d}t = \mathrm{d}a/(aH) = -a\mathrm{d}z/H$, the comoving distance becomes (with $K = 0$):

$$D_c(z) = \chi(z) = \int_0^z \frac{c\mathrm{d}z'}{H(z')}. \tag{1.26}$$

The comoving volume V_c is defined as the volume in which the number density of objects locked into Hubble flow is constant with the redshift; the comoving volume element in solid angle $d\Omega$ and redshift interval dz is (with $K = 0$)

$$\mathrm{d}V_{\rm C} = \frac{c\chi(z)^2}{H(z)}\mathrm{d}\Omega\mathrm{d}z. \tag{1.27}$$

1.1.2.5 Friedmann equations

Knowing the amount of matter and energy present, Einstein's equations (1.8) will describe the evolution and the dynamics of the Universe. Under the FLRW metric (1.18), Einstein's equations can be reduced to two independent equations, called the *Friedmann equations*:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3} \\ \frac{\ddot{a}}{a} &= -\frac{8\pi G}{6}(\rho + 3P) + \frac{\Lambda}{3} \end{aligned}, \tag{1.28}$$

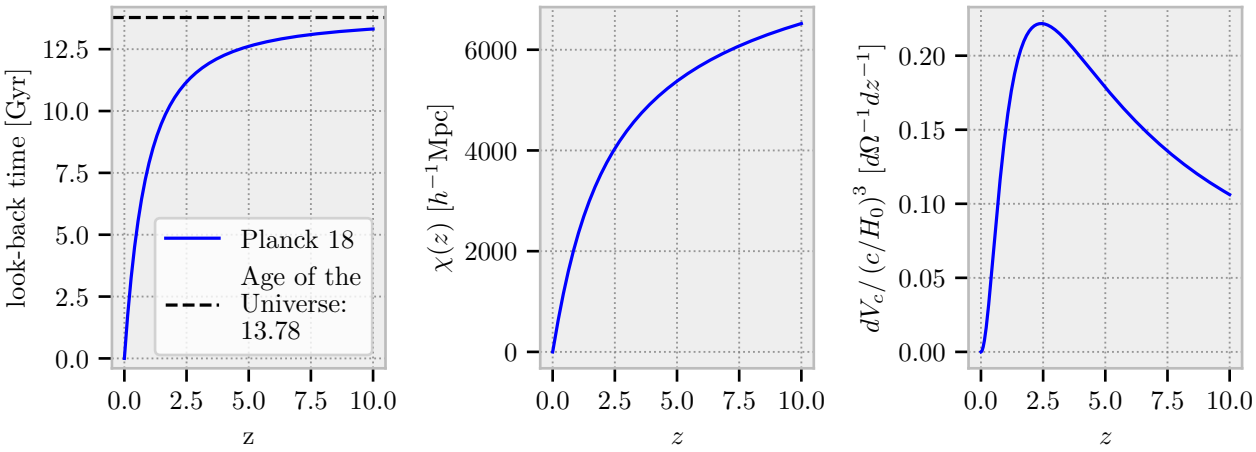


Figure 1.3: Look-back time, comoving distance and dimensionless comoving volume element as a function of the redshift for the fiducial Planck 2018 parameters (Planck Collaboration et al. 2020).

where the first (*resp.* second) one is obtained from the equation 00 (*resp.* *ii*) of (1.8).

The conservation equation (1.12) becomes

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + 3H(\rho + P) = 0. \tag{1.29}$$

Remark: The conservation equation is automatically verified with the Bianchi identity (1.13). Hence, (1.29) can be recovered simply by combining the two Friedmann equations (1.28).

1.1.3 Hot Big Bang model

1.1.3.1 The Universe dynamics

The Friedman equations (1.28) are used to describe the expansion of the Universe based on the amount and type of energy in it. However, these two independent equations have three unknowns and an additional equation is needed to solve them. The energy will be described by an equation of state of the following form

$$P = w\rho, \tag{1.30}$$

where w will be generally supposed constant. The energy and matter that fills the Universe can be classified into different species: the pressure-less matter, non-relativistic, with $w = 0$ which can be separated into baryonic matter (b) which interacts electromagnetically and dark matter (cdm) which interacts not or very little electromagnetically – the radiation, relativistic, with $w = 1/3$ which can be separated into photons (γ) and neutrinos (ν) – the curvature contribution that can be described by a fluid such as $w = -1/3$, see K term in (1.28), and will be neglected hereafter – the dark energy described by Λ which is assimilated to a fluid such as $w = -1$.

The evolution of the density of each species is given by injecting (1.30) in (1.29):

$$\rho \propto a^{-3(1+w)}, \tag{1.31}$$

then the scale factor can be obtained via the Friedmann equations (1.28). The behavior of density and scale factor for the different species, when they are considered as predominant, is given in Table 1.1.

Table 1.1: Summary of the behavior of density ρ, scale factor a and conformal time η as a function of time for the usual species, with an equation of state w, filling the Universe.

type	w	ρ	a	η
cosmological constant	-1	$\propto a^0$	$\propto e^{Ht}$	$\propto e^{-Ht}$
pressure-less matter	0	$\propto a^{-3}$	$\propto t^{2/3}$	$\propto t^{1/3}$
relativistic species	1/3	$\propto a^{-4}$	$\propto t^{1/2}$	$\propto t^{1/2}$

It is useful to define the critical density parameter

$$\rho_c = \frac{3H^2}{8\pi G}, \tag{1.32}$$

and the density parameter for each species as $\Omega_X = \rho_X/\rho_c$, defining for the curvature and dark energy components the following density $\rho_\Lambda = \Lambda/(8\pi G)$ and $\rho_K = -3K/(8\pi G a^2)$. Thus, the first Friedman equation (1.28) becomes

$$\sum_X \Omega_X + \Omega_\Lambda + \Omega_K = 1, \tag{1.33}$$

where typically $\sum_X \Omega_X = \Omega_m + \Omega_r = \Omega_b + \Omega_{cdm} + \Omega_\gamma + \Omega_\nu$. Since the density evolution for each species is known from (1.31), the above equation can be expressed as a function of the density parameters $\Omega_{X,0} = \Omega_X(z=0)$ and the redshift z, given the evolution of the Hubble parameter as

$$H^2(z) = H_0^2 \left[\sum_X \Omega_{X,0}(1+z)^{3(1+w_X)} + \Omega_{K,0}(1+z)^2 + \Omega_{\Lambda,0} \right]. \tag{1.34}$$

This expression of $H(z)$ is crucial since it enables us to compute analytically the distances and volume presented in Fig. 1.3 supposing a distribution of the different species today. The critical density parameter $\rho_{c,0}$ depends on the value of the Hubble constant today H_0. It is therefore natural to consider the reduced density parameters $\omega_{X,0} = \Omega_{X,0}h^2$ which is then independent of H_0.

The quantities Ω_X and ω_X can be expressed explicitly as a function of their value today ($z=0$) and z. Therefore, in the following and as usual, $\Omega_X = \Omega_{X,0}$ and $\omega_X = \omega_{X,0}$.

As shown in Table 1.1, the density of the species that fill the Universe evolves differently according to the scale factor a or to the redshift z, revealing phases of domination of one species over the others. The evolution of the density parameters as a function of z is shown in Fig. 1.4. Ω_X are chosen to be those measured in the latest Planck data release (Planck Collaboration et al. 2020) and will be the fiducial value for these parameters in this dissertation. The history of the Universe begins with a radiation-dominated area, then turns into a matter-dominated area, that can be well approximated by an Einstein-de Sitter Universe. The most recent part of the Universe's history is an accelerated expanding Universe where dark energy dominates. The moment when the radiation density is equal to the matter density is known as matter-radiation equality, and it occurs at a redshift of $z_{eq} \sim 3387$.

This description is only valid for fluids that interact only through gravity. Otherwise, one must take into account the coupling terms, see for example Section 2. Since the only non-gravity interaction is Compton scattering between photons and baryons, and since baryons represent only a very small fraction of matter, the above description is valid at least to first order.

According to the Planck 2018 measurement (Planck Collaboration et al. 2020) and assuming the ΛCDM model, the Universe is today made of $\sim 8 \cdot 10^{-3}\%$ of radiation, $\sim 31\%$ of matter

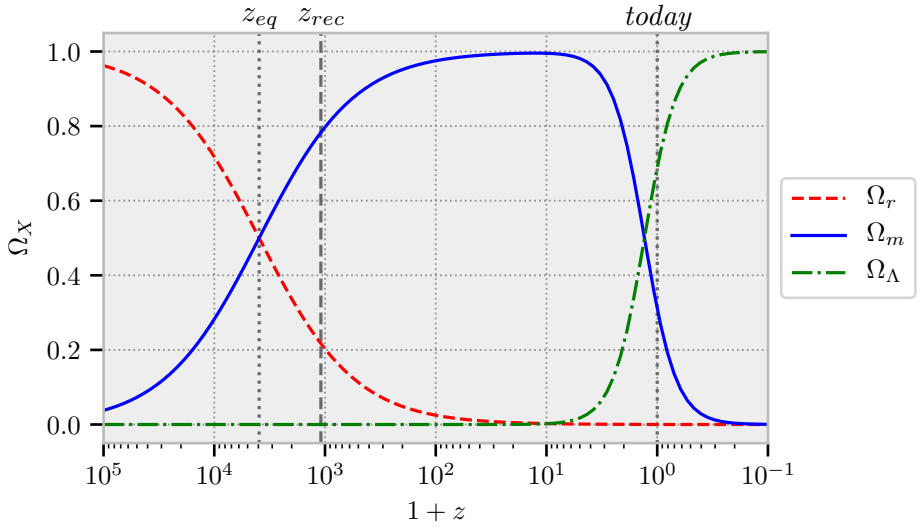


Figure 1.4: Evolution of the density parameters (radiation in red, matter in blue, dark energy in green) in a flat ΛCDM model for the fiducial Planck 2018 parameters (Planck Collaboration et al. 2020) ($\Omega_m = 0.311$, $\Omega_\Lambda = 0.689$, $\Omega_r = 7.91 \cdot 10^{-5}$, $H_0 = 67.66$). z_{eq} is the redshift of the radiation-matter equality *i.e.* when the radiation density is equal to the matter density. z_{rec} is the redshift of the recombination mentioned in 1.1.3.2. Today, the Universe is dominated by the dark energy component which appears very late in its history, well after the formation of all its large-scale structures.

(including $\sim$ 4.5% of baryonic matter) and $\sim$ 69% of dark energy. Supported by the same measurement, the Universe will be assumed flat in the following, at least for the scales observed in the various cosmological experiments: $K = 0$.

1.1.3.2 A brief history

For much of the 20th century, the idea that the Universe was expanding and therefore denser and hotter in the past was controversial. In fact, it was even mockingly referred to as the "Big Bang" by Fred Hoyle, who believed in a stationary Universe. However, the discovery of cosmic microwave background (CMB) radiation by Penzias and Wilson in 1965 confirmed a prediction made by Gamow, Alpher, and Herman in 1948.

As the Universe expanded and cooled, protons and electrons combined to form hydrogen preventing the photons present at that time to interact by Thomson scattering. Photons were then released into the Universe, creating cosmic microwave background radiation. This radiation can only exist in an expanding Universe in which the temperature decreases over time.

The timeline in Fig. 1.5 summarizes the history of the Universe according to the "Hot Big Bang" model. It can be quickly described as follows:

- The initial phase between the 'instant zero' and the real beginning of the "Hot Big Bang" model corresponds to the early Universe and is currently explained by cosmic inflation and will be described in Chapter 2. Note that before the Planck epoch ($t_{\rm pl} \sim 10^{-43}$ s) the actual theory of gravitation cannot be applied since the 4 fundamental forces should be considered in a uniform framework.

- At the very beginning ($t \sim 10^{-30}$ s), the Universe consisted of an extremely hot and dense plasma, known as the "primordial soup", made of the most fundamental particles,

as quarks and gluons. As the Universe expanded and cooled, the temperature dropped to a point where the first nuclei could form, limiting the existence of the most energetic particles.

- Then, the plasma was made almost exclusively of the lowest energetic particles as protons, electrons, neutrinos and photons. During a short period ($t \sim 10/10^3$ s), the plasma was hot enough to allow the creation of heavier nuclei. This is the Big Bang nucleosynthesis (BBN) which is responsible for the formation of most of the helium and deuterium in the Universe. The quantity observed of these two components in the galaxies (Aver et al. 2015, Cooke et al. 2018) can be only explained by the BBN and is also in favor of the "Hot Big Bang" model.
- As the universe continued to expand and cool, protons and electrons combined. This is the recombination ($t \sim 380\ 000$ yr). Photons decoupled from matter and propagated freely through the Universe.
- After the recombination, the Universe was not dense enough to emit new light. This is called the Dark Ages. Under the gravity effect, the matter fell into the over-dense regions creating the first stars and galaxies. These first structures emitted new first lights which are responsible for the reionization ($t \sim 250/900$ Myr) of the Universe (Bromm and Yoshida 2011, Wise et al. 2012).
- Then, galaxies continued to be formed under the competition between gravity and the expansion of the Universe creating the cosmic web as it is observed today with a filamentary structure and clusters.

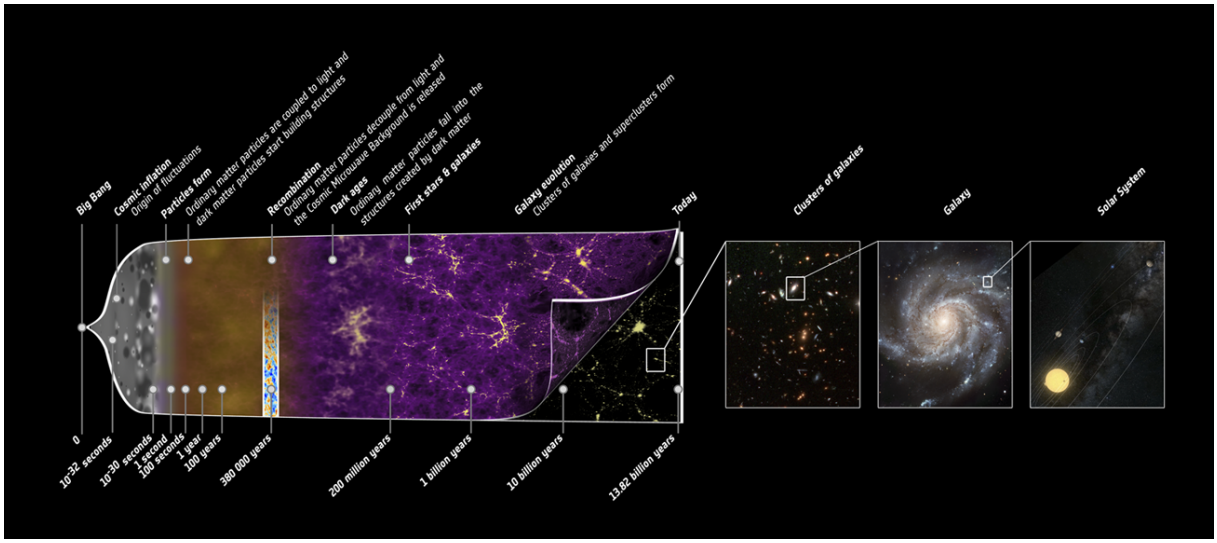


Figure 1.5: Cosmological timeline of the Hot Big Bang model which summarizes the almost 14-billion-year-long history of our universe. Starting from the initial phase of the cosmos where the properties of the Universe were almost uniform and punctuated only by tiny fluctuations up to the well-formed and clustered Universe that is observed today. Credits: ESA – C. Carreau

1.1.3.3 The Hot Big Bang problems

Although the "Hot Big Bang" model has received significant support from the discovery of cosmic microwave background radiation and the observed abundances of light elements produced during primordial nucleosynthesis, it still faces challenges that cannot be resolved without additional theories. For instance, the following problems should be explained:

► **Flatness problem** The curvature evolution can be written from (1.28) as

$$\frac{\mathrm{d}\Omega_K}{\mathrm{d}\ln a} = (3w+1)\,(1-\Omega_K)\,\Omega_K, \tag{1.35}$$

which can be integrated as a function of the redshift

$$\Omega_K = \frac{\Omega_{K0}}{(1-\Omega_{K0})\,(1+z)^{3w+1} + \Omega_{K0}} \tag{1.36}$$

The current observations are in favor of a flat Universe, typically $|\Omega_{K0}| = |\Omega_0 - 1| < 0.1$. Neglecting the dark energy contribution, this leads to radiation-matter equality ($w = 0$) to

$$|\Omega\,(z_{\mathrm{eq}}\,) - 1| < 3 \times 10^{-5},$$

and then ($w = 1/3$) to a very small flatness at the Planck time

$$|\Omega\,(z_{pl}) - 1| < 10^{-60}.$$

Therefore, the theory should be able to either explain why $\Omega_K = 0$ with high precision or identify a mechanism that can produce curvature of such a small magnitude.

► **Horizon problem** The observed temperature of the CMB is prodigiously isotropic, where the relative fluctuation is about 10^{-5}, see Fig. 1.1b. This leads us to the fundamental assumption of the cosmological principle.

To explain this isotropy simply, the regions that emit CMB photons at z_{rec} should be causally related, i.e. the light could have traveled between them before z_{rec}. Let's compute the *particle horizon* at z_{rec} which gives the maximum distance that light particles could have traveled to an observer from the beginning of the Universe. It is exactly the comoving distance at z_{rec}:

$$\chi(z_{rec}) = \int_{z_{rec}}^{\infty} \frac{\mathrm{d}z}{H(z)} \approx \frac{2}{\sqrt{\Omega_m} H_0} \frac{1}{\sqrt{1+z_{rec}}}. \tag{1.37}$$

Since the recombination occurs during the matter-dominated area, $H^2(z)$ is approximated to $\Omega_m H_0^2 (1+z)^3$ from (1.34). The corresponding angular distance is $d_A(z_{rec}) = a(z_{rec})\chi(z_{rec}) \approx 2^\circ$.

Finally, only patches with a diameter of about 2°, *i.e.* with a surface of about 0.0012 sr, on the celestial sphere could be causally connected at the time of recombination. Hence, the CMB should consist of $4\pi/0.0012 \approx 10^4$ causally independent patches which should have all the same temperature. An elegant solution will be to provide a mechanism that increases the particle horizon before the recombination.

► **Monopole problem** In the standard picture of the Hot Big Bang model, the temperature gets bigger as the Universe gets younger. Hence, in the very early Universe *i.e.* before the *Grand Unification scale* ($T \sim 10^{16}$ GeV, $t \sim 10^{-38}$ s), the electromagnetic force, weak force and strong force become equal in strength and should be described by a common theory known as the *Grand Unification Theory* (GUT) (De Boer 1994)

Unfortunately, many of these theories predict the creation of "*topological defects*" resulting from the phase transition due to symmetry breaking when the Universe cools as it expands. In particular, most GUT theories predict the creation of magnetic monopoles (see Zeldovich and Khlopov 1978, Preskill 1979, Guth and Tye 1980) already hypothesized by Maxwell or Dirac. In the case of GUT transition, magnetic monopoles should be the dominant matter in the Universe. However, no monopole has ever been observed, directly or indirectly, by humans. Monopoles are called *unwanted relics* since they are produced by the theory but not observed.

▶ **Inflation, a life-saving solution** These first two problems can be solved, with a great experimental pertinence, by adding an additional period of accelerated expansion at the beginning of the Universe's history: *inflation*. The monopole problem was one of the early motivations for inflationary physics since inflation allows us to control the energy scales at which particles are created in the Universe, and choosing an energy scale lower than the GUT transition solves this problem. This description of the early Universe was pioneered by Guth (1981), Albrecht and Steinhardt (1982), Linde (1982a).

The mechanisms that can create inflation will be the main topic of Chapter 2. Note that inflation will provide, naturally, an explanation for the size and shape of the primordial fluctuations. Today, inflation and the hot Big Bang model constitute the standard cosmological model although the notion of a standard model can be quite variable over time.

This thesis aims to bring some, unfortunately indirect, observations to this mechanism via the observation of the galaxy distribution.

1.2 Large-scale structures of the Universe

As described in the previous section and according to current observations, the Universe is very well described by a homogeneous and isotropic model. However, the discovery of galaxies in the early 20th century demonstrated that the universe is not perfectly homogeneous at small scales. The goal of this section is to describe the evolution of the matter perturbations that will lead to the appearance of galaxies and the large-scale structures of the Universe.

1.2.1 Statistical description of cosmological fields

Let's start by defining and describing the cosmological fields, along with providing a statistical characterization of them.

1.2.1.1 Contrast density field and Ergodicity

To investigate perturbations in a homogeneous and isotropic Universe, the main field used will be the contrast density field δ of any statistical field such as the matter density or the temperature. In the following, only real fields will be considered. For instance, the matter contrast density is defined at any spatial position $\mathbf{x}$ as

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}, \tag{1.38}$$

where $\rho(\mathbf{x})$ is the matter density and $\bar{\rho}$ the spatial average of the matter density. By construction, see hereafter, the first moment of the field δ is zero:

$$\langle \delta(\mathbf{x}) \rangle = 0. \tag{1.39}$$

In probability theory, the expected value of a random variable X, denoted by $\langle X \rangle$, is the generalization of the classic weighted average for a random variable. The law of large numbers enables the estimation of the expected value as the sample average of a large set of independent realizations of X. It is generally called the *ensemble average*. In physics, this average can be computed with a large set of independent simulations.

However, when observing the Universe, only one realization of the cosmological fields can be observed. In such cases, it is usually assumed that fields satisfy the property of *ergodicity*

which set the equality between the ensemble average and the volume average when the volume is sufficiently large. Although there is no formal proof of this assumption in cosmology, it is regarded as a commonsense axiom (Peacock 1998).

Thus, the property of ergodicity and the cosmological principle ensures that the ensemble average does not depend on the position, and it is equal to the mean value of the density observed in the Universe:

$$\langle \rho(\mathbf{x}) \rangle = \langle \rho \rangle_{\mathbf{x}} = \bar{\rho}.$$

1.2.1.2 2-point correlation function & power spectrum

The contrast density field δ will be statistically described by using its statistical moments. Fig. 1.6 shows two different fields. Both of them have the same expected value (first-order moment). Both of them have the same expected value (first-order moment). The one on the left is completely random and has no over-density. On the right, the field is called 'clustered', characterized by its second moment, the *two-point correlation function*.

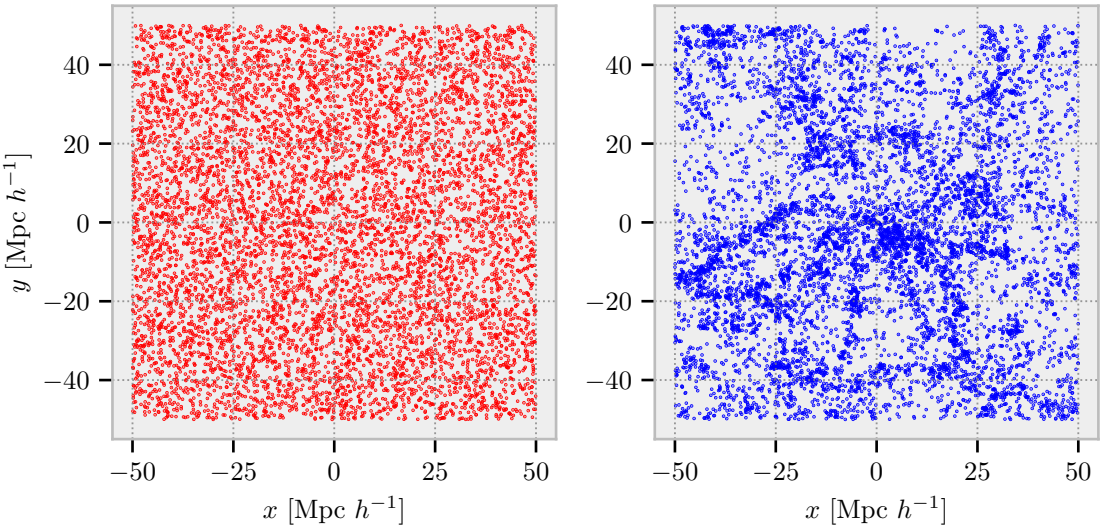


Figure 1.6: Left: Randomly distributed field with an associated power spectrum of zero. Right: Field generated with the Planck 2018 power spectrum. The points are clustered and not randomly distributed. The power spectra and the correlation functions of these two fields are shown in Fig. 1.7.

The 2-point correlation function, also called *autocorrelation* function, is defined as the ensemble average of the field taken in two positions:

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle.$$

In the case of a field with a null expected value, it is the covariance between the field $\delta(\mathbf{x})$ and the field $\delta(\mathbf{x}+\mathbf{r})$, which depends on the position $\mathbf{x}$. However, the statistical homogeneity of the Universe ensures that ξ depends only on $\mathbf{r}$. In addition, the correlation function depends only on $r = \|\mathbf{r}\|$ since the Universe is also assumed isotropic.

As often, it is useful to go in Fourier space:

$$\tilde{\delta}(\mathbf{k}) = \int \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3\mathbf{x} \quad \text{and} \quad \delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \tilde{\delta}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3\mathbf{k}. \tag{1.40}$$

Note that since the field δ is real, $\tilde{\delta}(\mathbf{k}) = \tilde{\delta}^*(-\mathbf{k})$. The *power spectrum* is defined as the Fourier transform of the 2-point correlation function:

$$P(\mathbf{k}) = \int \xi(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} \quad \text{and} \quad \xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int P(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{k}. \tag{1.41}$$

The statistical properties of the field δ will be completely described by the field $\tilde{\delta}$. For instance, the 2-point correlation function ξ can be expressed in function of $\tilde{\delta}$:

$$\begin{aligned}
\langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle &= \int \int \langle (\delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle e^{-i\mathbf{k}\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot(\mathbf{x}+\mathbf{r})} \mathrm{d}^3\mathbf{x} \mathrm{d}^3\mathbf{r} \\
&= \int \xi(\mathbf{r}) \left(\int e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{x}} \mathrm{d}^3\mathbf{x} \right) e^{-i\mathbf{k}'\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} \\
&= (2\pi)^3 \delta_D(\mathbf{k}+\mathbf{k}') \int \xi(\mathbf{r}) e^{-i\mathbf{k}'\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r},
\end{aligned}$$

where δ_D is the Dirac distribution. It appears via the Fourier transform of the identity distribution:

$$(2\pi)^3 \delta_D(\mathbf{k}) = \int e^{-i(\mathbf{k}\cdot\mathbf{x})} \mathrm{d}^3\mathbf{x}, \tag{1.42}$$

which will be very useful in the following. Since ξ depends only on r, P will also depend only on k. Thus,

$$\langle \tilde{\delta}(\mathbf{k}) \tilde{\delta}(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k}+\mathbf{k}') P(k), \tag{1.43}$$

and $\xi(r)$ can be only expressed in function of k. The spherical coordinates (k, θ_k, ϕ_k) of $\mathbf{k}$ are associated to the sphere where the z axis is defined along the direction of $\mathbf{x}$:

$$\begin{aligned}
\xi(r) &= \frac{1}{(2\pi)^3} \int P(k) e^{i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{k} \\
&= \frac{1}{(2\pi)^3} \int P(k) \left(\int e^{irk\cos(\theta)} \cos(\theta_k) \mathrm{d}\theta_k \right) \left(\int \mathrm{d}\phi_k \right) k^2 \mathrm{d}k \\
&= \frac{4\pi}{(2\pi)^3} \int P(k) \frac{\sin(kr)}{kr} k^2 \mathrm{d}k.
\end{aligned}$$

Since the 2-point correlation function depends only on r (isotropy), an equivalent definition of the power spectrum (1.41) is generally given, changing our usual definition of the Fourier transform:

$$P(k) = \int \xi(r) e^{i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r}, \quad \text{and} \quad \xi(r) = \frac{1}{(2\pi)^3} \int P(k) e^{-i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{k}.$$

The 2-point correlation function and associated power spectrum represent the second-order summary statistics. They describe at what scales the particles are clustered together. Fig. 1.6 shows two different fields with a similar first-order summary statistic: $\langle \delta(x,y) \rangle = 0$ but with two different second-order summary statistics. They are displayed in Fig. 1.7. The field on the left has no particular clustering and has a zero power spectrum, while on the right the field has strong over/under dense regions and the associated power spectrum is non-zero. This field was generated with the matter power spectrum measured from Planck data, mimicking in two dimensions the way galaxies are placed in the Universe.

Remark: Instead of computing the autocorrelation of a field δ, it can be useful sometimes to compute the *cross-correlation* between two fields δ_1, δ_2. This is the natural extension of what is defined in this section: $\xi_{12}(\mathbf{r}) = \langle \delta_1(\mathbf{x}) \delta_2(\mathbf{x}+\mathbf{r}) \rangle$.

The purpose of galaxy surveys, and in particular the DESI survey (see Section 3.2) which will be discussed in this dissertation, is to map the Universe and characterize the x-order summary

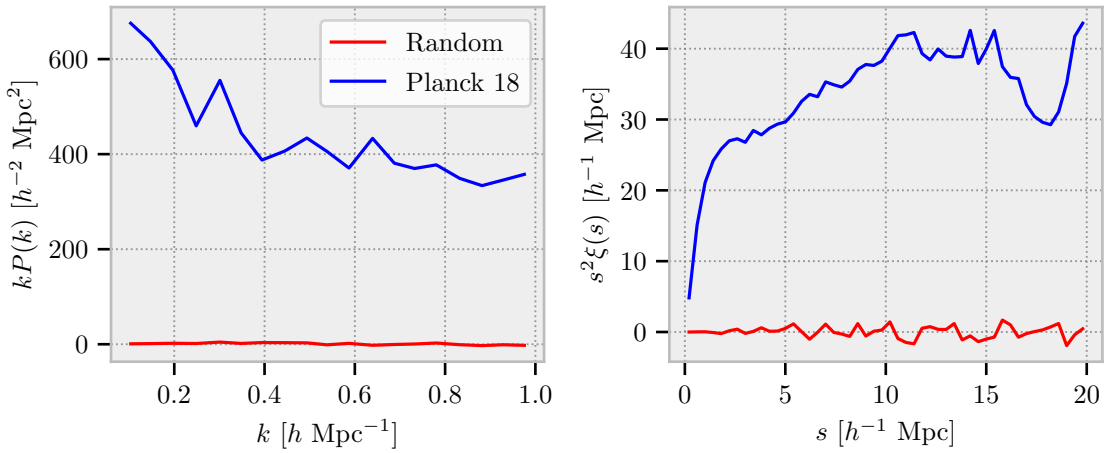


Figure 1.7: The associated power spectra (left) and the correlation functions (right) of the two fields display in Fig. 1.6. As expected, the random field exhibits no correlation.

statistics of the matter distribution. However, mapping the Universe in 3 dimensions is a challenging and costly task, as it requires observing individually the spectrum of billions of galaxies to determine their redshifts and hence their distances from us. It is therefore sometimes more practical, and less expensive, to look at the angular distributions and their associated summary statistics.

The angular 2-point correlation function $w(\theta)$ is then defined as $w(\theta) = \langle \delta(\phi)\delta(\phi+\theta) \rangle$ where θ is the angular separation between two positions on the sphere (for an explicit derivation, see for instance Kurki-Suonio 2019). It is related to $\xi(r)$ through:

$$w(\theta) = \int_0^\infty \int_0^\infty \xi(r_{12}) \tilde{S}(r_1) r_1^2 \mathrm{d}r_1 \tilde{S}(r_2) r_2^2 \mathrm{d}r_2, \tag{1.44}$$

where $r_{12} = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta)}$ and $\tilde{S}(r)$ represents the probability for a galaxy at distance r to be observed by the survey. It is defined as

$$\tilde{S}(r) = \frac{S(r)}{\int_0^\infty S(r) r^2 \mathrm{d}r}, \tag{1.45}$$

where $S(r)$ is the survey selection function giving the density of galaxies observed by the survey at distance r, $S(r) = \mathrm{d}N(r)/\mathrm{d}r$ with N the area density, and it is assumed here that it does not depend on the direction of the sphere.

The famous *Limber approximation* (Limber 1954) states that the correlation function is non-negligible only for small values of r *i.e.* that the angular correlation function is non-zero only for small values of θ, one can assume that $x = r_2 - r_1$ and $y = (r_1 + r_2)/2$, and hence derive *Limber's equation* valid only for small angles:

$$w(\theta) \simeq \int_0^\infty \tilde{S}(y)^2 y^4 \int_{-\infty}^\infty \xi\left(\sqrt{x^2 + y^2\theta^2}\right) \mathrm{d}x\mathrm{d}y. \tag{1.46}$$

Assuming then a power law for the 3D correlation function such that $\xi(r) = (r/r_0)^{-\gamma}$, the angular correlation function is reduced to

$$w(\theta) = \sqrt{\pi} \frac{\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} r_0^\gamma \theta^{1-\gamma} \int_0^\infty \tilde{S}^2(y) y^{5-\gamma} \mathrm{d}y \tag{1.47}$$

Similarly, the angular power spectrum can be derived from the 3D power spectrum

$$C_l = \frac{2}{\pi}\int_0^\infty \left[\int_0^\infty j_l(kr)\tilde{S}(r)r^2\mathrm{d}r\right]^2 P(k)k^2\mathrm{d}k, \tag{1.48}$$

where j_l is the spherical Bessel function. The angular power spectrum C_l is the main observable in the CMB experiment which probes the last scattering surface.

These two angular statistics are shown in Fig. 1.8. The power spectrum is the power spectrum from the Planck data and the selection function $S(z)$ is a Gaussian centered in 1.7 with a dispersion of 0.8. This selection function mimics the redshift distribution of observed quasars in DESI, see Section 4.2.4.5. In particular, one can clearly identify in the angular and 3d 2-point correlation functions, the peak generated by the baryonic acoustic oscillation, see Section 1.2.3.4.

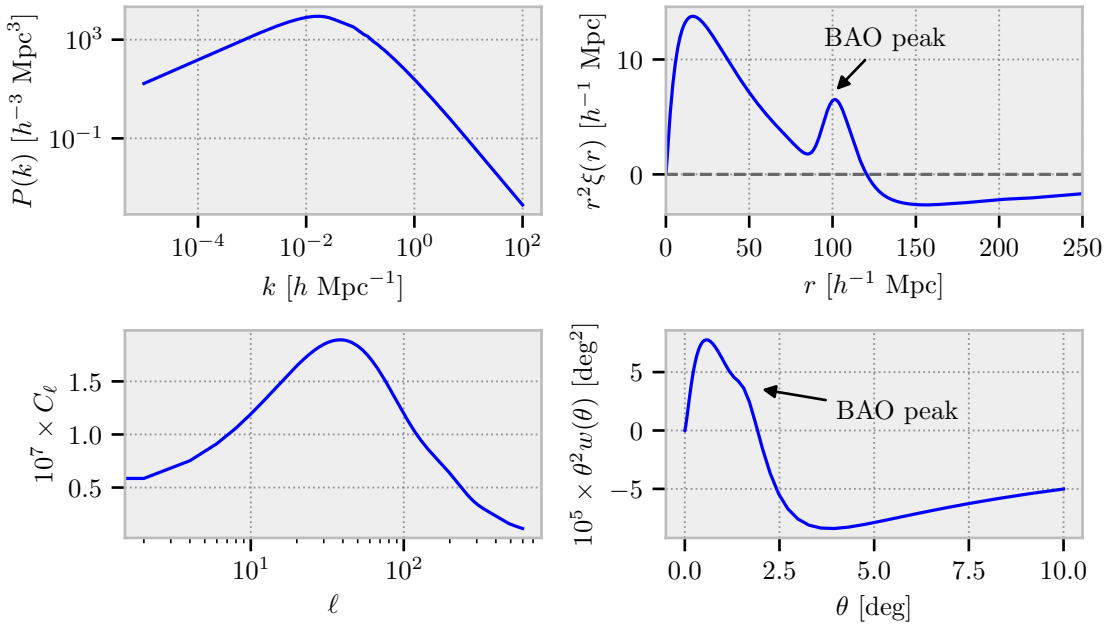


Figure 1.8: On top, power spectrum and 2-point correlation function from Planck data. On the bottom, *angular* power spectrum and correlation function. The selection function used to compute them is a Gaussian centered in z=1.7 with a dispersion of 0.8, mimicking the quasar redshift distribution observed with DESI, see Section 4.2.4.5. The BAO peak, see Section 1.2.3.4, can be identified in the 2-point correlation functions.

1.2.1.3 High order statistics and Wicks theorem

Like the variance, the 2-point correlation function does not allow the characterization of an entire cosmological field, fully described by the knowledge of all the *N-point correlation functions*. These high-order statistics are defined as

$$\xi^{(N)}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) \equiv \left\langle \delta\left(\mathbf{x}_1\right)\delta\left(\mathbf{x}_2\right)\ldots\delta\left(\mathbf{x}_N\right)\right\rangle. \tag{1.49}$$

Using $\mathbf{x}_N$ as the reference point and defining $\mathbf{x}_i = \mathbf{x}_N + \mathbf{r}_i$, the spatial homogeneity ensures that $\xi^{(N)}$ depends only on the separations between the points. Since $\xi^{(N)}$ is invariant under permutation of $\mathbf{x}_i$, $\mathbf{r}_i$ can be reordered by length. Note that the choice of reference point is free and other symmetries exit. Finally,

$$\xi^{(N)}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) = \xi^{(N)}\left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N-1}\right), \text{ with } |\mathbf{r}_1| \leq \ldots \leq |\mathbf{r}_{N-1}|. \tag{1.50}$$

It is more usual to work with the *connected N-point correlation functions* which are the remaining part of the N-point correlation functions minus all the contributions from the lower-order correlations:

$$\xi_C^{(N)}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) \equiv \xi^{(N)}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) - \sum_n \xi_C^{(N_1)} \ldots \xi_C^{(N_n)}, \tag{1.51}$$

where the sum is over all partitions of $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ into smaller subsets. Note that $\xi_C^1 = \langle\delta\rangle = 0$, $\xi_C^{(2,3)} = \xi^{(2,3)}$ and $\xi_C^{(n)} \neq \xi^{(n)}$ for $n \geq 4$.

If δ is assumed to be gaussian, then the Wick theorem can be applied to the N-point correlation function:

$$\xi^{(N)}\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) = \sum_n \xi_C^{(N_1)} \ldots \xi_C^{(N_n)}, \tag{1.52}$$

such that for Gaussian density perturbations, summary statistics are fully determined by $\xi(r)$. In particular, in the case of Gaussian fields, $\xi^{(N)} = 0$ when N is odd. Hence, the connected higher-order statistics measure the non-gaussianity.

Specifically, the 3-point correlation function $\xi^3(\mathbf{r}_1, \mathbf{r}_2)$, written as $\zeta(\mathbf{r}_1, \mathbf{r}_2)$ is a statistic of great interest to probe the non-gaussianity in the field δ. The associated Fourier transform is called the *bispectrum*, noted B such that

$$\langle\delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle = (2\pi)^3 \delta_D^3\left(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3\right) B\left(\mathbf{k}_1, \mathbf{k}_2\right), \tag{1.53}$$

it can be also written as $B\left(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\right)$ with the implicit condition $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$.

Remark: As for the power spectrum, while the isotropy *i.e.* the rotational invariance holds, the bispectrum will only depend on the magnitude of k_1, k_2 and not on their orientations. Additionally, the bispectrum is proportional to $\delta_D^3\left(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3\right)$ since the Universe is assumed homogeneous leading to translation invariance.

1.2.2 Linear perturbation theory

Assuming that galaxies or the CMB temperature correctly trace the matter, it seems possible to measure the density contrast field of the δ matter as well as the statistical moments that they describe it. This section aims to model theoretically the evolution of δ during the history of the Universe.

In the following, we will work at linear order in perturbation theory, neglecting all the second-order terms.

1.2.2.1 Perturbed metric

As presented in Section 1.1, the Universe is correctly described as a homogeneous and isotropic expanding Universe. The inhomogeneities can be then treated as perturbations of the smooth FLRW metric:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}, \tag{1.54}$$

where $\bar{g}_{\mu\nu}$ is the metric given by (1.18), $h_{\mu\nu}$ is a symmetric 4-tensor and $|h_{\mu\nu}| << \bar{g}_{\mu\nu}$.

The perturbed metric $h_{\mu\nu}$ can be decomposed with the usual *scalar-vector-tensor (SVT)*

decomposition (Bardeen 1980) as

$$h_{00} = -2a^2\Phi, \tag{1.55}$$

$$h_{i0} = a^2\left[2\frac{\partial B}{\partial x^i} + 2B_i\right], \tag{1.56}$$

$$h_{ij} = a^2\left[-2\Psi\delta_{ij} + 2\frac{\partial^2 E}{\partial x^i \partial x^j} + 2\frac{\partial E_i}{\partial x^j} + 2\frac{\partial E_j}{\partial x^i} + 2E_{ij}\right], \tag{1.57}$$

where Φ, Ψ, B and E are scalars, B_i and E_i are divergence-less vector fields and E_{ij} is a traceless, symmetric and divergence-less tensor field. The metric is a symmetric 4-tensor, which implies the existence of 10 degrees of freedom. These degrees of freedom are indeed recovered through the SVT decomposition, which yields $4+2\times2+2=10$. The different modes resulting from the SVT decomposition correspond to different physical phenomena: the scalar modes correspond to the gravitational potential, the vector modes to gravito-magnetism, while the tensor mode to gravitational waves.

The coordinate system used to describe the perturbed metric is not unique, and a different choice can lead to another expression for the metric tensor. However, the underlying physics remains the same. This is known as the *gauge-freedom*. In this case, out of the 10 potential degrees of freedom, only 6 are independent. To fix the remaining 4 degrees of freedom, we can set B, E and B_i to zero. This choice is called the *Newtonian gauge* or *longitudinal gauge*, and Φ and Ψ are known as the *Bardeen potentials* in this gauge. Hence, the perturbed metric can be expressed as follows[6]

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Phi)\mathrm{d}\eta^2 + (1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{1.58}$$

In this expression, only scalar modes are considered since they evolve independently of the vector and tensor modes. Indeed, the scalar, vector and tensor parts are not coupled to each other in the linear theory *i.e.* first order in perturbation theory. The vector mode, in the absence of a stress tensor, decays very quickly as a^2 and so disappears at the time of our observation, see Appendix 4.2. Additionally, they are not produced during the early Universe, see Chapter 2. The tensor mode is an order of magnitude smaller than the scalar ones.

The Newtonian gauge has the great advantage that scalar modes, described by the Bardeen potentials Φ and Ψ, can easily be related to the Newtonian limit of gravity. In this limit, $\Psi = \Phi$ and Φ/Ψ will be equal to the Newtonian gravitational potential verifying the *Poisson equation*[7]:

$$\Delta\Psi = 4\pi G a^2 \bar{\rho}\delta. \tag{1.59}$$

Several other gauges exist, see Appendix 4.1, with other specificities, for instance, *spatially-flat gauge* is very convenient for computing inflationary perturbations or *synchronous gauge* is a popular gauge for numerical implementation of the perturbation equations.

Remark: In the early Universe, the amplitude of the Bardeen potentials Φ in real space, is about 10^{-5} and justifies the linear order used in the following perturbation analysis.

[6]Here, we follow the convention used in Peter and Uzan (2013). Note that several other conventions exist! In particular, Dodelson and Schmidt (2020) use

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Psi)\mathrm{d}\eta^2 + (1+2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right].$$

[7]This will be derived in the next section. The additional factor a^2 is because the coordinates are in comoving units.

1.2.2.2 Perturbed Einstein equations

The Einstein equations for the perturbed part of the metric is simply

$$\delta G_{\mu\nu} = \kappa \delta T_{\mu\nu}, \tag{1.60}$$

where $\delta G_{\mu\nu}$ is computed carefully from (1.58) - (1.9), and $\delta T_{\mu\nu}$ is the perturbed stress energy tensor derived in Appendix 1. In the Newtonian gauge (1.58) and considering only the scalar modes, it simplifies as

$$\begin{aligned} \delta T_{00} &= \bar{\rho} a^2 (\delta + 2\Phi) \\ \delta T_{0i} &= -\bar{\rho} a^2 (1+w) \partial_i v \\ \delta T_{ij} &= \bar{P} \left(h_{ij} + \frac{\delta P}{\bar{P}} \bar{g}_{ij} + \Pi^S_{ij} \right) \end{aligned}, \tag{1.61}$$

where δ is the usual contrast density $\delta = \delta\rho/\bar{\rho}$, $\bar{P} = w\bar{\rho}$ and v (*resp.* Π^S_{ij}) the scalar part of the 3-velocity (*resp.* anisotropic pressure).

Finally, at linear order, the perturbed Einstein equations (1.60) lead to[8]

$$\begin{aligned} &\Delta\Psi = 4\pi G a^2 \bar{\rho} [\delta - 3\mathcal{H}(1+w)v] \\ &\Psi - \Phi = 8\pi G a^2 \bar{P} \Pi \\ &\Psi' + \mathcal{H}\Phi = -4\pi G a^2 \bar{\rho} (1+w) v \\ &\Psi'' + 3\mathcal{H}\left(1 + c_s^2\right)\Psi' + \left[2\mathcal{H}' + \mathcal{H}^2\left(1 + 3c_s^2\right)\right]\Psi - c_s^2 \Delta\Psi \\ &= -\left(\mathcal{H}^2 + 2\mathcal{H}'\right)\left[\frac{\Gamma}{2} + \left(3\mathcal{H}^2 + 2\mathcal{H}'\right)\Pi + \mathcal{H}\Pi' + \frac{1}{3}\Delta\Pi\right] - 9c_s^2\mathcal{H}^4\Pi \end{aligned}, \tag{1.62}$$

where $'$ corresponds to the derivative respect to the conformal time η and $\mathcal{H} = a'/a = aH$ is the *conformal Hubble parameter*. Using the Friedmann equation (1.28), it can be expressed as

$$\frac{3}{2}\mathcal{H}^2 = 4\pi G a^2 \bar{\rho}. \tag{1.63}$$

The *speed of sound* c_s and the *entropy perturbation* Γ are defined as

$$\delta P = c_s^2 \delta\rho + \bar{P}\Gamma. \tag{1.64}$$

The two last equations describe the time evolution of the two scalars Ψ, Φ. The first equation is the general relativity version of the Poisson equation and looks like the usual form with $\delta^C = \delta - 3\mathcal{H}(1+w)v$. The second equation sets the equality between Ψ and Φ if the scalar mode of the anisotropic stress is null. This case appears when only perfect fluids are considered, and simplifies a lot the above equations.

The energy-momentum continuity equations (1.12) can be used instead of some Einstein equations, in particular, to replace the two-time evolution equations above. The background (order 0) continuity equation is given in (1.29) and the first-order perturbation continuity equations are in the Newtonian gauge:

$$\delta' + 3\mathcal{H}\left(c_s^2 - w\right)\delta = -(1+w)\left(\Delta v - 3\Psi'\right) - 3\mathcal{H}w\Gamma, \tag{1.65}$$

$$v' + \mathcal{H}\left(1 - 3c_s^2\right)v = -\Phi - \frac{c_s^2}{1+w}\delta - \frac{w}{1+w}\left[\Gamma + \frac{2}{3}\Delta\Pi\right]. \tag{1.66}$$

The equation (1.65) is generally referred to the *energy continuity equation* and (1.66) to the *Euler equation.*

[8] The first equation is obtained with $\delta G^0_0 + 3\mathcal{H}\partial_i^{-1}\delta G^i_0$, the second one with the traceless part of δG^j_i, the third is $\partial_i^{-1}\delta G^i_0$ and the fourth is from $\delta G^j_i + 3c_s^3\delta G^0_0$.

1.2.2.3 The real Universe

Let's derive the final set of equations to describe the Universe we observe today. According to the latest measurement, the Universe is filled with 5 main fluid components: photons, neutrinos, cold dark matter, baryons (including electrons) and dark energy. The electrons and the nuclei are tightly coupled through Coulomb scattering and then can be treated as a single component, the baryons. Fig. 1.9 gives a summary of the different components constituting the Universe and the different interactions between them and the metric. Dark energy perturbations will be neglected in the following because even if they exist, we will observe the Universe at a time when the dark energy is never dominant, see Fig. 1.4. Hence, the overall density is

$$\rho = \rho_b + \rho_{cdm} + \rho_\gamma + \rho_\nu + \rho_{DE}, \tag{1.67}$$

and the pressure (both for background and perturbations) of cold dark matter and baryons can be ignored: $w_{cdm} = w_b = 0 = c_{cdm} = c_b$. For simplicity here, the neutrinos are assumed massless[9]: $w_\nu = 1/3$, and we assume no entropy perturbation for each component: $\Gamma = 0$ such that $c_s^2 = \delta P/\delta\rho = w$.

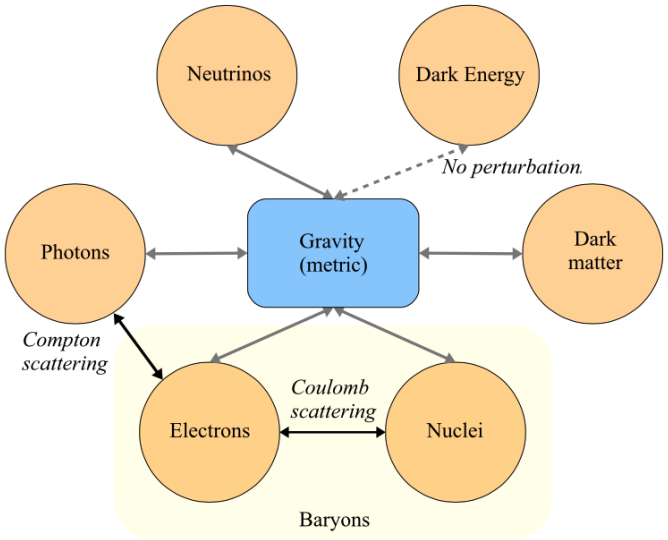


Figure 1.9: Different components that constitute the Universe and the interaction between them (Dodelson and Schmidt 2020). Due to their tight coupling through Coulomb scattering, electrons and nuclei will be considered as a single component (baryons). The perturbation of the Dark Energy will be neglected.

The cold dark matter does not interact with the other components such that the conservation equations are

$$\begin{aligned} \delta'_{cdm} &= -\Delta v_{cdm} + 3\Psi' \\ v'_{cdm} &= -\mathcal{H} v_{cdm} - \Phi \end{aligned} \tag{1.68}$$

The interaction of the neutrinos will be ignored since we are only interested in times much after neutrino decoupling which occurs when the Universe was just $\sim$ 1s. old. Then,

$$\begin{aligned} \delta'_\nu &= -\frac{4}{3}\Delta v_\nu + 4\Psi' \\ v'_\nu &= -\frac{1}{4}\delta_\nu - \frac{1}{6}\Delta\Pi_\nu - \Phi. \end{aligned} \tag{1.69}$$

[9]If neutrinos are considered massive, $w_\nu \neq 1/3$ and the equation of state should be estimated from P/ρ where P and ρ are computed by integrating the probability density function $f(\mathbf{x}, \mathbf{p}, t)$. However, some accurate approximations exist to circumvent this computation *e.g.* Nascimento (2023).

However, the baryons and photons interact via the *Compton scattering*, which creates a collision term between these two components. The collision term for δ_b' and δ_γ' can be neglected and only the momentum transfer between photons and baryons is relevant. Thus, the conservation equations for the baryons are

$$\begin{aligned} \delta_b' &= -\Delta v_b + 3\Psi' \\ v_b' &= -\mathcal{H} v_b - \Phi + \text{collision term}, \end{aligned} \tag{1.70}$$

and the ones for the photons are

$$\begin{aligned} \delta_\gamma' &= -\frac{4}{3}\Delta v_\gamma + 4\Psi' \\ v_\gamma' &= -\frac{1}{4}\delta_\gamma - \frac{1}{6}\Delta \Pi_\gamma - \Phi + \text{collision term}. \end{aligned} \tag{1.71}$$

To be completely rigorous, each component should be described from a statistical point of view, using its probability density function $f(\mathbf{r}, \mathbf{p}, t)$ which respects the *Boltzmann equation* (see, for instance, Dodelson and Schmidt 2020, for a full derivation). In particular, the continuity equation is obtained by taking the zeroth moment of the Boltzmann equation, while the Euler equation is obtained by taking the first moment.

In the Newtonian gauge and for the scalar modes, the Boltzmann equation at linear order is

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{p}{E}\frac{\hat{p}^i}{a}\frac{\partial f}{\partial x^i} - \left[H - \dot{\Psi} + \frac{E}{ap}\hat{p}^i \frac{\partial \Phi}{\partial x^i}\right] p \frac{\partial f}{\partial p} = C\left[f\right], \tag{1.72}$$

where $C\left[f\right]$ is the collision term and $E/p = 1$ for massless particles.

Considering Compton scattering between the baryons and the photons, the Boltzmann equation for the photons is[10]

$$\Theta' + \hat{p}^i \left(\frac{\partial \Theta}{\partial x^i} + \frac{\partial \Phi}{\partial x^i}\right) = \Psi' + \tau' \left[\Theta_0 - \Theta + \hat{\mathbf{p}} \cdot \mathbf{v}_\mathrm{b}\right], \tag{1.73}$$

where Θ is the temperature contrast defined by $T(\mathbf{x}, \hat{\mathbf{p}}, t) = T(t)\left[1 + \Theta(\mathbf{x}, \hat{\mathbf{p}}, t)\right]$, Θ_0 is the temperature monopole

$$\Theta_0(\mathbf{x}, t) \equiv \frac{1}{4\pi}\int d\Omega' \Theta\left(\hat{\mathbf{p}}', \mathbf{x}, t\right), \tag{1.74}$$

and $\tau' = a n_e \sigma_T$ is the inverse of the photon mean path, with n_e is the free electron density and σ_T the Thomson cross-section.

The temperature is usually decomposed into its spherical harmonic components in the Fourier space, a method known as the *Boltzmann hierarchy*. Here, for simplicity, only the first two multipoles are considered. They correspond to the familiar quantities: $\Theta_0 = \frac{1}{4}\delta_\gamma, \Theta_1 = \frac{1}{3}v_\gamma$. This approximation is particularly relevant in the case of the *thight-coupling regime* where $v_\gamma - v_b = 0$ at zeroth order and then, allowing the $\ell \geq 2$ multipoles to be ignored.

Hence, the Euler equation for the photons (1.71) becomes

$$v_\gamma' = -\Phi - \frac{1}{4}\delta_\gamma + \tau'\left(v_b - v_\gamma\right). \tag{1.75}$$

Similarly, the Euler equation for the baryons (1.70) is

$$v_b' = -\mathcal{H} v_b - \Phi + \frac{\tau'}{R}\left(v_\gamma - v_b\right), \text{where } R(\eta) = \frac{3}{4}\frac{\Omega_b(\eta)}{\Omega_\gamma(\eta)}. \tag{1.76}$$

[10] Here, we neglected the angular dependence of Compton scattering and ignored the contribution of the polarization field.

Finally, the perturbation of the metric is described by four independent equations given by (1.62). However, when using the continuity and Euler equations to describe the evolution of perturbations in the fluid, only two of these equations remain independent. In practice, the first two equations of (1.62) are chosen.

1.2.3 Linear perturbation evolution

The above set of equations looks horrific and will be solved numerically. In only a few cases, some analytical derivations can be performed, see Section 1.2.3.4.

1.2.3.1 Transfer function and linear power spectrum

As described in Section 1.2.1.2, the contrast density fields δ will be represented by its statistical moments such that the 2-point correlation function. In the following, we will only be interested in the associated power spectrum, since the evolution equations (1.68)–(1.71) are simpler in Fourier space where $\boldsymbol{\nabla} = ik$ and $\Delta = -k^2$. Hence, in Fourier space and at linear order, each mode k evolves independently.

Therefore, it is relevant to use *transfer functions* which describe the evolution of perturbed quantities from the initial conditions. A generic transfer function T is defined by

$$A(k,z) = T_{B\to A}(k,z) \times B(k, z_{prim}), \tag{1.77}$$

where A and B are any perturbed quantities and z_{prim} is the redshift at which the initial conditions are fixed.

The most useful is the transfer function which describes the evolution of the matter contrast density: $\delta_m(k,z) = T_m(k,z)\delta_m(k,z_{prim})$ which can be computed from the equations given in Section 1.2.2.3. Thus, the matter linear power spectrum is

$$P_{lin}(k,z) = T_m(k,z)^2 P_{ini}(k), \tag{1.78}$$

where P_{ini} is the initial matter power spectrum given by the initial conditions.

1.2.3.2 Initial conditions

To solve the set of equations given in Section 1.2.2.3, we need to specify the initial density and velocity perturbations for each species. The initial conditions are set during the radiation-dominated epoch ($\Omega_r \approx 1$ and $\Omega_m \ll 1$), sufficiently early that all scales k of *interest* are outside the horizon. These modes are called *super-Hubble* and respect $k \ll \mathcal{H}$.

As described in Chapter 2, the simple model of inflation (single field inflation) generates *adiabatic perturbations* such that all component velocity perturbations are equal: $v_i = v$, and the density perturbations are related:

$$\frac{\delta_i}{1+w_i} = \frac{\delta}{1+w}. \tag{1.79}$$

Therefore, the initial perturbations can be specified *for only* one specie or equivalently, via the Einstein equations, for *one* perturbed metric quantity. Since the *comoving curvature perturbation* $\mathcal{R}$, see (A.38), is constant for adiabatic perturbations and modes k outside the horizon (A.42), it is convenient to use it to describe the initial conditions.

The comoving curvature perturbation is related to the Bardeen potential Φ for modes outside the horizon through (A.47),

$$\mathcal{R}_{ini} = -\frac{3}{2}\Phi_{ini} \tag{1.80}$$

Finally, the initial conditions are set via $\mathcal{R}_{ini}$ such that $\Phi = const.$ and $\Phi' = 0$. Then, the initial density and velocity perturbations are computed from (1.79) and from (1.68)–(1.71),

$$v_{ini} = -\frac{\Phi_{ini}}{2\mathcal{H}_{ini}}, \quad \delta_{r,ini} = -2\Phi_{ini}, \quad \delta_{m,ini} = -\frac{3}{2}\Phi_{ini}. \tag{1.81}$$

Remark: As explained in Chapter 2, single field inflation will generate the following initial curvature power spectrum

$$P_{\mathcal{R}}(k) = \frac{2\pi^2}{k^3} A_s \left(\frac{k}{k_{\text{pivot}}}\right)^{n_s - 1}, \tag{1.82}$$

where n_s is expected to be closed to 1 and A_s is the amplitude of the initial power spectrum at k_{pivot} which is generally chosen to be 0.05 Mpc. Hence, to compute the matter power spectrum (1.78), one only needs to compute the transfer function T_m since $\delta_{m,ini}$ can be related to $\mathcal{R}_{ini}$.

1.2.3.3 Numerical solution

The prediction of the linear power spectrum (1.78) depends on the primordial power spectrum (1.82) and on the transfer function $T(k, z)$ which is computed from the equation set given in Section 1.2.2.3. To assess these quantities, a numerical approach were developed during the past 20 years to resolve the complete set of Boltzmann equations. The most popular Boltzmann solver codes are `CAMB`[11] and `CLASS`[12][13]. In the following, we will use `CLASS` with the friendly-user wrapper `cosmoprimo`[14].

A Boltzmann solver computes numerically the time evolution of the perturbed quantities in Fourier space. For instance, the evolution of the density perturbations and the Bardeen potentials for the mode $k = 0.1$ [Mpc^{-1}] in the fiducial Planck 18 cosmology are displayed in Fig. 1.10. The velocity perturbations are not shown here for convenience. The following section aims to explain quantitatively the shape of the curves displayed in this figure.

In general, in galaxy surveys, the evolution of cosmological perturbations is only useful to compute the linear matter power spectrum or the transfer function, in order to compare it with the one measured. Fig. 1.11 shows the linear matter power spectrum $P_{lin}(k)$ at redshift $z = 0$ and the associated transfer function $T(k, z = 0)$ for the fiducial Planck 18 cosmology and two other cosmologies close to the fiducial one.

1.2.3.4 Qualitative study

Let us derive the behavior of $T(k, z)$ displayed in Fig. 1.11. Since in linear theory, each k-mode evolves independently, we need to trace the history of a k-mode that undergoes different

[11] https://camb.info/

[12] http://class-code.net/

[13] Note that `CLASS` follows the convention of Ma and Bertschinger (1995). In particular, the Newtonian gauge is defined as

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Psi)\mathrm{d}\eta^2 + (1-2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right],$$

swapping $\Psi \leftrightarrow \Phi$ in our notation (1.58).

[14] https://github.com/cosmodesi/cosmoprimo

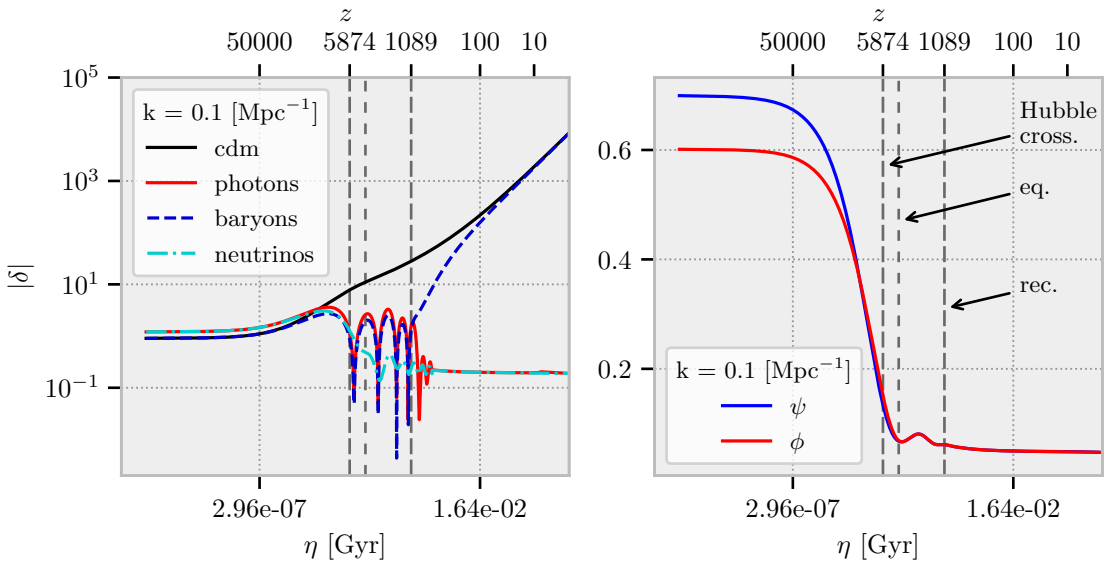


Figure 1.10: Time evolution of the density perturbations (right) and the Bardeen potentials (left) for the mode $k = 0.1$ [Mpc^{-1}] in the fiducial Planck 18 cosmology. The vertical gray dashed lines represent, from left to right, the moment when the mode k crosses the Hubble horizon, the radiation-matter equality and the recombination.

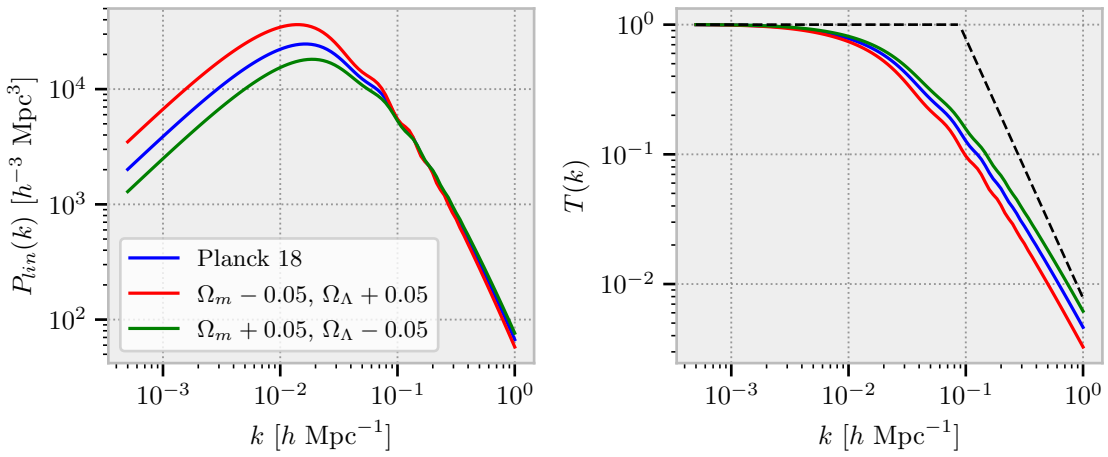


Figure 1.11: Linear power spectrum and associated transfer function (normalized to 1 at $k = 0$) at redshift $z = 0$ for the fiducial Planck 2018 parameters (Planck Collaboration et al. 2020) and some fluctuations around the fiducial values. The black dashed line represents the general form of the transfer function given by (1.100).

configurations: modes remaining super-Hubble ($k \ll \mathcal{H}(z = 0)$) *i.e.* never entering the horizon, modes entering the horizon during matter area, modes entering the horizon during radiation area and endure the transition between the radiation and the matter area.

In the following qualitative study, we assume that the Universe consists only of matter and photons which do not interact through Compton scattering, and neglecting the multipoles $\ell \geq 2$ in the Boltzmann hierarchy (tight coupling during radiation area and irrelevant on the evolution

of the matter distribution during the matter area). We introduced the evolution variable

$$y \equiv \frac{a}{a_{\rm eq}} = \frac{\rho_{\rm m}}{\rho_{\rm r}}, \tag{1.83}$$

to quantify the transition between the radiation area ($y \ll 1$) and the matter area ($y \gg 1$) which happens when $\Omega_{m,0} a_{eq}^{-3} = \Omega_{r,0} a_{eq}^{-4}$. The corresponding Fourier mode is

$$k_{eq} = \mathcal{H}_{eq} = H_0 \sqrt{2\Omega_{m,0} a_{eq}^{-1}} = H_0 \Omega_{m,0} \sqrt{2\Omega_{r,0}^{-1}}. \tag{1.84}$$

► **Large scales** Let's study modes that stay super-Hubble ($k \ll \mathcal{H}$) during the radiation area and the epoch of equality, and then enter the horizon in the matter area. In this case, k terms can be neglected compared to the conformal time derivate $'$ in (1.68) and in (1.71). Hence, the contrast density δ_m and δ_r are related by $\delta_m = 3/4\delta_r$ and

$$\begin{aligned} \delta_{\rm m} &= 3\Phi - \frac{9}{2}\Phi_{ini} \\ \delta_{\rm r} &= 4\Phi - 6\Phi_{ini} \end{aligned}, \tag{1.85}$$

where the initial conditions are given in Section 1.2.3.2.

Using (1.62), the evolution of Φ as a function of y while $k \ll \mathcal{H}$ is given by (Kodama and Sasaki 1984)

$$\Phi(\mathbf{k}, y) = \Phi_{ini}(\mathbf{k}) \frac{1}{10y^3} \left[16\sqrt{1+y} + 9y^3 + 2y^2 - 8y - 16 \right]. \tag{1.86}$$

Notably, in the matter area

$$\Phi = \frac{9}{10}\Phi_i, \quad \delta_r = -\frac{12}{5}\Phi_i, \quad \delta_m = -\frac{9}{5}\Phi_i. \tag{1.87}$$

The gravitational potential Φ decreases by a factor $9/10$ because of the transition between the radiation and the matter area. This behavior can be also predicted from the fact that $\mathcal{R}$ remains constant for adiabatic perturbations and super-Hubble modes as shown in Appendix 4.1.5. The gravitational potential evolves, see (A.47), between the radiation and the matter domination as

$$\Phi(\mathbf{k}, \eta)|_{\text{super-Hubble}} = \begin{cases} -\frac{2}{3}\mathcal{R}(\mathbf{k}), & \text{radiation domination,} \\ -\frac{3}{5}\mathcal{R}(\mathbf{k}), & \text{matter domination.} \end{cases} \tag{1.88}$$

Additionally, one can show[15] that while the Universe is matter dominated, the gravitational potential remains constant as the mode crosses the horizon, and then $\Phi = (-3/5)\mathcal{R}$ even inside the horizon. Thus, the evolution of matter density perturbations for these modes is controlled by the Poisson equation ($\propto a$).

► **Small scales** In contrast to the large scales, let's consider modes entering the horizon during the radiation area, and then passing the epoch of equality as sub-Hubble modes. This is the case for the mode k used in Fig. 1.10. These modes are such that $k \gg k_{eq}$.

[15] In the matter-dominated Universe, the pressure can be ignored: $\bar{P} = w = c_s^2 = 0$ and $\delta P = \Pi = 0$, such that the last equation of (1.62) leads to $\Phi'' + 3\mathcal{H}\Phi + \left(2\mathcal{H}' + \mathcal{H}^2\right)\Phi = 0$. The two equations of Friedmann (1.28) give $2\mathcal{H}' + \mathcal{H}^2 = 0$, and Table 1.1, $\mathcal{H} = 2/\eta$. Finally, the gravitational potential respects $\Phi'' + (6/\eta)\Phi'$ whose the solution is $\Phi(\eta, \mathbf{x}) = C_1(\mathbf{x}) + C_2(\mathbf{x})\eta^{-5}$. Unless $C_1(\mathbf{x})$ is surprisingly small *i.e.* very special initial conditions, one can always neglect the decaying part such that $\Phi(\mathbf{x}, \eta) = \Phi(\mathbf{x})$.

The evolution of the matter density contrast δ_m is given by (1.68) and can be written as

$$\delta_m'' + \mathcal{H}\delta_m' = 3\Phi'' + 3\mathcal{H}\Phi' - k^2\Phi, \tag{1.89}$$

where from (1.62) the gravitational potential follows the evolution equation

$$k^2\Phi + 3\mathcal{H}\left(\Phi' + \mathcal{H}\Phi\right) = -4\pi G a^2\left(\rho_m\delta_m + \rho_r\delta_r\right), \tag{1.90}$$

or the Poisson equation

$$k^2\Phi = -4\pi G a^2\left[\rho_m\delta_m + \rho_r\delta_r + \frac{3\mathcal{H}}{k}\left(\rho_m v_m + \frac{4}{3}\rho_r v_r\right)\right]. \tag{1.91}$$

• **Radiation-dominated epoch** In this case, the radiation dominates ($\mathcal{H} = \eta^{-1}$) and Φ can be computed from only the radiation components. Inserting (1.71) in the perturbed Einstein equations gives

$$\Phi'' + \frac{4}{\eta}\Phi' + \frac{1}{3}k^2\Phi = 0, \tag{1.92}$$

where the growing mode solution is given by

$$\Phi(\eta) = 3\Phi_{ini}\frac{\sin\left(\frac{k\eta}{\sqrt{3}}\right) - \frac{k\eta}{\sqrt{3}}\cos\left(\frac{k\eta}{\sqrt{3}}\right)}{\left(\frac{k\eta}{\sqrt{3}}\right)^3}. \tag{1.93}$$

Hence, for sub-horizon scales (Solution (1.93) is in fact true for all scales during the radiation area), the radiation perturbations oscillate with constant amplitude such that

$$\begin{aligned} \delta_r &\approx -\frac{2}{3}(k\eta)^2\Phi \approx 6\cos\left(c_s k\eta\right)\Phi_{ini} \\ v_r &= \frac{1}{2}\left(k\eta^2\Phi' + k\eta\Phi\right) \approx \frac{9c_s}{2}\sin\left(c_s k\eta\right)\Phi_{ini} \end{aligned}, \tag{1.94}$$

where the speed of sound is $c_s = 1/\sqrt{3}$.

The matter perturbations are obtained from (1.89). The associated homogeneous equation admits two solutions $\delta = const.$ and $\delta = ln(k\eta)$. Then, Green's method ensures that

$$\delta_m(\eta) = C_1 + C_2\ln(k\eta) + \int_0^\eta \mathrm{d}u\left[\ln(ku) - \ln(k\eta)\right]u\left(k^2\Phi - 3\Phi'' - \frac{3}{u}\Phi'\right), \tag{1.95}$$

where $C_1 = -3/2\Phi_{ini}$, $C_2 = 0$ in order to match the initial value given in (1.81). For $k\eta \gg 1$, the integral stops changing as a function of its upper limit and can be approximated numerically using (1.93) leading to

$$\delta_m(k,\eta) \approx \begin{cases} -A\ln(Bk\eta)\Phi_{ini} & k\eta \gg 1 \\ -3/2\Phi_{ini}(k) & k\eta \ll 1 \end{cases}, \tag{1.96}$$

where $A = 9.0$ and $B = 0.62$.

The matter and radiation perturbations have two clearly different behaviors. The radiation perturbations undergo a competition between gravity and pressure, causing the oscillation of (1.94) around zero with constant amplitude. While the matter perturbations are only affected by the gravity created by the radiation. The matter falls towards the initial gravity well, then accelerates or slows down as the potential begins its oscillation (1.93). However, the amplitude of oscillation decreases and the potential is not able to stop the collapse before the first reversal of Φ. Thus, the matter density perturbations continue to grow (1.96).

• **Matter-dominated epoch** Once the modes are crossing the horizon during the radiation area, they will see the transition to the matter as sub-Hubble modes $k \ll \mathcal{H}$.

As described above, during the radiation dominated epoch δ_r oscillates without growing while δ_m grows logarithmically so that $\delta\rho_m = \delta_m\rho_m$ will become larger than $\delta\rho_r = \delta_r\rho_r$ even though ρ_m remains smaller than δ_r. Hence, Φ will be determined only by matter perturbations from some $y = y_{\delta eq} << 1$ where the radiation still dominates.

Combining (1.89) and (1.91) in the limit of sub-Hubble modes ($k \gg \mathcal{H}$) and neglecting the radiation perturbations ($y \gg y_{\delta eq}$), gives the equation of evolution for δ_m in a radiation-matter background known as the *Meszaros equation* (Meszaros 1974):

$$\frac{d^2\delta_m}{d^2y} + \frac{2+3y}{2y(y+1)}\frac{d\delta_m}{dy} - \frac{3}{2y(y+1)}\delta_m = 0. \tag{1.97}$$

This equation remains valid until dark energy becomes non-negligible. It admits two solutions:

$$\begin{aligned} D_1(y) &= y + \frac{2}{3} \\ D_2(y) &= \frac{15}{8}(2+3y)\ln\left(\frac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - \frac{45}{4}\sqrt{1+y}, \end{aligned} \tag{1.98}$$

where D_1 (*resp.* D_2) is the growing (*resp.* decaying) solutions with late time ($y \gg 1$) behavior: $D_1(y) \propto y$ and $D_2(y) \propto y^{-3/2}$.

To match the solution in radiation-dominated epoch (1.96) and the solution in the matter-dominated epoch $\delta_m(y) = C_1D_1(y) + C_2D_2(y)$, we will assume that the behavior of δ_m is not so dramatically different when $y \approx y_{\delta eq}$. The decaying mode is neglected since we are interested in late-time behavior. Hence, the density perturbation growing mode for $y \gg y_{\delta eq}$ is

$$\delta_m(k,y) \approx -\frac{3}{2}A\ln\left(\frac{4Be^{-3}}{y_k}\right)\left(y+\frac{2}{3}\right)\Phi_{ini}(k), \tag{1.99}$$

where y_k is the value of y when mode k enters the horizon $k = \mathcal{H}$.

• **Baryonic acoustic oscillations** For the moment, there is no interaction between the matter and the radiation. In the real Universe, a fraction of the matter is baryonic and interacts with photons through Compton scattering. The oscillations of radiation density perturbations during the radiation area (1.94) let a similar signature into the baryonic density perturbations, and then through gravity into the matter density perturbations. This oscillation is known as the *baryonic acoustic oscillations* (BAO).

These oscillations are visible in Fig. 1.10. This figure shows the time evolution of the density perturbations and the Bardeen potentials for a mode k that crosses the Hubble horizon during the radiation domination and is described by the equations above. In particular, the damped oscillations of Φ (1.93) are discernible during the Hubble crossing moment and the recombination, and as expected, the potential is constant well before the Hubble crossing moment and after the recombination. The difference between Φ and Ψ at the very early time is due to neutrinos in the fiducial Planck 18 cosmology, which have been neglected in this qualitative study.

In the same manner, the photon oscillations (1.94) are noticeable during the Hubble crossing time and the recombination in the density perturbations (here, for convenience, we plot $|\delta|$) and lead to the baryonic oscillations, while the cold dark matter density perturbations growth following (1.96) in the radiation-dominated area and (1.99) in the matter-dominated area. Once the recombination occurs, cold dark matter will attract the baryons.

These oscillations let a characteristic signature into the linear matter power spectrum and in the transfer function, typically around $k \sim 0.1\ h\ \mathrm{Mpc}^{-1}$, as shown in Fig. 1.11.

▶ **Transfer function** The previous studies have attempted to describe the behavior of δ_m at large and small scales in a simplified case, and help us to understand the general form of the transfer function, represented in the black dashed line in the right panel of Fig. 1.11.

The large scales ($k \ll k_{eq}$) study shows that the transfer function, as defined in (1.106) is very closed to 1 (at 9/10 factor according to the definition) for all scales entering the horizon after the equality: $k \ll k_{eq}$.

The small scales ($k \gg k_{eq}$) study shows that when the mode remains super-Hubble, $\delta_m \propto a^2$ (Poisson equation), then when the mode is sub-Hubble during the radiation-dominated epoch, $\delta_m \propto ln(a)$ and until the equality. Finally, during the matter-dominated epoch, $\delta_m \propto a$. The only difference between the modes is then when it enters the horizon and stops growing. Thus, a mode k entering the horizon at $\mathcal{H}_k$ does not grow by a factor of $(\mathcal{H}_k/\mathcal{H}_{eq})^2$ compared to a mode remaining super-Hubble. The transfer function is roughly proportional to $(k/k_{eq})^2$.

Hence, the transfer function, left panel of Fig. 1.11, should be closed to

$$T_m(k) \sim \begin{cases} 1 & k \ll k_{eq} \\ (k_{eq}/k)^2 & k \gg k_{eq} \end{cases}. \tag{1.100}$$

1.2.3.5 Late time evolution

To take advantage of the late time evolution of the sub-Hubble scales ($k \gg \mathcal{H}$) in the matter-dominated area, one can write the perturbations of the non-relativistic matter in the Newtonian limit (see Adamek et al. 2016, for the impact of general relativity). In this context, perturbations of a perfect fluid are described by the well-known continuity, Euler and Poisson perturbed equations in comoving coordinates[16] (Rieutord 2014):

$$\begin{cases} \delta' + \nabla \cdot ((1+\delta)\mathbf{u}) &= 0 \\ \mathbf{u}' + (\mathbf{u}\cdot\nabla)\,\mathbf{u} + \mathcal{H}\mathbf{u} &= \dfrac{1}{\rho}\nabla\delta P - \nabla\Phi \\ \Delta\Phi &= 4\pi G a^2 \bar{\rho}\delta \end{cases}, \tag{1.102}$$

where $c_s^2 = \delta P/\delta\rho$, and the time component and derivative were replaced by the conformal time η ($\cdot \to \prime$). These equations are the Newtonian limit of (1.68) and (1.62). During gravitational collapse and before the virialization, $\nabla\delta P$ can be neglected.

Similarly to the SVT decomposition, the velocity u can be decomposed[17] into its divergence $\theta = \boldsymbol{\nabla}\cdot\mathbf{u}$ and its vorticity $\mathbf{w} = \boldsymbol{\nabla}\times\mathbf{u}$. As mentioned in Section 1.2.2.1, the vector modes are diluted in expansion. Here, from the linear order of the Euler equation, $\mathbf{w} \propto a^{-1}$. Hence, only the scalar mode θ is generally considered *i.e.* $\mathbf{u}$ is supposed curl-free. In Fourier space, $\mathbf{u} = -i\theta\mathbf{k}/k^2$.

[16] Applying carefully the chain rules with (1.20), the derivatives become in comoving coordinates:

$$\begin{cases} \boldsymbol{\nabla}_{\mathbf{r}} &\longrightarrow \dfrac{1}{a}\boldsymbol{\nabla}_{\mathbf{x}} \\ \partial_t f(\mathbf{r},t) &\longrightarrow \partial_t f(\mathbf{x},t) - H\,(\mathbf{x}\cdot\boldsymbol{\nabla}_{\mathbf{x}})\,f(\mathbf{x},t) \end{cases} \tag{1.101}$$

[17] Helmholtz decomposition.

At the linear order ($\delta, u_i \ll 1$), (1.102) becomes

$$\begin{aligned} \delta' + \theta &= 0 \\ \theta' + \mathcal{H}\theta &= -\Delta\Phi' \end{aligned} \tag{1.103}$$

which lead, with the Poisson equation to

$$\delta'' + \mathcal{H}\delta' - \frac{3}{2}\mathcal{H}^2\Omega_m\delta = 0. \tag{1.104}$$

This 2^{nd} order linear partial differential equation admits two independent solutions:

$$\delta(\mathbf{k}, \eta) = D^+(\eta)\delta^+_{late}(\mathbf{k}) + D^-(\eta)\delta^-_{late}(\mathbf{k}), \tag{1.105}$$

where δ^+_{late} (growing modes), δ^-_{late} (decaying modes) are set at redshift z_{late} when $\Omega_m = 1$ after the recombination. In the presence of dark energy, D should be computed numerically. However, in an *Einstein-deSitter* Universe ($\Omega_m = 1$),

$$D^+(t) = t^{\frac{2}{3}} = a, \quad D^-(t) = t^{-1} = a^{-3/2}.$$

In the following, only growing modes (+) will be considered and D^+ is called the linear *growth factor*.

To take advantage of this late evolution, it is usual to define T_m only between the primordial fluctuations and the fluctuations at z_{late} after the recombination ($z_{late} \ll z_{rec}$)such that

$$\delta_m(k, z) = T_m(k, z_{late}) \times \frac{D^+(z)}{D^+(z_{late})} \times \delta_{prim}(k). \tag{1.106}$$

Hence, the growth of the structure formation during the matter-dominated area is completely encoded in the growth factor D^+. Note that generally, the transfer function is normalized to 1 at $k = 0$ and therefore does not depend on the choice of z_{late}. This late-time evolution in EdS Universe is only true when neutrinos are neglected. Otherwise, the evolution function continues to depend on k.

For growing modes,

$$\delta' = \frac{D^{+\prime}}{D^+}\delta = f\mathcal{H}\delta, \tag{1.107}$$

where f is the *growth rate* defined by

$$f = \frac{d\ln D^+}{d\ln a}. \tag{1.108}$$

The growth rate can be accurately parametrized by ${\Omega_m}^\gamma$. For instance, Linder (2005) shows that in a flat ΛCDM Universe with $\Omega_m \in [0.22, 1]$, $f \simeq \Omega_m^{0.55}$ to better than 0.05%.

From (1.103), the linear matter velocity can be expressed in Fourier space according to the matter density perturbation,

$$\mathbf{u}(\mathbf{k}) = if\mathcal{H}\frac{\mathbf{k}}{k^2}\delta(\mathbf{k}). \tag{1.109}$$

This description, at linear order, does not hold for small scales $k > 0.1\ h\ \mathrm{Mpc}^{-1}$, and one needs to use the full set of equations (1.102) without assuming the linear approximation as it is done in Appendix 3.

1.3 Observational cosmology

In the previous sections, we discussed the evolution of large-scale structures in the Universe, starting from primordial fluctuations and progressing to the late-time universe. Chapter 2 will explain the creation of primordial fluctuations. In order to validate this description and constrain the fundamental parameters of the ΛCDM model, the primary goal of observational cosmology is then to observe the statistical distribution of the matter in the Universe, which can be accomplished, for instance, through observations of light emitted by galaxies.

The first subsection will describe how the underlying matter field can be linked to observable objects. The second one will provide an overview of the different standard observations used to validate the description presented in this chapter and constitute the current state-of-the-art of our knowledge.

1.3.1 Cosmological tracers

Galaxy survey, as it will be used in this dissertation, observes millions of galaxies to track the matter distribution of the Universe. Galaxies are expected to be formed in the denser regions of the Universe. However, galaxies do not measure the matter density field itself but are the result of a complex formation process in which the matter density field plays a crucial role.

Let's describe how the observed distribution of galaxies δ_g^{obs} is linked to the matter one δ:

$$\delta_g^{obs} \xrightarrow[1.3.1.2]{1.3.1.3} \delta_g \xrightarrow{1.3.1.1} \delta$$

1.3.1.1 Tracer bias

Galaxies form within the gravitational potential wells created by dark matter in high-density excursions of the matter density field (Kaiser 1984). The *galaxy bias* relates the galaxy density field to the underlying density matter field. Since the formation process is affected by various physical processes, such as gas cooling, star formation, and feedback from supernovae and black holes, each type of galaxy is more or less clustered than the underlying matter and has a different bias.

The galaxy bias can be described in a perturbative bias expansion theory

$$\delta_g(\mathbf{x}, z) = \sum_O b_O(z) O(\mathbf{x}, z), \tag{1.110}$$

where O are statistical fields describing the environment properties on which galaxy density can depend. For instance, $(\delta,\ \delta^2,\ \ldots)$ or the tidal field $K_{ij}(\mathbf{k}) = \left[(k_i k_j)/k^2 - 1/3\delta_{ij}\right] \delta(\mathbf{k})$, ... have to be considered in this expansion.

The rigorous description[18] of the bias expansion is out of the scope of this dissertation which will be focused on scales where the linear theory can be applied. In linear order, the galaxy bias reads as

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + \epsilon, \tag{1.111}$$

where ϵ is the stochastic contribution which corresponds to the presence of random small-scale fluctuations, not correlated to the long-wavelength fluctuations described by δ and to the fact that galaxies are a discrete sampling of the matter density field. This contribution is modeled as with noise *i.e.* $P_\epsilon(k) = const.$ and $\langle \epsilon \delta \rangle = 0$.

[18] See Section 2 of Desjacques et al. (2018) for a complete derivation.

Remark: The expansion bias needs to be performed up to the third order (δ^3) when the non-linear power spectrum is computed with the 1-loop correction since $\delta^{(2)} \propto \left(\delta^{(1)}\right)^2$.

1.3.1.2 Redshift space distortions (RSD)

As explained in Section 1.1.2.4, the redshift observed is composed of several contributions. Hence, when transforming redshifts into distances, the observed distance $\mathbf{s}$ is slightly different from the real distance $\mathbf{r}$ without the effect of the peculiar velocity $\mathbf{u}$. They are linked by

$$\mathbf{s} = \mathbf{r} + \frac{(\mathbf{u} \cdot \hat{\mathbf{z}})}{\mathcal{H}} \hat{\mathbf{z}} = \mathbf{r} + u_z \hat{\mathbf{z}}, \tag{1.112}$$

where $\hat{\mathbf{z}}$ is the line-of-sight (LOS).

At large scales, galaxies fall coherently into deep gravity well and the density field becomes squashed along the (LOS) (we are only sensitive to the velocity along the LOS). The clustering amplitude is then higher along the LOS. This is known as the *Kaiser effect* but was first mentioned by Sargent and Turner (1977). Large scales are correctly described by the linear theory. This effect is shown on the left of Fig. 1.12, the observed density field (full line) looks asymmetric.

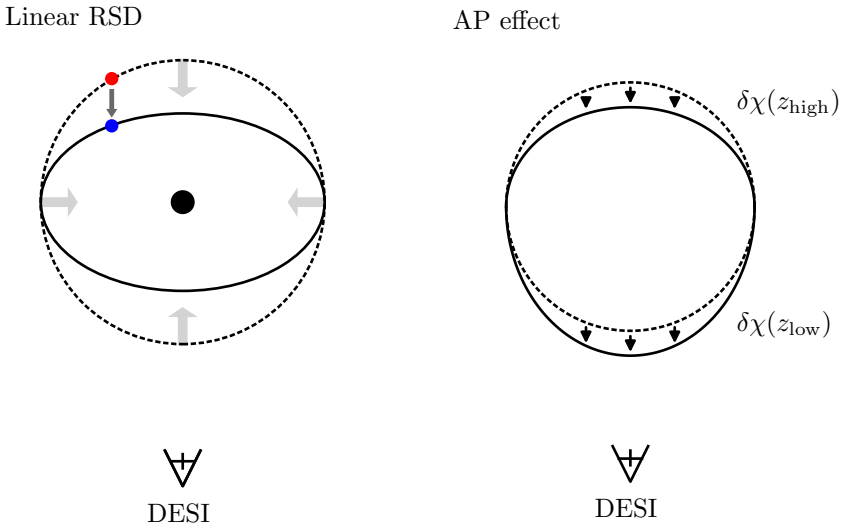


Figure 1.12: Left: Redshift space distortion effects for linear and non-linear cases. In both cases, the dashed (*resp.* full) line represents a contour of constant density (*resp.* the observed contour for an observer in the vertical direction which looks asymmetric) around a central overdensity *i.e.* deep gravity well. The Kaiser effect (linear order) is displayed on the left. The coherent velocities squash the density field along the line of sight. A grey arrow has a typical size of 1 h^{-1}Mpc. Right: The use of wrong cosmology in the redshift-distance conversion caused the Alcock-Paczyński distortion. Since the comoving distance depends on the redshift, the additional displacement is not the same for all the redshifts. The observed tracer field is then asymmetric (from the dashed circle if the displacement is the same for all redshifts to ellipse). Inspired from Dodelson and Schmidt (2020).

► **Linear order** First, let's try to link the density contrast in redshift space δ^s to the density contrast in real space δ using the linear order of the perturbation theory. The peculiar velocity in Fourier space is then given by (1.109). The density conservation between the redshift and the

real space implies

$$\delta^s(\mathbf{s}) = \left|\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{r}}\right|^{-1} [1+\delta(\mathbf{r})] - 1, \tag{1.113}$$

and in Fourier space

$$\begin{aligned}\delta^s(\mathbf{k}) &= \int \mathrm{d}^3 r \left(1+\delta(\mathbf{r}) - \left|\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{r}}\right|\right) e^{-i\mathbf{k}\cdot\mathbf{s}} \\ &= \int \mathrm{d}^3 r \left(\delta(\mathbf{r}) - \partial_z u_z\right) e^{-i(k\mu u_z + \mathbf{k}\cdot\mathbf{r})}\end{aligned} \tag{1.114}$$

where $\mu \equiv \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{z}}$ is the directional cosine between the line-of-sight $\hat{\mathbf{z}}$ and the wavelength $\mathbf{k}$. Using (1.109) and assuming $|\mu k u_z \ll 1|$, (1.114) becomes

$$\begin{aligned}\delta^s(\mathbf{k}) &\approx \int \mathrm{d}^3 r \left(\delta(\mathbf{r}) - \partial_z u_z(\mathbf{r})\right) e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &= \delta(\mathbf{k}) - \int \mathrm{d}^3 r e^{-i\mathbf{k}\cdot\mathbf{r}} \partial_z \left[\int \frac{\mathrm{d}^3 k'}{(2\pi)^3} u_z(\mathbf{k}') e^{i\mathbf{k}'\cdot\mathbf{r}}\right] \\ &= \delta(\mathbf{k}) - \int \frac{\mathrm{d}^3 k'}{(2\pi)^3} ik'\mu \frac{ifk'\mu\delta(\mathbf{k}')}{k'^2} \int \mathrm{d}^3 r e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} \\ &= \delta(\mathbf{k}) + f\mu^2 \int \mathrm{d}^3 k' \delta_D(\mathbf{k}-\mathbf{k}')\delta(\mathbf{k}') \\ &= \left(1+f\mu^2\right)\delta(\mathbf{k}).\end{aligned} \tag{1.115}$$

This is known as the *Kaiser formula* (Kaiser 1987). Hence, the power spectrum in redshift space,

$$P^s(\mathbf{k}) = P^s(k,\mu) = \left(1+f\mu^2\right)^2 P_{lin}(k), \tag{1.116}$$

depends on k, μ and not just on k as in the real space. This is the redshift space distortion, at least for the linear order.

► **Multipole decomposition for the linear order** The galaxy contrast density can be expressed at linear order as $\delta_g = b_1\delta$, where we neglect the stochastic contribution which is not impacted by the RSD. Since the velocity is caused by the gravity of the total matter field, galaxies are assumed to have a similar velocity to the underlying matter field such that $u_g = u$. Hence, the Kaiser formula (1.115) for galaxy contrast reads as

$$P^s(\mathbf{k}) = \left[b_1 + f\mu^2\right]^2 P_{lin}(k) = b_1^2\left[1+\beta\mu^2\right]^2 P_{lin}(k), \tag{1.117}$$

where $\beta = f/b_1$, and we have dropped $_g$ in P_g^s.

To compress the information of $P^{(s)}(\mathbf{k})$, it is useful to use the Legendre decomposition, given by

$$P^s(k,\mu) = \sum_\ell P_\ell^s(k)\mathcal{L}_\ell(\mu), \tag{1.118}$$

where L_ℓ are the Legendre polynomes and P_ℓ^s are the multipoles. By construction, they are[19]

$$P_\ell^s(k) \equiv \frac{2\ell+1}{2}\int_{-1}^{1} \mathrm{d}\mu \mathcal{L}_\ell(\mu) P^s(k,\mu). \tag{1.119}$$

[19] One can also integrate over the angle θ between $\hat{\mathbf{k}}$ and $\hat{\mathbf{z}}$, and then divide by 4π instead of integrate over μ and divide by 2.

One can show that only the monopole ($\ell = 0$), quadrupole ($\ell = 2$), and hexadecapole ($\ell = 4$) are non-zero:

$$\begin{aligned} P_0^{(z)}(k) &= \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) P(k) \\ P_2^{(z)}(k) &= \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) P(k) \\ P_4^{(z)}(k) &= \frac{8}{35}\beta^2 P(k) \end{aligned} \quad . \tag{1.120}$$

Similarly to (1.118) – (1.119), the correlation function in redshift space can be also decomposed with Legendre polynomes and its multipoles are related to the power spectrum multipoles via[20]

$$\xi_\ell(r) = \frac{i^\ell}{2\pi^2} \int k^2 dk j_\ell(kr) P_\ell(k), \tag{1.121}$$

where j_l is the spherical Bessel function. Hence, defining $J_\ell(x) \equiv \int_0^x \xi(y) y^{\ell-1} dy$, the multipoles for the correlation function are:

$$\begin{aligned} \xi_0^{(z)}(s) &= \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \xi(s) \\ \xi_2^{(z)}(s) &= \left(\frac{4}{3}\beta + \frac{4}{7}\beta^2\right) \xi(s) + \left(-4\beta - \frac{12}{7}\beta^2\right) \frac{J_3(s)}{s^3} \\ \xi_4^{(z)}(s) &= \frac{8}{35}\beta^2 \xi(s) + \frac{12}{7}\beta^2 \frac{J_3(s)}{s^3} - 4\beta^2 \frac{J_5(s)}{s^5} \end{aligned} \tag{1.122}$$

Fig. 1.13 shows the 2-point statistics in redshift space, power spectrum (left) and correlation function (right), for $b = 2,\ \beta = 0.355,\ z = 0.7$.

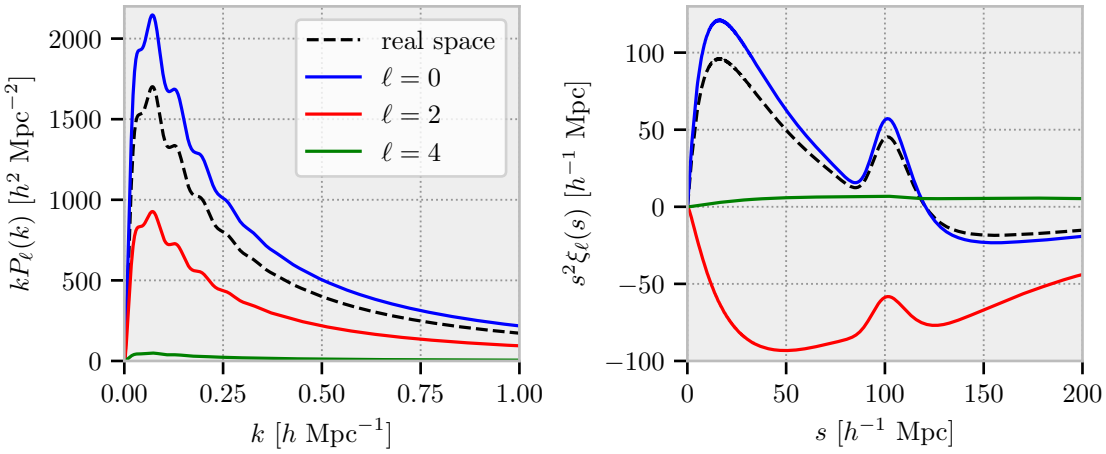


Figure 1.13: Monopole (blue), quadrupole (red), hexadecapole (green) of the power spectrum (left) and correlation function (right) in redshift space with $b = 2.3$, $\beta = 0.355$, $z = 0.7$. The real space function (black) is displayed for comparison. The odd multipoles are null.

Remark: In this case, the hexadecapole ($\ell = 4$) does not depend on the bias of the tracer and the odd multipoles are zero. They can be non-zero when considering cross-correlation. Finally, $l > 4$ are zero because the decomposition was performed using the linear order *i.e.* the Kaiser formula (1.115).

[20] The Rayleigh expansion of a plane wave is $e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_\ell i^\ell (2\ell + 1) j_\ell(kx) \mathcal{L}_\ell(\hat{k} \cdot \hat{x})$.

▶ **Improving the linear order** Besides the Kaiser effect, the density field is also stretched at small scales, typically below 1 h^{-1}Mpc, when galaxy velocities become too important respect to the size of the gravity well. This effect is called the *Finger of God* (FoG) effect. Additionally, when galaxies virialize and cease to have coherent velocities for only random motions, the clustering amplitude is damped along the LOS.

To be completely rigorous, we need to use the exact expression of P^s. Starting from given by (1.114), the redshift space power spectrum is

$$P^s(\mathbf{k}) = \int \mathrm{d}^3 r e^{-i\mathbf{k}\cdot\mathbf{r}} \left\langle e^{-ik\mu\Delta u_z} \left[\delta(\mathbf{x}) - \partial_z u_z(\mathbf{x})\right] \left[\delta\left(\mathbf{x}'\right) - \partial_z u_z\left(\mathbf{x}'\right)\right] \right\rangle . \tag{1.123}$$

At the linear order, we have omitted the term $e^{-ik\mu\Delta u_z}$. However, at small scales, *i.e.* after virialization, the random motions create out-of-phase velocities and the exponential term damps the redshift space power spectrum.

Motivated by (1.123), Scoccimarro (2004), Percival and White (2009) have proposed phenomenological models to describe the competition between Kaiser and Finger of God effects. The redshift space power spectrum can then be expressed as follows

$$P^s(k,\mu) = D_{\mathrm{FoG}}\left(k\mu\sigma_v\right) P^s_{\mathrm{Kaiser}}(k,\mu), \tag{1.124}$$

where σ_v^2 is the velocity dispersion defined as

$$\sigma_v^2 = \frac{1}{3}\int \frac{\mathrm{d}^3 q}{(2\pi)^3} \frac{P_{\theta\theta}(q)}{q^2} \approx \frac{f^2}{6\pi^2}\int \mathrm{d}q P_{lin}(q). \tag{1.125}$$

$P_{\theta\theta}$ is the velocity power spectrum. Here, the velocity is normalized by $\mathcal{H}$ (1.112). The velocity dispersion is generally treated as a free parameter. However, it can be directly computed from (A.12), in particular, the above approximation ($\approx$) is for the linear order. The Finger of Gods damping is modeled either as a Gaussian or Lorentzian damping:

$$D_{\mathrm{FoG}}(x) = \begin{cases} \exp\left(-x^2\right) & \text{(Gaussian)} \\ 1/\left(1+x^2\right) & \text{(Lorentzian)} \end{cases} \tag{1.126}$$

Finally, the power spectrum can be either modeled with the Kaiser effect or can be expanded to treat the non-linearity:

$$P^s_{\mathrm{Kaiser}}\left(k,\mu\right) = \begin{cases} \left(1+\beta\mu^2\right)^2 P_{\delta\delta}(k) & \text{(linear)} \\ P_{\delta\delta}(k) + 2f\mu^2 P_{\delta\theta}(k) + f^2\mu^4 P_{\theta\theta}(k) & \text{(non-linear)} \end{cases}, \tag{1.127}$$

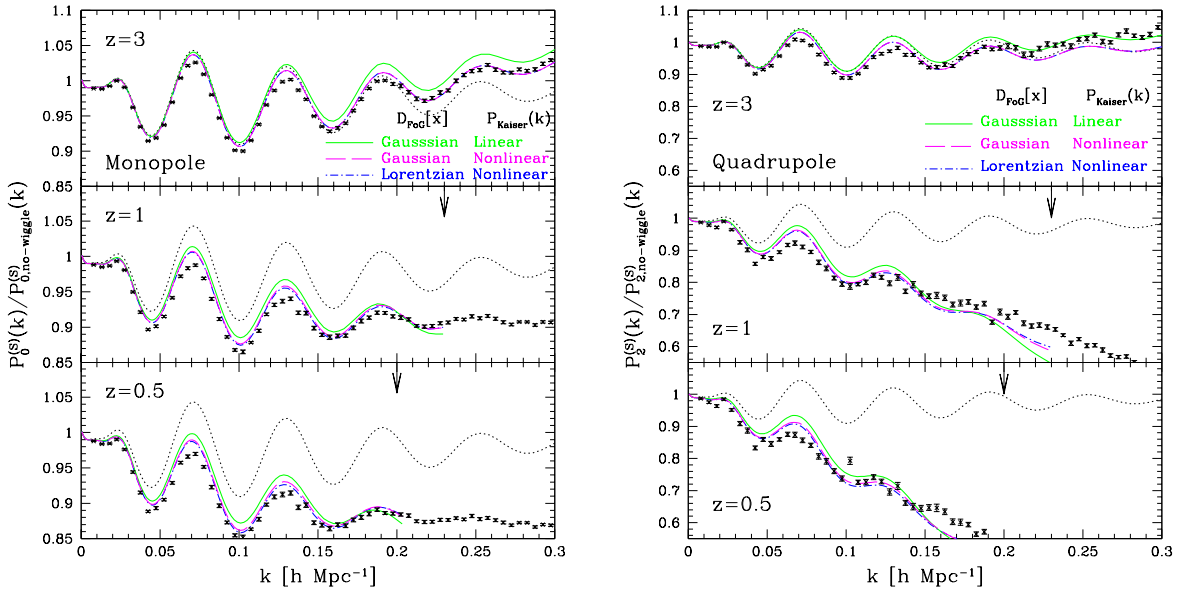
where $P_{\delta\theta}$ is the density-velocity cross power spectrum.

The phenomenological approach presented above provides a reasonably accurate description of the redshift space distortions, as shown in Fig. 1.14, up to relatively small scales. At a redshift of $z = 0.5$, it only deviates by 1% from simulations for k greater than 0.2 h Mpc^{-1}. This simple approach does not take into account the intermediate-scale interaction between the Kaiser and FoG effects. A more complex description, known as the TNS model (Taruya et al. 2010, Zheng and Song 2016), can address this issue. However, for this dissertation, the phenomenological model will be retained.

1.3.1.3 Alcock-Paczyński effect

In observational cosmology, we derive distances through the redshift measurement. To apply the conversion with (1.26), we need to assume a cosmology. A wrong one will then displace artificially the objects from their true position

$$\chi_{fid}(z) = \chi(z) + \delta\chi(z). \tag{1.128}$$



(a) **(b)**

Figure 1.14: Comparison between N-body simulations and phenomenological models (Taruya et al. 2010). The ratio of power spectra to smoothed linear power spectra *i.e.* without wiggles are plotted for monopole (right) and quadrupole (left). Several combinations of (1.126) and (1.127) are displayed. Here, the velocity dispersion σ_v is fitted to match the prediction to the N-body simulations.

Fig. 1.12 illustrates this phenomenon. The artificial displacement is unfortunately not the same for all redshifts, creating an artificial asymmetry in the observed density field. This effect is known as the *Alcock-Paczyński distortion*.

The observed position is (θ_1, θ_2, z) where (θ_1, θ_2) is the position in a sphere of unit radius and z the observed redshift centered around $\bar{z}$ the mean redshift of observed positions. Hence, the observed 3D position is

$$\mathbf{x}^{obs} = \chi_{fid}(z) \times \left(\theta_1, \theta_2, 1 - \frac{\chi_{fid}(\bar{z})}{\chi_{fid}(z)}\right). \tag{1.129}$$

The real 3D position in the correct cosmology is then,

$$\mathbf{x}\left(\mathbf{x}^{obs}\right) = \left(\alpha_\perp {x^{obs}}^1,\ \alpha_\perp {x^{obs}}^2,\ \alpha_\parallel {x^{obs}}^3\right), \tag{1.130}$$

where

$$\alpha_\perp = \left.\frac{\chi}{\chi_{fid}}\right|_{\bar{z}};\quad \alpha_\parallel = \left.\frac{H_{fid}}{H}\right|_{\bar{z}}. \tag{1.131}$$

We have use (1.128) for (x^1, x^2) and, since the observed tracers are generally in a narrow redshift slice, $x^3 = \chi(z) - \chi(\bar{z}) = (z - \bar{z})/H(\bar{z})$ where H is either from the true cosmology or from the fiducial one. For a similar reason and to remove the redshift dependence, the quantity $\alpha_\perp$ is also taken at $\bar{z}$.

Since the density is independent of the choice of the coordinates, the observed density contrast and the real one are related together (similarly for the power spectrum)

$$\delta_g^{obs}\left(\mathbf{k}^{obs}\right) = \delta_g(\mathbf{k})\big|_{\mathbf{k}=\left(\alpha_\perp^{-1}{k^{obs}}^1,\ \alpha_\perp^{-1}{k^{obs}}^2,\ \alpha_\parallel^{-1}{k^{obs}}^3\right)}. \tag{1.132}$$

Hence, even without galaxy peculiar velocity and so RSD effects, the galaxy power spectrum can have some anisotropy, since $\alpha_\perp$ and $\alpha_\parallel$ are different. This was first noticed by Alcock and Paczynski (1979) and it is known as the *Alcock-Paczyński effect* (AP effect). This effect can be disentangled from the RSD effects since the amplitude depends on the shape of the power spectrum. Section 1.3.2.4 describes how we can infer cosmological information from these parameters.

1.3.2 Cosmological probes

The description given in this chapter is supported by various cosmological observations. The following sections aim to give a brief overview of the current pillars of cosmology based on the most recent experiments and not to give a complete description of each field. It will deliberately omit certain aspects of observational cosmology such as weak-lensing, intensity mapping, galaxy clusters, ... This dissertation will be focused on galaxy survey whose standard measurements are described in Section 1.3.2.4.

1.3.2.1 Supernovae

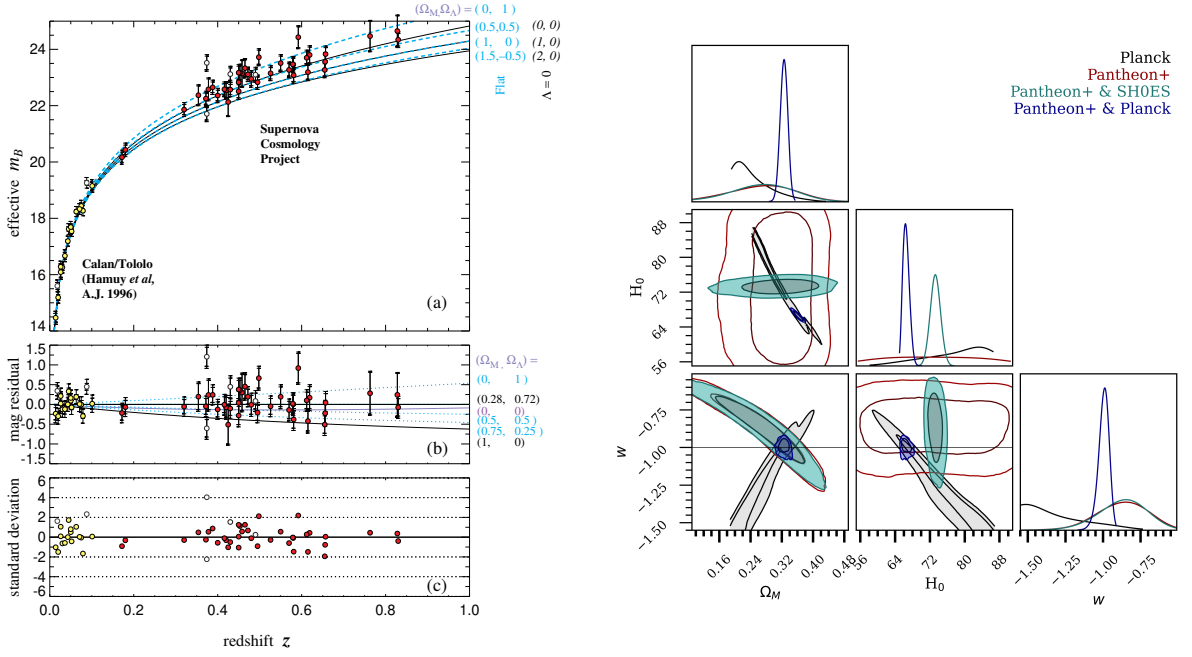
► **Discovery the dark energy** One cornerstone of cosmology is the discovery of the accelerated expansion of the Universe, which we attribute today to the presence of dark energy. This discovery was achieved at the end of the 20th century by two independent groups (Riess et al. 1998, Perlmutter et al. 1999), using type Ia supernovae and their magnitude-redshift diagram. The Supernova Cosmology Project[21] diagram is shown in Fig. 1.15a. The best fit with a flat ΛCDM model is highly in favor of the existence of dark energy $(\Omega_\Lambda, \Omega_m) = (0.72, 0.28)$, where[22] $\Omega_m = 0.28^{+0.09}_{-0.08}(\text{stat})^{+0.05}_{-0.04}(\text{sys})$ for statistical and systematic errors.

► **Measuring the Hubble constant** In addition to the detection of the accelerated expansion, supernovae can also be very useful in deriving the value of H_0 with the third rung of the *distance ladder* method led by the SHOES collaboration, as described in detail in Riess et al. (2022) (see Fig. 12 for the three rungs of the ladder). The ladder is necessary to determine the intrinsic supernovae magnitude and thus to measure the value of (H_0) in the magnitude-redshift diagram.

Fig. 1.15b shows the famous H_0 tension at a 5σ level between the local value of H_0 derived with the distance ladder method ($H_0 = 73.04 \pm 1.04$ km s^{-1} Mpc^{-1}) and the value inferred from the cosmic microwave background (CMB), as discussed in Section 1.3.2.2, measured by the Planck mission ($H_0 = 67.74 \pm 0.46$ km s^{-1} Mpc^{-1}). The fitted model in this figure is a flat cosmology with a free equation of state parameter w for the dark energy. Planck alone is not able to constrain this parameter, and the above value is given for fixed parameters $w = -1$. Current measurements are in favor of $w = -1$.

This 5σ tension is today the greatest disagreement between two independent cosmological measurements. Both measurements are well explained by the ΛCDM model, but not with the same parameters. A major community effort is underway to reconcile these two measurements, either by validating the CMB or the local measurement or by proposing a refinement to the model, as described in Di Valentino et al. (2021).

[21] https://supernova.lbl.gov/
[22] In a flat universe, neglecting the radiation contribution: $\Omega_m + \Omega_\Lambda = 1$.
[23] http://astro.vaporia.com/start/calantololosurvey.html



(a) **(b)**

Figure 1.15: (a) Magnitude-redshift diagram, also called Hubble diagram, of 42 high-redshift (*resp.* 18 low-redshift) type Ia supernovae from the Supernova Cosmology Project (*resp.* the Calàn/Tololo Supernova Survey[23]) published in Perlmutter et al. (1999). The best fit with ΛCDM model is obtained for $(\Omega_\Lambda, \Omega_m) = (0.72, 0.28)$. (b) Posteriors for flat wCDM Universe with Ω_m, H_0 and w (equation of state of Λ) as cosmological parameters (Brout et al. 2022). Planck is $TT, TE, EE+lowE$ likelihoods and $0.2 < \Omega_m < 0.4$ is set for computational speed. Pantheon+ does not include the distance ladder method and cannot fit correctly the value of H_0.

1.3.2.2 Cosmic microwave background

The discovery of the cosmic microwave background (CMB) by Penzias and Wilson in 1964 provided clear evidence of the expansion of the Universe. Since then, three dedicated space missions - COBE (Mather et al. 1990, Smoot et al. 1990), WMAP (Bennett et al. 2003; 2013, Hinshaw et al. 2013) and Planck (Tauber et al. 2010, Planck Collaboration et al. 2020;?) - have been launched to measure temperature fluctuations, or anisotropies, in the CMB. Currently, the Planck mission provides the most accurate measurements.

The ΛCDM model and perturbation theory, which are discussed in this chapter, can be described using only six parameters. As shown in Fig. 1.16, the angular power spectrum of the temperature measured by Planck is exactly reproduced by this six-parameter model. The blue line in the figure represents the best fit obtained with TT, TE, EE + low E + lensing likelihood, where TT (*resp.* TE, EE) are the angular power spectrum of the temperature (T) (*resp.* cross-power spectrum of t temperature and E mode polarization, power spectrum of the E mode polarization), low E includes the low multipoles of the angular power spectrum of the E mode. Finally, lensing adds the power spectrum of the lensing potential and the lensing field map itself. This is weak gravitational lensing. In the presence of mass, lines of sight in the Universe are slightly deflected, such that the foreground mass distribution can be detected by systematically aligning background sources around the lens mass.

This observation provides unparalleled validation of the theory described in this chapter. Table 1.2 gives the best constraint today on the six ΛCDM parameters using the TT, TE,

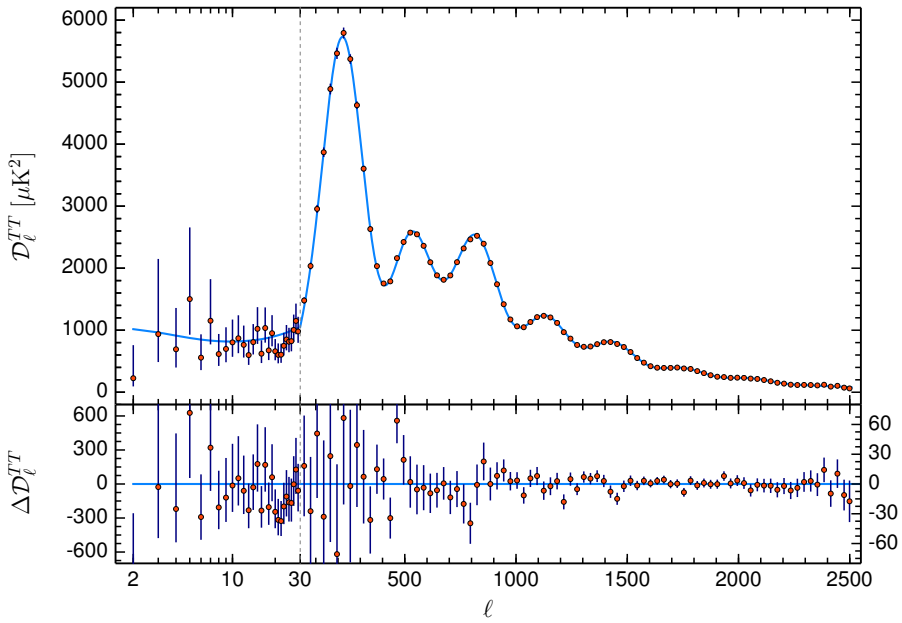


Figure 1.16: Planck18 temperature angular power spectrum at the top and the residuals to the best fit at the bottom (Planck Collaboration et al. 2020). Red points are the measurement and the error bars are 1σ uncertainties, containing cosmic variance and not the uncertainties in the foreground model for $\ell \geq 30$. The blue line is the base ΛCDM theoretical spectrum best fit to the Planck TT,TE,EE+lowE+lensing likelihoods. Note the difference of horizontal scale before (logarithmic) and after (linear) the dashed line and of vertical scale for the residual.

EE+lowE+lensing likelihood, and also some useful derived quantities. In this table, Ω_m includes the contribution from one neutrino with a mass of 0.06 eV, θ_{MC} (*resp.* θ_*) is an approximation to (*resp.* the full numerical result of) the acoustic scale angle.

The CMB continues to be the subject of significant study, particularly for the observation of the B-modes of the CMB polarization, which would be generated during inflation.

1.3.2.3 Big Bang nucleosynthesis

The Big Bang nucleosynthesis (BBN), also known as primordial nucleosynthesis, describes the formation of heavy nuclei during the first few dozen minutes of the Universe's history. This process produced the majority of deuterium (^{2}H), helium-3 (^{3}He), and helium-4 (^{4}He), as well as a small fraction of lithium, beryllium, and boron. Other elements in the Universe were created through stellar nucleosynthesis, which occurs later in the Universe's history and is still ongoing. Spallation is responsible for much of the current lithium, beryllium, and boron in the Universe.

The most significant reactions during primordial nucleosynthesis are displayed in Fig. 1.17a. The abundance of each element depends on the time it takes for lighter elements to form heavier ones. Additionally, the Universe expands and cools as time passes, making the formation of heavy nuclei impossible after a certain moment. Knowing the reaction rates, one can theoretically predict[24] the abundance of elements produced during primordial nucleosynthesis as a function of time or temperature. This can be computed using numerical codes such as `PArthENoPE`[25]

[24] See Pitrou et al. (2018) for a recent review and all the references therein.
[25] `https://parthenope.na.infn.it/`

Table 1.2: Base-ΛCDM cosmological parameters from Planck TT,TE,EE+lowE+lensing best fit. The errors are the 1σ uncertainties. ($k_{pivot} = 0.05$ Mpc^{-1}).

Parameter	TT,TE,EE+lowE+lensing
$\Omega_{\rm b}h^2$	0.02237 ± 0.00015
$\Omega_{\rm c}h^2$	0.1200 ± 0.0012
$100\theta_{\rm MC}$	1.04092 ± 0.00031
τ	0.0544 ± 0.0073
$\ln(10^{10}A_{\rm s})$	3.044 ± 0.014
$n_{\rm s}$	0.9649 ± 0.0042
$\Omega_{\rm m}h^2$	0.1430 ± 0.0011
H_0 [km s^{-1} Mpc^{-1}]	67.36 ± 0.54
$\Omega_{\rm m}$	0.3153 ± 0.0073
Age [Gyr]	13.797 ± 0.023
σ_8	0.8111 ± 0.0060
$S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$	0.832 ± 0.013
$z_{\rm re}$	7.67 ± 0.73
$100\theta_*$	1.04110 ± 0.00031
$r_{\rm drag}$ [Mpc]	147.09 ± 0.26

(Pisanti et al. 2008, Consiglio et al. 2018, Gariazzo et al. 2022) or `PRIMAT`[26] (Pitrou et al. 2018; 2020). Fig. 1.17b shows the abundance of the lightest elements, with deuterium, helium-3, and helium-4 being the most commonly produced during BBN.

The nuclear rates required to form heavier nuclei depend on the density of baryonic matter. The more baryons there are, the more interactions occur. Therefore, observing the primordial abundance of elements is a direct measurement of the amount of baryonic matter. Since (^{2}H, ^{3}He, ^{4}He) are mostly produced during primordial nucleosynthesis, they play a major role in determining the baryonic density (Aver et al. 2015, Cooke et al. 2018). Fig. 1.17c gives the latest constraints on $\Omega_b h^2$ from (Pitrou et al. 2020), with an inferred baryonic density being 1.8σ lower than the Planck measurement.

BBN plays a crucial role in cosmology since it can be combined with BAO scale measurements (see Fig. 1.19c) to provide a competitive and independent constraint on the ΛCDM parameters compared to Planck 18 (Planck Collaboration et al. 2020).

1.3.2.4 Galaxy survey

If the description of the Universe's history presented in this chapter is correct, the oscillations observed in the angular temperature spectrum at $z \sim 1100$, see Fig. 1.16, should be imprinted in the distribution of matter that we observe today through the galaxy distribution. The BAO peak (Fourier transform of the oscillations) was first detected in 2005 by the SDSS collaboration[27] in the luminous red galaxy (LRG) sample (Eisenstein et al. 2005), and then in 2007 in the main galaxy sample with a combined clustering analysis with 2dFGRS[28] (Percival et al. 2007). Fig. 1.18 shows the BAO peak detected for the time in 2005 with the SDSS LRG sample. This

[26] http://www2.iap.fr/users/pitrou/primat.htm
[27] https://classic.sdss.org/
[28] http://www.2dfgrs.net/

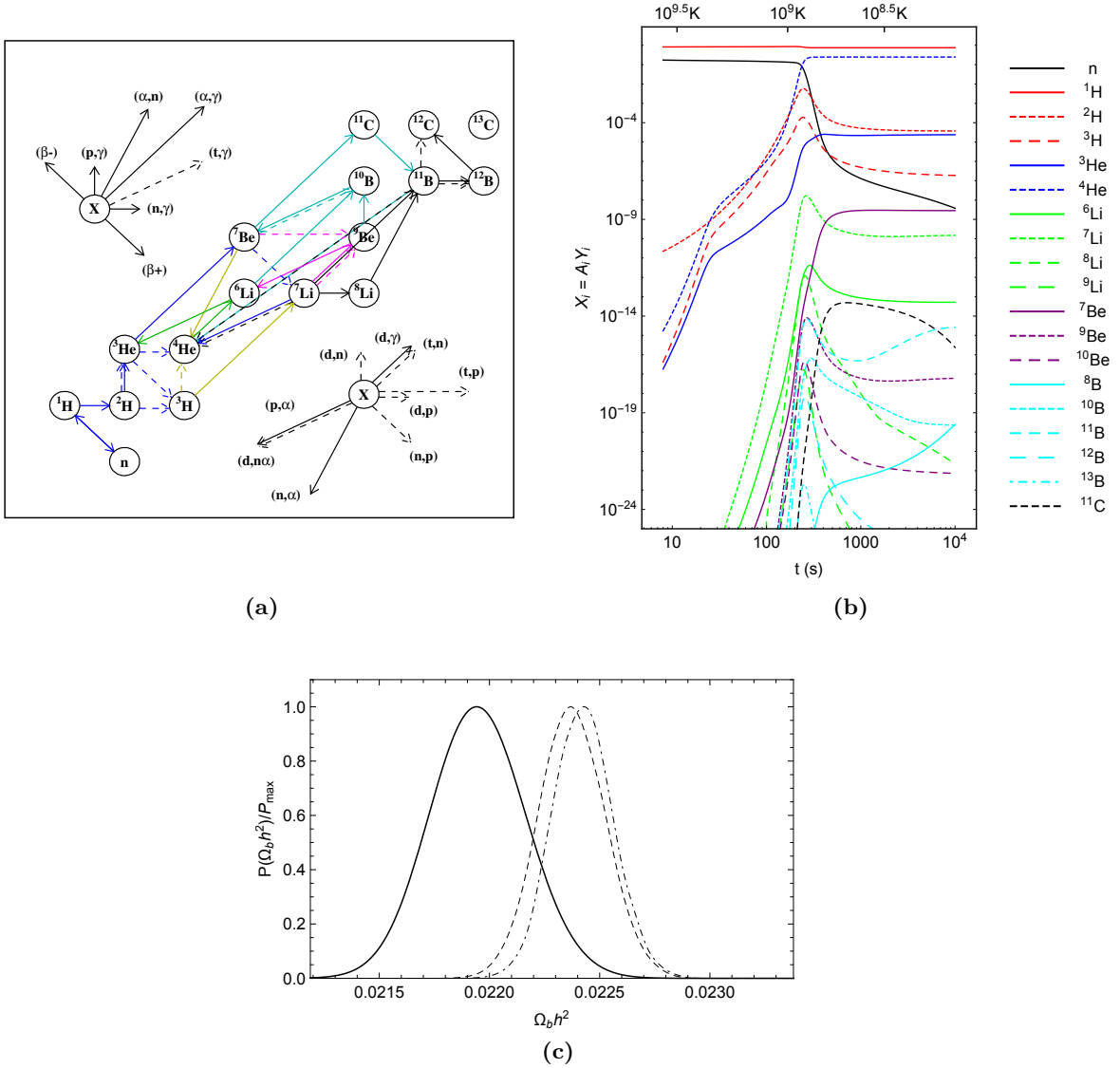


Figure 1.17: (a) Sketch of the most important reactions (on 424 available) during the BBN (Pitrou et al. 2018). In particular, it shows the production of deuterium, ^{3}He and ^{4}He which are mainly produced during the BBN. (b) Evolution of the lightest element abundances as a function of time/temperature (Pitrou et al. 2018). The current observations are focused on deuterium, ^{3}He, ^{4}He and ^{7}Li. (c) Inferred posterior distribution of baryon density ($\Omega_b h^2$) from BBN (full line), Planck (dashed line) and Planck+BAO (dashed-dotted line) (Pitrou et al. 2020). The difference between BBN and Planck is about 1.8σ and it is mainly due to the primordial abundance of deuterium, highlighting the importance of this probe.

detection opened the road for 20 years of sky mapping with SDSS through the BOSS (Dawson et al. 2013) / eBOSS (Dawson et al. 2016) programs and still today with the next survey generation as DESI (DESI Collaboration et al. 2016a) and Euclid (Scaramella et al. 2022).

The acoustic oscillation during radiation domination, see Section 1.2.3.4 and (1.94), imprints in the matter distribution a characteristic scale r_d which represents the distance that the baryons contrast density δ_b, under radiation pressure, have traveled while the two components were coupled through Compton scattering. r_d is the sound horizon at baryons/photons decoupling time z_d

$$r_d = \int_0^{t_d} \mathrm{d}t \frac{c_s(t)}{a(t)} = \int_{z_d}^{\infty} \mathrm{d}z \frac{c_s(z)}{H(z)}, \tag{1.133}$$

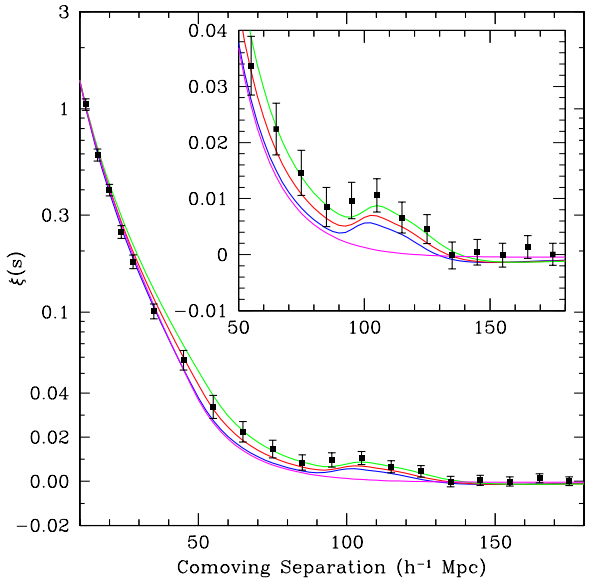


Figure 1.18: Redshift space correlation function of the SDSS luminous red galaxy sample (Eisenstein et al. 2005). The points are correlated together. The color lines are pure CDM models with $n_s = 0.98$ and $\Omega_b h^2 = 00.24$. The magenta line is without the acoustic peak. The others have different values of $\Omega_m h^2$. The bump, known as the BAO peak, at 100 h^{-1} Mpc is statistically significant, rejecting the Universe without BAO (magenta line) with confidence of $\sigma = 3.4$.

where $c_s = c\left[3\left(1 + 3\rho_b/4\rho_\gamma\right)\right]^{-1/2} \sim c/\sqrt{3}$ is the speed of sound in the baryons-photons fluid.

Since the decoupling between baryons and photons is not instantaneous and photons are much more numerous than baryons, the baryons/photons decoupling time z_d is slightly different than the "CMB time" $z_\star = 1089.80 \pm 0.21$ defined as the moment when photons have interacted on average with baryons for the last time. A fraction of the photons, which continue to oscillate, interact longer with the baryons and drag ($_d$) baryons a bit further. Therefore, z_d corresponds to the moment when the baryons stop interacting on average with the photons: $z_d = 1060.01 \pm 0.29$. Hence, the sound horizon is slightly different $r_* = 144.57 \pm 0.22$ Mpc and $r_d = 147.21 \pm 0.23$ Mpc. Numbers are from the best fit of Planck18 (Planck Collaboration et al. 2020). One can find a numerical estimation of r_d, see for instance equation 16 in Aubourg et al. (2015).

This BAO scale r_d can be observed with the preferred correlation in the galaxy distribution either in the LOS or transverse direction:

$$\Delta z = H(z) r_d, \quad \Delta\vartheta = \frac{r_d}{D_M(z)}, \tag{1.134}$$

where Δ_z is the redshift difference and $\Delta\vartheta$ is the angular separation of the preferred distance in the 2-point correlation. In a flat universe, the transverse comoving distance $D_M(z)$ is equal to the comoving distance $\chi(z)$. Hence, by setting the value of r_d either with the CMB or with the BBN, one can extract cosmological information at different redshifts.

Unfortunately, these two distances cannot be measured directly due to the AP effect, as described in Section 1.3.1.3. Besides, we need to generate a template of the power spectrum or correlation function in order to measure the position of the BAO peak. This template is computed from a given cosmology that will be supposed to be the fiducial one for practical reason[29]. Hence, the template value of r_d is then not exactly the correct one and we have to

[29] In the most general case, we can use another cosmology to compute the template, such that $r_d^{temp} = r_d^{fid}$.

add an additional factor to the AP parameters $\alpha\|, \alpha\perp$ (1.131):

$$\alpha_\perp \equiv \frac{\chi(z) r_d^{fid}}{\chi_{fid}(z) r_d}, \quad \alpha_\parallel \equiv \frac{H_{fid}(z) r_d^{fid}}{H(z) r_d}, \tag{1.135}$$

renaming generally the pure AP parameters as $q_\parallel, q_\perp$.

Sometimes, one can prefer to use the isotropic dilation factor:

$$\alpha_{\text{iso}} = \frac{D_V(z) r_d^{fid}}{D_V^{fid}(z) r_d}, \tag{1.136}$$

where D_V is the "spherically averaged" distance defined as $D_V(z) = \left[z\chi^2(z)c/H(z)\right]^{1/3}$.

Fig. 1.19a gives the 7 independent 2-point clustering measurements using different types of galaxies from the latest eBOSS data release DR16 (Alam et al. 2021), on which we can extract the $\alpha_\parallel, \alpha_\perp$. The middle line corresponds to the ratio between the isotropic measurement (for Ly-α, it is a combination of radial and transverse measurements) and the theoretical prediction assuming Planck18 parameters.

In Section 1.3.1.2, RSD was linked to the growth rate f via the Kaiser formula (1.115). However, f completely degenerates with the primordial amplitude A_s of the power spectrum used to compute P_{lin}. To be agnostic to A_s which is a free parameter in the ΛCDM model, the parameter σ_8 defined as

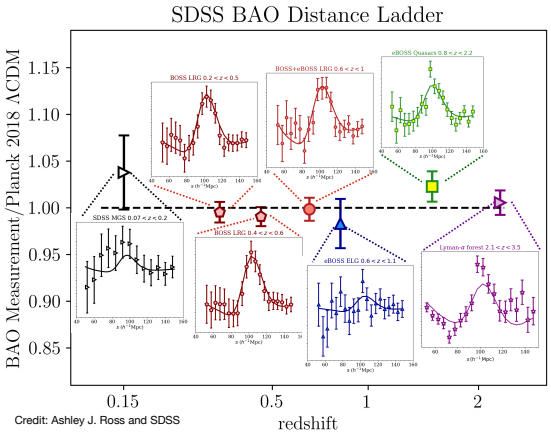
$$\sigma_8^2(z) = \frac{1}{2\pi^2} \int_0^\infty \mathrm{d}k k^2 P_{lin}^m(k, z) W^2(rk), \tag{1.137}$$

where W is the top-hat window function and $r = 8\ h^{-1}$Mpc. Hence, the Kaiser formula becomes,

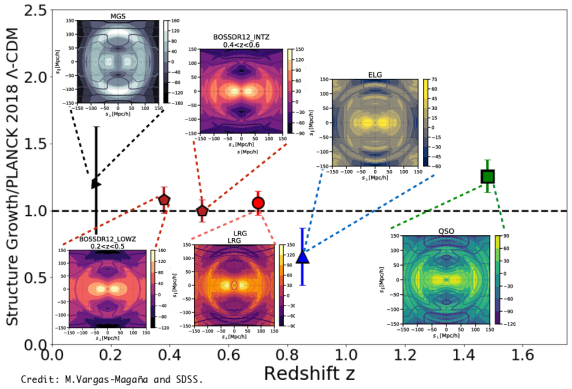
$$P(\mathbf{k}) = \left(1 + \beta\mu^2\right)^2 P_{lin}(k) = \left(\sigma_8 b_1 + \sigma_8 f \mu^2\right)^2 \frac{\tilde{P}_{lin}^m(k)}{\sigma_8^2}, \tag{1.138}$$

and is independent of A_s, but the growth rate is measured as $f\sigma_8$. Similarly to Fig. 1.19a, Fig. 1.19b shows the 2D redshift space function from eBOSS DR16 and the middle line the ratio between the measurement and the prediction from Planck18.

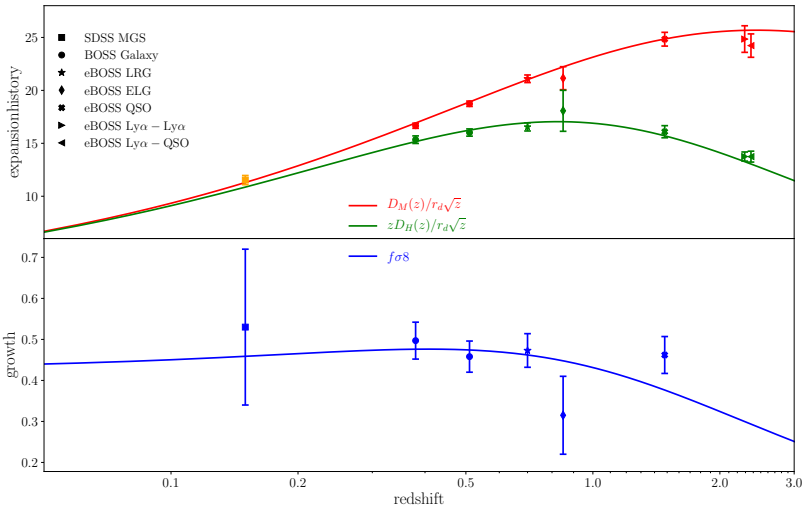
Finally, Fig. 1.19c sums up the eBOSS DR16 $\alpha_\parallel, \alpha_\perp, f\sigma_8$ measurements. The full lines correspond to the theoretical prediction assuming the Planck 18 values as ΛCDM parameters. Expect the value from the ELG, the BAO/RSD measurements are remarkably in agreement with the prediction. The aim of the new generation of galaxy surveys is to get similar sensitivity to the ΛCDM parameters than Planck by increasing the number of points and decreasing the error bars on this figure. See Alam et al. (2021) for additional figures and inferred cosmological constraints as the sum of neutrinos mass, alternative models to ΛCDM, etc ...



(a) BAO scale $\alpha_{\rm iso}$



(b) Growth rate f



(c) eBOSS DR16

Figure 1.19: (a) 2-point correlation functions for the 7 independent tracers covering different redshift ranges used in BOSS/eBOSS program. The middle line shows the ratio between the isotropic BAO measurement (1.136) (for Ly-α, it is a combination of radial and transverse measurements) and the prediction from Planck 18. Expect the ELG sample, 6 others are in perfect agreement. From A. Ross and SDSS collaboration. (b) Same as (a) but for the 2D redshift space power spectrum. From M. Vargas-Magaña and SDSS collaboration. (c) Distance measurements from SDSS DR16 BAO measurements as a function of the redshift (Alam et al. 2021). Red is for the transverse BAO, green is for the radial BAO and blue is for the growth rate. Full lines are the Planck 18 best-fit prediction for the flat ΛCDM model

Chapter 2

The early Universe

Contents

And if inflation is wrong, then God missed a good trick. But, of course, we've come across a lot of other good tricks that nature has decided not to use.

Jim Peebles, interview at Princeton (1994)

Although some problems in the Hot Big Bang model may emerge as the flatness, horizon and monopole problems, it seems that, on the whole, it correctly describes the Universe. In particular, the primordial nucleosynthesis, the discovery of the CMB and the theoretical description of its anisotropies are strong predictions of the simple flat ΛCDM model, see Section 1.3.2.

Hence, it is natural to try to complete this model instead of changing it completely. The most distant observation that we can make by collecting light comes from the CMB, and it remains upstream a whole part of the history of the Universe that is still unknown. A phase of accelerated expansion at the very beginning of the Universe, as mentioned in Section 1.1.3.3, could be the miracle cure for all our ills.

This chapter will first explain the mechanism of inflation and how this paradigm can also generate primordial density fluctuations that will be the seed of the large-scale structures of the Universe. We then turn our attention to the possible observables of inflation, in order to classify the various candidates for this paradigm. In particular, we will study the primordial non-gaussianity and the imprint left on galaxy clustering.

2.1 Inflation

2.1.1 The missing piece in the jigsaw puzzle

2.1.1.1 An accelerated expansion

The flatness and horizon problems, raised in Section 1.1.3.3, are both related to the fact that the *comoving Hubble radius* $(aH)^{-1}$ increases over time. The curvature density parameter (B.2) and the particle horizon (1.37) can be written as

$$\Omega(a) - 1 = -\Omega_K = \frac{K}{(aH)^2}, \qquad \chi(a) = \int_{\ln a_i}^{\ln a} \frac{\mathrm{d}\ln a}{aH}, \tag{2.1}$$

where a_i is the scale factor at the beginning of the Universe. Assuming the Universe consists of a single fluid with an equation of state $P = w\rho$, the comoving Hubble radius becomes

$$(aH)^{-1} = H_0^{-1} a^{(1+3w)/2}, \tag{2.2}$$

and it increases over time if $1 + 3w > 0$ that is the case for radiation or matter dominated area, see Table 1.1. Hence, any small space curvature in the early Universe ($K \neq 0$) is amplified throughout the history of the Universe. Additionally, since $(aH)^{-1}$ is a time-increasing function, the particle horizon at the recombination can be only increased by pushing back the beginning of the Universe t_i by integrating longer.

Both of these problems can be solved by adding a period to the early Universe in which the Hubble sphere $(aH)^{-1}$ shrinks over time. This means that the time derivate of the comoving Hubble radius should be negative

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{aH}\right) = -\frac{\ddot{a}}{(aH)^2} < 0 \quad \Longrightarrow \quad \ddot{a} > 0, \tag{2.3}$$

corresponding to a period of accelerated expansion, as first proposed by Guth (1981), Albrecht and Steinhardt (1982), Linde (1982a). As for the recent accelerated expansion, a fluid that can describe this period should respect $w < -1/3$ (2.2).

Let's introduce the *first slow roll parameter*, $\varepsilon \equiv -\dot{H}/H^2$, which verifies

$$\frac{\ddot{a}}{a} = H^2\,(1 - \varepsilon)\,. \tag{2.4}$$

The accelerated expansion ($\ddot{a} > 0$) implies that $\varepsilon < 1$. Assuming a flat universe with a single fluid component and using the Friedmann equations (1.28), ε can be related to w by

$$1 - \varepsilon = -(1 + 3w)/2. \tag{2.5}$$

If $\varepsilon \to 0$ *i.e.* the Hubble parameter varies slowly with time ($\dot{H} \simeq 0$), then $w \to -1$ such as this period is similar to a de Sitter Universe which consists of a Universe with only a cosmological constant Λ. Hence, the scale factor respects

$$a(t) \propto e^{Ht}, \tag{2.6}$$

and the expansion increases exponentially in a way that this period is called *inflation*.

2.1.1.2 With an appropriate energy

As mentioned in Section 1.1.3.3, if the Hot Big Bang starts at very high temperature, some unwanted relics can be produced when the Universe cools as it expands. To circumvent the production of these relics, all we need to ensure is that the amount of energy at the beginning of the Hot Big Bang is low enough for the initial temperature to be below that of the phase transition. Typically, the monopole problem will be solved requiring a temperature below the Grand Unification scale ($\sim 10^{16}$ GeV).

Nevertheless, the observed quantities of helium and deuterium in the Universe can only be explained by the primordial nucleosynthesis, see Section 1.3.2.3, *i.e.* the temperature should be higher than 1 MeV.

Finally, from the current observations, the initial amount of energy should produce a temperature such as

$$1 \text{ MeV} < T_{\text{reh}} < 10^{16} \text{ GeV}. \tag{2.7}$$

2.1.1.3 For a sufficiently long period

To solve the two main issues mentioned above, inflation should exist over a sufficiently long period. The length of the inflation is controlled by the number of e-folds N defined by

$$N \equiv \ln\left(\frac{a_e}{a_i}\right) = \int_{t_i}^{t_e} \mathrm{d}t H(t), \tag{2.8}$$

where a_e (*resp.* a_i) is the scale factor at the end (*resp.* beginning) of the inflation. N corresponds to the growth of the scale factor during the inflation, expressed in the exponential basis e.

From (B.2) and neglecting Ω_{K0},

$$\left|\frac{\Omega_K(t_e)}{\Omega_K(t_i)}\right| \simeq \left(\frac{a_e}{a_i}\right)^{-2} = \mathrm{e}^{-2N}. \tag{2.9}$$

The curvature at Planck time is $|\Omega_K(t_{\text{pl}})| < 10^{-60}$, see Section 1.1.3.3. Then, to have $|\Omega_K(t_i)| \sim \mathcal{O}(1)$ and inflation that stops at Planck time, the number of e-fold should be

$$N \geq 70. \tag{2.10}$$

From (2.2), the particle horizon can be calculated analytically

$$\chi(a) = \frac{2H_0^{-1}}{(1+3w)}\left[a^{(1+3w)/2} - a_i^{(1+3w)/2}\right] \approx \frac{2}{(1+3w)}(aH)^{-1}, \tag{2.11}$$

where the early time contribution can be neglected for $w > -1/3$. Hence, to solve the horizon problem, the observable Universe should be causally connected at least at the recombination as observed by the CMB experiment. Therefore, at the very least, we need

$$(aH)^{-1}(t_{\text{rec}}) < (aH)^{-1}(t_i). \tag{2.12}$$

Neglecting the relative recent matter and Dark energy contribution, the Universe can be assumed to be in the radiation-dominated area since the end of the inflation. Hence, with $w = 1/3$,

$H \propto a^{-2}$ (see Table 1.1), one can link the end of the inflation and the recombination through

$$\frac{aH(t_{\rm rec})}{aH(t_e)} = \frac{a(t_e)}{a(t_{\rm rec})} = \frac{T(t_{\rm rec})}{T(t_e)}^{1} \sim \frac{3000\ K}{10^{16}\ \text{GeV}} \sim 10^{-26}, \tag{2.13}$$

where 1 eV $= k_B/1.6 \times 10^{19}$ K and the end of the inflation t_e is chosen to be at the Grand Unification scale ($T \sim 10^{16}$ GeV), see Section 2.1.1.2. Finally, (2.12) requires that

$$(a_i H_i)^{-1} > (a_{\rm rec} H_{\rm rec})^{-1} \sim 10^{26} (a_e H_e)^{-1} \quad \Rightarrow \quad N = \ln\left(\frac{a_e}{a_i}\right) \geq 60, \tag{2.14}$$

where the Hubble parameter is supposed constant during inflation.

To achieve this long period of inflation, ε should remain small for a sufficiently large number of Hubble times

$$|\eta| < 1, \quad \text{with } \eta \equiv \frac{d\ln\varepsilon}{dN} = \frac{\dot{\varepsilon}}{H\varepsilon}. \tag{2.15}$$

η is known as the *second slow roll parameter*.

Finally, having inflation in the early Universe solves the two major issues of the Hot Big Bang model. It remains to derive physical mechanisms which can lead to an inflationary phase and to connect them with the Hot Big Bang model. First, let's describe briefly the Lagrangian mechanics which will be the standard approach in the following.

2.1.2 Lagrangian mechanics

In 1780, Joseph-Louis Lagrange introduced a concise and elegant description of classical mechanics through the use of the *stationary-action principle*. This description can easily be adapted to the theory of field whether it is non-relativistic, relativistic or even quantum case (Landau and Lifshitz 1976; 1975). In this description, a physical system is depicted by its Lagrangian $\mathcal{L}$. The temporal evolution of the system is given by its action $\mathcal{S}$ defined as the temporal integral of the Lagrangian of the system. The equations of motion are then obtained thanks to the principle of least action which states $\delta\mathcal{S} = 0$ and is known as the *Euler–Lagrange equations*.

In a manner completely equivalent to the description given in Section 1.1.1, the Einstein equations (1.8) can be derived from the stationary-action principle. Einstein and Hilbert proposed to use the scalar curvature R as Lagrangian to depict the metric behavior:

$$\mathcal{S}_{\rm EH} = \frac{c^4}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}R, \tag{2.16}$$

where the multiplicative factor is set to recover the classical mechanics in the Newtonian limit and g is the determinant of the metric in which the coordinates x are written such as $g = \det(g_{\mu\nu})$.

Adding the dark energy component and the Lagrangian of matter $\mathcal{L}_m$, the action becomes

$$\mathcal{S} = \mathcal{S}_{\rm EH} + \mathcal{S}_\Lambda + \mathcal{S}_m = \frac{c^4}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}(R - 2\Lambda) + \int \mathrm{d}^4x\sqrt{-g}\mathcal{L}_m, \tag{2.17}$$

[1] The *Stephan-Boltzmann law* ensures that the energy density $u \propto T^4$. In an expanding Universe, $u \propto a^{-4}$. Hence, during the radiation-dominated area, the Universe is in thermal equilibrium and its temperature evolves as $T \propto a^{-1}$. Note that after the recombination, matter and radiation are decoupled and evolve independently. The temperature of the remaining photons is still $T_\gamma \propto a^{-1}$. For matter, one can assume an ideal gas. Since the expansion of the Universe is an adiabatic process, the *Laplace law* ensures that $PV^\gamma = const.$ where γ is the adiabatic index, and since there is no matter creation, $P \propto \rho^\gamma$ during the expansion. Similarly, from the *ideal gaz law*, $P \propto T \times \rho$. In an expanding Universe, $\rho \propto a^{-3}$ and therefore the temperature of the matter evolves as $T_m \propto a^{-3\gamma+3}$. Typically, for a monoatomic gaz, $\gamma = 5/3$ and $T_m \propto a^{-2}$.

and $\delta\mathcal{S} = 0$[2] ensures that

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{2.18}$$

where the stress–energy tensor is defined from the Lagrangian of matter $\mathcal{L}_m$ as

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}} \frac{\delta\sqrt{-g}\mathcal{L}_m}{\delta g^{\mu\nu}} = -2\frac{\delta\mathcal{L}_m}{\delta g^{\mu\nu}} + g_{\mu\nu}\mathcal{L}_m. \tag{2.19}$$

Remark: Here, there is only one field, the $g_{\mu\nu}$ metric, and therefore only one set of Euler-Lagrange equations that describe its dynamics. In general, there are as many Euler–Lagrange equations as there are different fields to describe each of their dynamics.

This formalism is very useful to test various extensions or modifications of the theory of gravity since it provides a simple formalism to derive the Einstein equations. Further, the same formalism is used in quantum field theory and many similarities will appear in the following.

Remark: One can introduce the Planck mass as $M_{\rm pl} = \sqrt{\frac{\hbar c}{8\pi G}} \approx 2.43 \times 10^{18}$ Gev[3] to simplify the above actions. Hence, with $\hbar = 1$ and $c = 1$, (2.16) becomes

$$\mathcal{S}_{\rm EH} = \frac{M_{\rm pl}^2}{2} \int \mathrm{d}^4x\sqrt{-g}R.$$

2.1.3 Single field slow-roll inflation

The simplest model to describe an inflation phase that respects the conditions mentioned above is to postulate the existence of an additional scalar field $\phi(t, \mathbf{x})$, called *inflaton*, and its associated potential energy $V(\phi)$. If the field varies with time, it will also carry kinetic energy. Once the stress-energy associated with inflaton dominates, the inflaton dynamics will source the evolution of the background metric of the Universe.

Remark: Although no scalar field had been observed at that time, the use of a scalar field was appealing because its use a few years earlier led to the creation of a mechanism that provides masses to particles in the standard model (Higgs 1964, Englert and Brout 1964). 50 years later, the Higgs boson was detected, which was the first observed manifestation of a scalar field (ATLAS Collaboration et al. 2012, CMS Collaboration et al. 2012).

2.1.3.1 Source of inflation

Using the simple form for the kinetic term, called the canonical kinetic term, the Lagrangian for a scalar field reads

$$\mathcal{L}_\phi = -\frac{1}{2} g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi). \tag{2.20}$$

The associated stress-energy tensor is

$$T_{\mu\nu} \equiv -\frac{2}{\sqrt{-g}} \frac{\delta\sqrt{-g}\mathcal{L}_\phi}{\delta g^{\mu\nu}} = {}^4\partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left[\frac{1}{2} g^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + V(\phi)\right]. \tag{2.21}$$

[2]Here, δ means the functional derivative as a function of $g_{\mu\nu}$ and should be written $\frac{\delta\mathcal{S}}{\delta g_{\mu\nu}}$ if other fields are present.

[3]The mass are quoted in Gev/c^2, but here $c = 1$.

[4]With the *Jacobi formula* $\delta g = g g^{\mu\nu}\delta g_{\mu\nu}$, we have $\delta\sqrt{-g} = \sqrt{-g} g^{\mu\nu}\delta g_{\mu\nu}/2$, and from $g_{\mu\alpha}g^{\alpha\nu} = \delta_\mu^\nu$, we have $\delta g_{\mu\nu} = -g_{\mu\alpha}g_{\nu\beta}\delta g^{\alpha\beta}$.

By analogy to a fluid, one can define the associated density and pressure to the scalar field where they are defined in the Minkowski space by $\rho \equiv T^{00}$ and $P \equiv 1/3(T^{11} + T^{22} + T^{33})$. Hence,

$$\begin{aligned} \rho_\phi &= \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi) = \frac{1}{2}\dot{\phi}^2 + V(\phi) \\ P_\phi &= \frac{1}{2}\dot{\phi}^2 - \frac{1}{6}(\nabla\phi)^2 - V(\phi) = \frac{1}{2}\dot{\phi}^2 - V(\phi) \end{aligned}, \tag{2.22}$$

where ϕ is assumed to be homogeneous such as $\partial_i\phi = 0$. This is consistent with the fact that in the following we will use the FLRW metric that describes a homogeneous Universe. In this case, the associated equation of state for the scalar field is given by

$$w_\phi = \frac{\dot{\phi}^2 - 2V(\phi)}{\dot{\phi}^2 + 2V(\phi)}. \tag{2.23}$$

If the potential term $V(\phi)$ dominates *i.e.* $V(\phi) \gg \dot{\phi}^2$ then $w \to -1$ and the inflation is close to a de Sitter Universe. When the kinetic term dominates, $w \to 1$ and the inflation stops.

At early time, the Universe should stay homogeneous and can be assumed flat as it can still be described by the FLRW metric. In addition, the matter, radiation and dark energy contribution can be neglected. The dynamics of the Universe is then obtained by the Euler-Lagrange equations of

$$\mathcal{S} = \mathcal{S}_{\rm EH} + \mathcal{S}_\phi = \frac{M_{\rm pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}R + \int d^4x\sqrt{-g}\mathcal{L}_\phi. \tag{2.24}$$

The first set of equations, $\delta\mathcal{S}/\delta g_{\mu\nu} = 0$, leads to the Friedmann equations (1.28) with fluid quantities given by ρ_ϕ and P_ϕ:

$$\begin{aligned} H^2 &= \frac{1}{3M_{\rm pl}^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right) \\ \dot{H} &= -\frac{1}{2M_{\rm pl}^2}\dot{\phi}^2 \end{aligned}, \tag{2.25}$$

where the second Friedmann equation was simplified using the first one[5].

The second set, $\delta\mathcal{S}/\delta\phi = 0$, gives the dynamics of ϕ. Since ϕ and $g_{\mu\nu}$ are only coupled through $\sqrt{-g}$, $\delta\mathcal{S}/\delta\phi = \delta\mathcal{S}_\phi/\delta\phi$, and ϕ will satisfy the well-known *Klein-Gordon equation*

$$\Box\phi + V'(\phi) = 0, \tag{2.26}$$

where $V' = \frac{\mathrm{d}V}{\mathrm{d}\phi}$ and $\Box$ is the *d'Alembertian* operator

$$\Box\phi \equiv \frac{1}{\sqrt{-g}}\partial_\nu\left(\sqrt{-g}g^{\mu\nu}\partial_\mu\phi\right). \tag{2.27}$$

Finally, with $\sqrt{-g} = a^3$, ϕ is ruled by

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \tag{2.28}$$

This is the traditional equation of motion with force derived from a potential ($V'(\phi)$) with friction ($3H\dot{\phi}$).

[5] $\dot{H} = \ddot{a}/a - (\dot{a}/a)^2$

2.1.3.2 How does inflation work?

Initially, the ϕ field will have to be dominated by its potential energy and this for a sufficiently long time for inflation to take place (2.23). Then, it is sufficient to find a potential which according to (2.28) can accelerate the field enough so that its kinetic energy becomes predominant compared to the potential energy. Inflation then naturally stops at this moment (2.23).

A potential that makes all this description possible is shown in Fig. 2.1. Inflation occurs in the shaded parts of the potential. To ensure that inflation lasts long enough, the potential energy must predominate throughout the entire period *i.e.* the potential must not accelerate too much the field during inflation. This is possible if the derivative of the potential is close to zero (2.28). This period is legitimately called *slow-roll inflation*. Then, the field rolls slowly along the potential before rolling into the well and oscillating there as it heats up. This is the *reheating* period and the inflation stops.

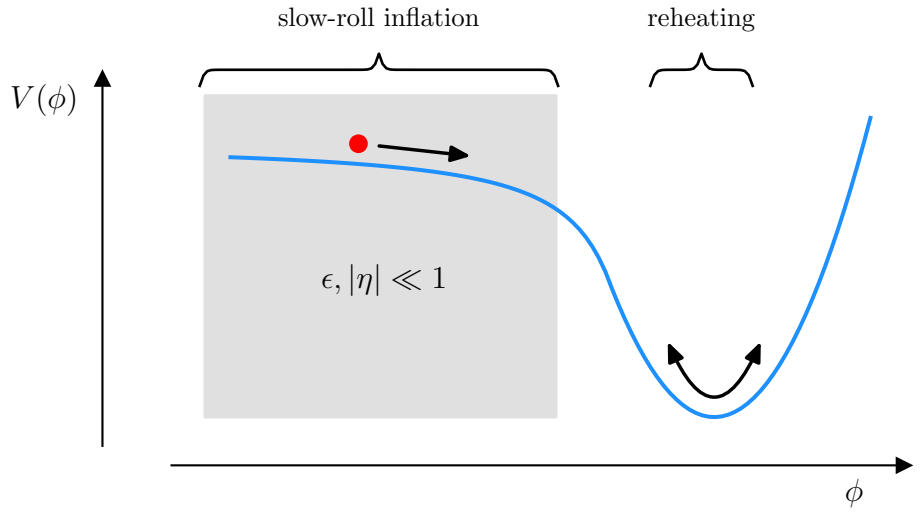


Figure 2.1: Example of potential V which respects the slow-roll approximation. Inflation starts at the red points and occurs in all the grey-shaded areas. Then, the field rolls slowly along the potential before rolling into the well and oscillating there as it heats up. Inflation stops at this point.

► **Slow-roll approximation** In (2.28), the friction term tends to slow down ϕ, and the slow-roll inflation can be easily reach if

$$\dot{\phi}^2 \ll V(\phi), \quad |\ddot{\phi}| \ll |3H\dot{\phi}|. \tag{2.29}$$

From (2.28) and (2.25), these slow-roll conditions lead to

$$\begin{aligned} 3H\dot{\phi} &= -V'(\phi) \\ H^2 &= \frac{V(\phi)}{3M_{\rm pl}^2}, \end{aligned} \tag{2.30}$$

and the slow-roll parameters can be written as

$$\varepsilon(\phi) = \frac{M_{\rm pl}^2}{2}\left(\frac{V'(\phi)}{V(\phi)}\right)^2, \quad \eta(\phi) = M_{\rm pl}^2\frac{V''(\phi)}{V(\phi)}. \tag{2.31}$$

Hence, the slow-roll conditions (2.29) imply

$$\varepsilon \ll 1, \quad |\eta| \ll 1 \tag{2.32}$$

and the condition mentioned in Section 2.1.1 are respected. Note that the conditions on ε, η are only conditions on the shape of the considered potential V so that it may have a period in which the slow-roll approximation is correct. This is a necessary but not sufficient condition to have a slow-roll period. Furthermore, the slow-roll approximation guarantees inflation, but it is not a necessary condition. It is perfectly possible to produce inflation without a slow-roll period.

► **Toy model** The simplest model of inflation is described by single field inflation driven by a mass term,

$$V(\phi) = \frac{1}{2} m^2 \phi^2. \tag{2.33}$$

This potential is displayed in Fig. 2.2 and leads to the following slow-roll parameters

$$\varepsilon = \eta = 2 \left(\frac{M_{\rm pl}}{\phi} \right)^2. \tag{2.34}$$

The slow-roll regions ($\varepsilon, |\eta| \ll 1$) should be such as $\phi^2 \gg 2M_{\rm pl}^2$. The condition is represented by the gray shaded areas in Fig.2.2. However, this condition is not sufficient to have a slow-roll regime. The slow-roll existence will depend on the initial value of the inflaton $(\phi, \dot{\phi})$.

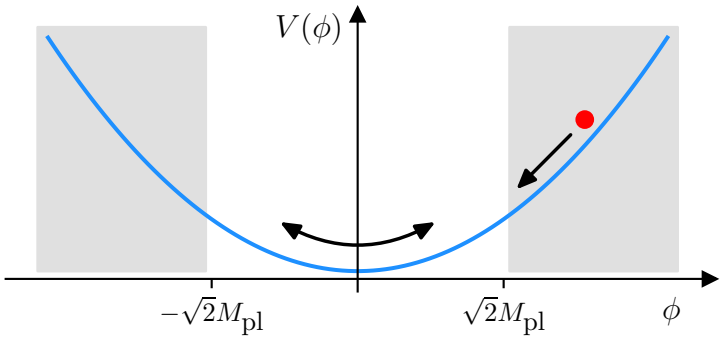


Figure 2.2: Potential for a single field inflation driven by a mass term. The gray shaded areas represent where the slow-roll conditions can be respected. At the bottom of the well, the oscillation during the reheating period is familiar to the well-known harmonic oscillator.

► **Initial conditions** Since (2.28) is a second order differential equation, it is completely determined by the initial conditions $(\phi_i, \dot{\phi}_i)$. Even if ϕ_i is in the possible slow-roll region, the slow-roll approximation may not hold. Fortunately, the solution in the slow-roll approximation given by (2.30) is an attractor for the full solution given by (2.25) and (2.28) *i.e.* whatever the initial conditions $(\phi_i, \dot{\phi}_i)$ the solution will converge to the slow-roll solution.

For instance, (2.28) can be written for our toy model as

$$\dot{\phi} \times \frac{{\rm d}\dot{\phi}}{{\rm d}\phi} = -\sqrt{\frac{3}{2}} \frac{1}{M_{\rm pl}^2} \left(\dot{\phi}^2 + m^2 \phi^2 \right)^{1/2} \dot{\phi} + m^2 \phi. \tag{2.35}$$

This is a first-order differential equation in ϕ which can be easily solved numerically. Fig 2.3 shows in the $(\phi, \dot{\phi})$ plane, solutions for several initial conditions. Most solutions will converge

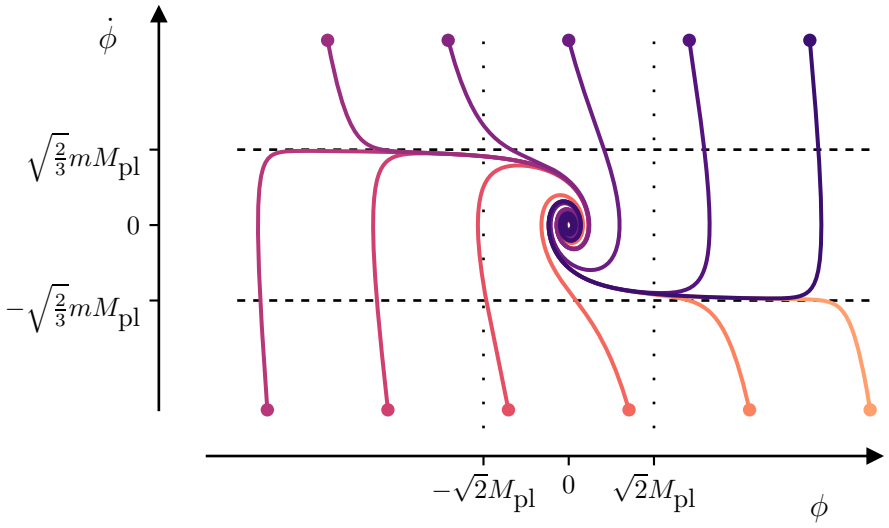


Figure 2.3: Phase portrait for several initial conditions in a single field inflation driven by a mass term. Before rotating around $(0,0)$ (reheating phase), most solutions will first converge towards $\dot{\phi} = \pm\sqrt{2/3}mM_{\rm pl}$ (slow-roll phase). This illustrates the attraction toward the slow-roll solution.

towards $\dot{\phi} = \pm\sqrt{2/3}mM_{\rm pl}$ and to the same slow-roll dynamics. Then, they will fall into the well of the potential V, rotating around $(0,0)$ and signaling the end of the inflation.

The initial conditions cannot be completely arbitrary since they must be within the *basin of attraction* of the slow-roll solution. However, if the basin of attraction is large enough, the Universe without any particular initial conditions for $(\phi, \dot{\phi})$ can reach itself the basin so that if the inflaton is the major contribution to the total energy, the Universe starts to inflate. The Universe then becomes very flat and homogeneous, and its dynamic is governed by the slow-roll solution. Thus, inflation erases the memory of the initial conditions and the history of the Universe is entirely described by the shape of the V potential.

► **Reheating: end of the inflation** In this description, all the energy in the Universe during the inflation comes from the potential energy $V(\phi)$ ($\dot{\phi}^2 \ll V(\phi)$). At the end of the inflation, the inflaton falls into a potential well and starts to oscillate. Its energy is then yielded to particles produced at this moment, known as the *reheating* period, producing all the content of the observable Universe. For instance, if inflaton is decaying into fermions (Linde 1982b, Abbott et al. 1982), the energy density evolves as

$$\dot{\rho}_\phi + (3H + \Gamma_\phi)\,\rho_\phi = 0, \tag{2.36}$$

where Γ_ϕ is a phenomenological decay term.

As described in Section 2.1.1.2, the temperature of the Universe after the end of the inflation should be large enough to produce the standard Big-Bang nucleosynthesis and sufficiently low to not produce *unwanted relics*. Since all the energy in the Universe is from the potential energy of the inflaton, (2.7) will constrain the amplitude of the potential V. For instance, in our toy model, the slow-roll regime breaks when $\phi_e \approx \sqrt{2}M_{\rm pl}$, and the total available energy is then $V(\phi_e) = m^2 M_{\rm pl}^2$. Hence,

$$\begin{aligned} &(1\text{ MeV})^4 < m^2 M_{\rm pl}^2 < \left(10^{16}\text{ GeV}\right)^4 \\ \Rightarrow\quad &4\cdot 10^{-22}\text{ MeV} < m < 4\cdot 10^{13}\text{ GeV} \end{aligned}. \tag{2.37}$$

This very wide range offers only minimal constraints on the mass m of the inflaton. For comparison, the Higgs field, the only scalar field observed in the Universe, has mass of 125.25 ± 0.17 GeV.

► **Duration of the inflation** As seen in Section 2.1.1, the duration of inflation is quantified by the number of e-fold N and $N \geq 60$ is needed to solve the horizon problem. In the slow-roll approximation and using (2.30), this number becomes

$$N(\phi) = \int_{t_i}^{t_e} H(t)\mathrm{d}t = \int_{\phi_i}^{\phi_e} \frac{H}{\dot{\phi}}\mathrm{d}\phi \approx \frac{1}{M_{\mathrm{pl}}^2}\int_{\phi_e}^{\phi_i} \frac{V(\phi)}{V'(\phi)}\mathrm{d}\phi. \tag{2.38}$$

In our toy model, it is

$$N(\phi) = \frac{1}{M_{\mathrm{pl}}^2}\int_{\phi_e}^{\phi_i} \frac{\phi}{2}\mathrm{d}\phi = \frac{1}{4M_{\mathrm{pl}}^2}\phi_i^2 - \frac{1}{2}, \tag{2.39}$$

where $\phi_e \approx \sqrt{2}M_{\mathrm{pl}}$.

For instance, one can assume that the largest possible initial value of ϕ will lead to the Planck density such as $V(\phi_i) = M_{\mathrm{pl}}^4$, and so $N(\phi) \approx M_{\mathrm{pl}}^2/2m^2$. Even with the highest mass $m \sim 4 \cdot 10^{16}$ GeV, the number of e-folds $N \sim 10^{11}$ is very large *i.e.* we need to wait a very long time with this model to recover an acceptable temperature when the initial value of ϕ leads to the Planck density.

2.1.4 Origin of primordial fluctuations

The success of inflation is not only due to the resolution of the flatness, horizon and unwanted relics problems but mainly because inflation is an elegant mechanism to explain the origin of the metric fluctuations (Hawking 1982, Starobinsky 1982, Guth and Pi 1982, Bardeen et al. 1983). The origin of the perturbations will be understood as quantum-mechanical vacuum fluctuations during inflation. Inflation predicts the production of scalar and tensor perturbations and no vector perturbations.

2.1.4.1 Perturbed equations

As in Section 1.2.2.1, the scalar field ϕ will de decomposed as

$$\phi(t, \mathbf{x}) = \phi_0(t) + \delta\phi(t, \mathbf{x}), \tag{2.40}$$

where ϕ_0 is described by (2.25) and (2.28). Note that in the following, $'$ will refer to the conformal time derivative and not to the derivate with respect to ϕ as above, and for simplicity, V will correspond to $V(\phi_0)$. Furthermore, in conformal time, Friedmann equations (2.25) are

$$\begin{aligned} \mathcal{H}^2 &= \frac{1}{3M_{\mathrm{pl}}^2}\left(\frac{1}{2}\phi_0'^2 + a^2V\right) \\ \mathcal{H}' &= -\frac{1}{3M_{\mathrm{pl}}^2}\left(\phi_0'^2 - a^2V\right) \end{aligned}. \tag{2.41}$$

We will work with the Newtonian gauge in similarity to the previous chapter, although this is not the usual gauge for solving perturbations during inflation. The following equations are then obtained as in section 1.2.2.2. In particular, the perturbed stress-energy tensor for a scalar field is written

$$\begin{aligned} \delta T_{\mu\nu} = {} & 2\partial_\mu\delta\phi\partial_\nu\phi_0 - \delta g_{\mu\nu}\left[\frac{1}{2}g^{\alpha\beta}\partial_\alpha\phi_0\partial_\beta\phi_0 + V\right] \\ & - g_{\mu\nu}\left[\frac{1}{2}\delta g^{\alpha\beta}\partial_\alpha\phi_0\partial_\beta\phi_0 + g^{\alpha\beta}\partial_\alpha\phi_0\partial_\beta\delta\phi + \frac{\partial V}{\partial\phi}\delta\phi\right]. \end{aligned} \tag{2.42}$$

In the Newtonian gauge, it becomes

$$\begin{aligned}\delta T_{00} &= \phi_0'\delta\phi' + 2a^2V\Phi + a^2\frac{\partial V}{\partial\phi}\delta\phi \\ \delta T_{0i} &= (\partial_i\delta\phi)\cdot\phi_0' \\ \delta T_{ij} &= \left[\phi_0'\delta\phi' - (\Psi+\Phi){\phi_0'}^2 + 2a^2V\Psi - a^2\frac{\partial V}{\partial\phi}\delta\phi\right]\delta_{ij}\end{aligned} \quad . \tag{2.43}$$

► **Scalar modes** Since (2.43) does not have any anisotropic pressure[6], $\Phi = \Psi$, and the other Einstein equations are

$$\begin{aligned}\Phi' + \mathcal{H}\Phi &= \frac{1}{2M_{\rm Pl}}\phi_0'\delta\phi \\ k^2\Phi &= -\frac{1}{2M_{\rm Pl}^2}\left[{\phi_0'}^2\Phi + \phi_0'\delta\phi' + a^2\frac{\partial V}{\partial\phi}\delta\phi + 3\mathcal{H}\phi_0'\delta\phi\right]\end{aligned}, \tag{2.44}$$

and the dynamics of $\delta\phi$ is given by the perturbed Klein-Gordon equation

$$\delta\phi'' + 2\mathcal{H}\delta\phi' + k^2\delta\rho = -2a^2\frac{\partial V}{\partial\phi}\Phi + 4\phi_0'\Phi' - a^2\frac{\partial^2 V}{\partial\phi^2}\delta\rho. \tag{2.45}$$

► **Vector modes** Scalar field (2.43) does not have any vector components. Therefore, there is no associated Klein-Gordon equation for the vector mode and the only Einstein equations are

$$E_i' + 2\mathcal{H}E_i = 0, \tag{2.46}$$

such that the vector modes from the metric follow $E_i \propto a^{-2}$. Hence, for any model of inflation based on a scalar field, vector modes have disappeared at the end of the inflation.

► **Tensor modes** As for the vector modes, scalar field (2.43) does not have any tensor components and the equations for the metric tensor modes are,

$$E_{ij}'' + 2\mathcal{H}E_{ij}' - \Delta E_{ij} = 0. \tag{2.47}$$

The solution, see Appendix 4.3, $E_{+,\times}$ are the primordial gravitational waves.

2.1.4.2 From quantum perturbations to metric perturbations

Since the comoving curvature perturbation $\mathcal{R}$ (A.38) is constant under an adiabatic evolution for super-Hubble scales, the initial conditions will be given by the amount of perturbations when the mode k crosses the Hubble radius during the inflation. Fig. 2.4 shows the history of a mode k from the inflation period until today.

During the inflation, the Universe is dominated by a scalar field and the comoving curvature perturbation (A.38) can be written as[7]

$$\mathcal{R} = -\Phi - \frac{\mathcal{H}\delta\phi}{\phi_0'}. \tag{2.48}$$

[6] $\delta T_i^j \propto \delta_i^j$

[7] $\mathcal{R}$ is a gauge invariant quantity, and it is, therefore, sufficient to determine the value of $v+B$ as a function of ϕ in the Newtonian gauge. The perturbed stress-energy tensor for a perfect fluid (A.2) gives $\delta T_{0i} = -a^2(\bar\rho+\bar P)\partial_i v$ while the one for a scalar field (2.43) gives $\delta T_{0i} = (\partial_i\delta\phi)\cdot\phi_0'$. Finally, with $\bar\rho+\bar P = {\phi_0'}^2/a^2$, we have $v+B = -\delta\phi/\phi_0'$.

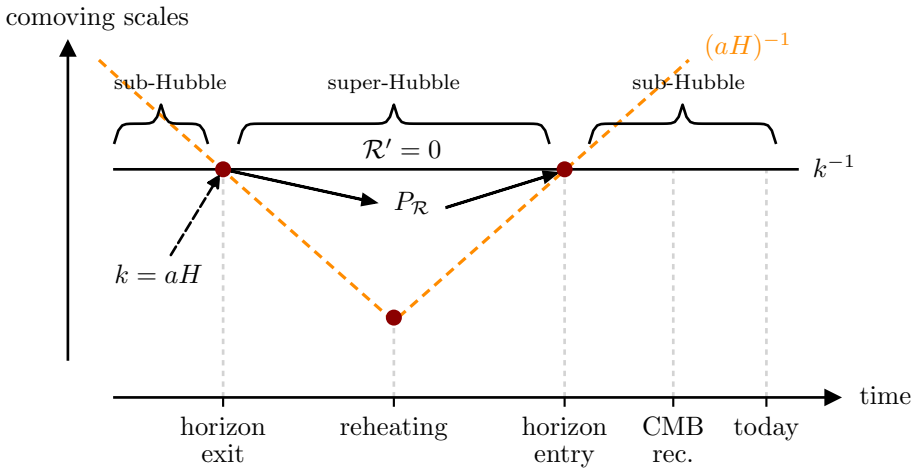


Figure 2.4: Perturbation dynamics for a mode k during the inflation and beyond. Thanks to inflation, at very early times, all modes of interest are far inside the horizon and can be modulated by inflation. During the inflation, the Hubble sphere shrinks and the mode becomes super-Hubble and its evolution is frozen ($\mathcal{R}' = 0$), fixing the initial conditions for the post-inflationary Universe. Finally, during the standard Big-Bang evolution, the Hubble sphere increases and the mode re-enters the horizon following the dynamics described in Section 1.2.3.

Motivated by the slow-roll limit, one needs to consider

$$u = -z\mathcal{R} \quad \text{with} \quad z = \frac{a\phi_0'}{\mathcal{H}}, \tag{2.49}$$

which verifies, astonishingly, a very simple equation of motion derived from (2.44) and (2.45) (see, for instance, Mukhanov et al. 1992):

$$u'' + \left(k^2 - \frac{z''}{z}\right) u = 0. \tag{2.50}$$

This is the equation of a free field with a variable mass z''/z.

For super-Hubble modes, the dominant solution of 2.50 is $u = z\mathcal{R}_0$ where $\mathcal{R}_0$ is constant. Hence, after the horizon exit, the curvature perturbation $\mathcal{R}$ is, as expected, constant. The perturbations induced during the inflation will then persist after the end of inflation and will be the initial condition used in Section 1.2.3.2.

► **Slow-roll approximation** The shape of the initial power spectrum will be given by the solution of (2.50). Unfortunately, it is not possible to solve it in the general case. However, under the slow-roll approximation, there are two different time scales: one fast for the accelerated expansion (a) and one slow for the variation of the Hubble parameter and the scalar field such that $u \to a\delta\phi$, $z''/z \to a''/a \underset{s-r}{=} 2/\eta^2$, and (2.50) becomes

$$u'' + \left(k^2 - \frac{2}{\eta^2}\right) u = 0. \tag{2.51}$$

Similarly to how one quantize the Klein-Gordon equation (see Peskin and Schroeder 1995), u is promoted to operator $\hat{u}$

$$\hat{u}(\eta, \mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \left[\hat{a}_{\mathbf{k}} \psi_k(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + \hat{a}^\dagger_{\mathbf{k}} \psi^*_k(\eta) e^{-i\mathbf{k}\cdot\mathbf{x}}\right], \tag{2.52}$$

where $\hat{a}_{\mathbf{k}}$ (*resp.* $\hat{a}^{\dagger}_{\mathbf{k}}$) is the annihilation (*resp.* creation) operator and $\psi_k(\eta)$ is found for each mode k by solving (2.51):

$$\psi_k(\eta) = \frac{1}{\sqrt{2k}}\left(1 - \frac{i}{k\eta}\right)e^{ik\eta}. \tag{2.53}$$

Note that at small scales (far inside the horizon) *i.e.* $k\eta \to \infty$, (2.53) converge to the known solution in the Minkowski space validating the normalization of ψ_k.

For a sufficiently long period of inflation, $k\eta$ becomes very small, and the mode exits the horizon. Hence, for super-Hubble mode, $\psi_k \propto 1/\eta \propto a$ and $\delta\phi \propto \psi_k/a$ is constant. In this limit, $k\eta \ll 1$

$$\psi_k(\eta) = \frac{e^{-ik\eta}}{\sqrt{2k}}\frac{-i}{k\eta}, \tag{2.54}$$

and the power spectrum of $\delta\phi$ is[8]

$$P_{\delta\phi}(k) = \frac{|\psi_k(\eta)|^2}{a^2} = \frac{H^2}{2k^3}, \tag{2.55}$$

where the last equality is also from the slow-roll approximation[9]:

$$\eta \equiv \int_{a_e}^{a}\frac{\mathrm{d}a}{Ha^2} \simeq \frac{1}{H}\int_{a_e}^{a}\frac{\mathrm{d}a}{a^2} \simeq -\frac{1}{aH} \tag{2.56}$$

► **Primordial power spectrum** From (2.30), the comoving curvature perturbation reads as[10]

$$\mathcal{R} = \frac{3H^2}{\partial V/\partial\phi}\delta\phi, \tag{2.57}$$

and the power spectrum of the scalar perturbation is then

$$P_{\mathcal{R}}(k) = \frac{H_*^2}{2k^3}\frac{V_*^2}{(\partial V/\partial\phi)_*^2} = \frac{H_*^2}{4k^3\varepsilon_*}, \tag{2.58}$$

where $*$ corresponds to the value of the quantity when the mode exits the horizon *e.g.* $k = aH$ since the comoving curvature perturbation is conserved for super-Hubble modes.

The condition $k = aH$ implies

$$\mathrm{d}\log k = \mathcal{H}'\mathrm{d}\eta = -\frac{V}{M_{\mathrm{pl}}^2\,\partial V/\partial\phi}\mathrm{d}\phi, \tag{2.59}$$

and the derivative of the primordial power spectrum can be expressed with the slow-roll parameters ε, η

$$\frac{\mathrm{d}\log P_{\mathcal{R}}(k)}{\mathrm{d}\log k} = -3 - 6\varepsilon + 2\eta. \tag{2.60}$$

[8] Contrary to particle physics, observables in cosmology correspond to equal time correlators between positions in space. Thus, the two-point correlation of u is the expectation value of the observable $u(\eta,x)u(\eta,x')$ on a given state of the Universe, *a priori* during the inflation, to the cosmological vacuum, which will be immediately associated to $|0\rangle$ *i.e.* to the kernel of the annihilation operator. This identification is known as the *Bunch-Davies vacuum*. Hence, the two-point correlation is then $\xi(\eta,r) = \langle 0|u(\eta,\mathbf{x})u(\eta,\mathbf{x}+\mathbf{r})|0\rangle$ as in Section 1.2.1.2, and the associated power spectrum is then $P_u(k) = \left\langle|\hat{u}|^2\right\rangle = |\psi_k(\eta)|^2$.

[9] H is roughly constant during the inflation, and the scale factor at the end of the inflation is much larger than during the inflation, $a \ll a_e$

[10] During the inflation, Φ is negligible compare to the last term in (2.48), since the inflation aims to generate non-negligible metric perturbations.

Traditionally, the *spectral index* is then defined as

$$n_s \equiv 4 + \frac{\mathrm{d}\log P_{\mathcal{R}}(k)}{\mathrm{d}\log k} = 1 - 6\varepsilon + 2\eta, \tag{2.61}$$

Finally, the primordial power spectrum, see Section 1.2.3.2, is generally parametrized as

$$P_{\mathcal{R}}(k) = \frac{2\pi^2}{k^3} A_s \left(\frac{k}{k_{\mathrm{pivot}}}\right)^{n_s - 1}, \tag{2.62}$$

were A_s sets the amplitude of the power spectrum at k_{pivot} chosen to be 0.05 Mpc and n_s is intended to be closed to 1 in the slow-roll approximation (2.61).

Remark: (2.62) is a strong prediction of a single field slow roll inflation. To validate it, one can introduce α_s, the *running* of the spectral index, defined as

$$P_{\mathcal{R}}(k) = \frac{2\pi^2}{k^3} A_s \left(\frac{k}{k_{\mathrm{pivot}}}\right)^{n_s - 1 + \frac{1}{2}\alpha_s \ln(k/k_{\mathrm{pivot}})}, \tag{2.63}$$

where α_s is expected to be zero.

► **Primordial gravitational waves** The two polarization solutions $E_{+,\times}$ of (2.47) evolve and will be quantified independently. The correct variable is

$$v_p = \frac{a(\eta) E_p}{2M_{\mathrm{pl}}}, \tag{2.64}$$

where $p = +, \times$. Fortunately, v_p respects the same equation (2.50) than u and therefore can be quantified in the same manner. The power spectrum of the primordial tensor perturbations, counting the two polarization modes, is then

$$P_{\mathcal{T}}(k) = 4\frac{H_*^2}{k^3} \tag{2.65}$$

In the slow-roll approximation, the ratio of the tensor and scalar amplitudes is expressed by the first slow-roll parameter

$$r = \frac{P_{\mathcal{T}}(k)}{P_{\mathcal{R}}(k)} = 16\varepsilon. \tag{2.66}$$

Today, r is an important quantity to classify the different models of inflation, since gravitational waves should be produced during inflation.

Similarly to the scalar spectral index, one can define a tensor spectral index for the gravitational waves

$$n_{\mathcal{T}} = \frac{\mathrm{d}\log \mathcal{P}_{\mathcal{T}}(k)}{\mathrm{d}\log k} = -2\varepsilon. \tag{2.67}$$

► **Toy Model** For our toy model (2.33), the slow-roll parameters are $\varepsilon = \eta = 2\left(M_{\mathrm{pl}}/\phi\right)^2$ and the primordial fluctuations are parametrized by

$$n_s = 1 - 8\left(\frac{M_{\mathrm{pl}}}{\phi}\right)^2, \quad r = 32\left(\frac{M_{\mathrm{pl}}}{\phi}\right)^2, \quad n_{\mathcal{T}} = -4\left(\frac{M_{\mathrm{pl}}}{\phi}\right)^2. \tag{2.68}$$

To solve the horizon problem, one needs at least $N \sim 60$, and from (2.39)

$$\phi^2 = 4M_{\mathrm{pl}}^2\left(N + \frac{1}{2}\right) \approx 4M_{\mathrm{pl}}^2 N. \tag{2.69}$$

The primordial fluctuations are then controlled by

$$n_s = 1 - \frac{2}{N} \approx 0.966, \quad r = \frac{8}{N} \approx 0.133, \quad n_{\mathcal{T}} = -\frac{1}{N} \approx -0.016. \tag{2.70}$$

With this toy model, the amount of gravitational waves can be reduced by increasing the number of e-folds to the detriment of the value of n_s, bringing it closer to 1.

2.1.5 Model for inflation

The above description is generic and agnostic on the shape of the potential. If the potential respects the slowing down approximation, we can simplify the inflation paradigm, and it is, therefore, attractive to find potentials that verify this approximation. Many models were proposed, and we briefly present those used in Fig. 2.5. A complete and updated review is available in Martin et al. (2014).

► **Power-law inflation** Power-law inflation (Lucchin and Matarrese 1985) refers to model with a potential that has the form

$$V(\phi) = m^4 e^{-\alpha\phi/M_{\rm pl}}, \tag{2.71}$$

where α is a dimensionless parameter. These models were extensively studied since they lead to an exact solution in power-law such that $a \propto t^{2/\alpha^2}$ and $w = -1+\alpha^2/3$. The slow-roll parameters are

$$\varepsilon = \frac{\alpha^2}{2}, \quad \eta = \alpha^2. \tag{2.72}$$

► **Natural inflation** Natural inflation (Freese et al. 1990, Adams et al. 1993) was proposed to solve the well-known 'fine-tuning' problem of inflation *i.e.* we need to choose very particular values so that the inflationary predictions correspond to what the Universe looks like. This model used a potential written as

$$V(\phi) = m^4 \left[1 + \cos\left(\frac{\phi}{f}\right)\right], \tag{2.73}$$

where m is chosen to match the CMB normalization and f is an unknown free scale. The slow-roll parameters are

$$\epsilon = \frac{M_{\rm pl}^2}{2f^2}\frac{\sin^2 x}{(1+\cos x)^2}, \quad \eta = \frac{M_{\rm pl}^2}{f^2(1+\cos x)}\left(\frac{\sin^2 x}{1+\cos x} - 1\right), \tag{2.74}$$

with $x = \phi/f$.

► **R^2 inflation** Starobinsky (1980) noticed that the quantum corrections to general relativity should be important for the early Universe. These corrections lead to the modification of the Einstein-Hilbert action which then takes a similar form to the $f(R)$ modified gravity with a R^2 term. The new action becomes

$$\mathcal{S} = \frac{M_{\rm g}^2}{2}\int {\rm d}^4x\sqrt{-g}\left(R + \frac{R^2}{\mu^2}\right), \tag{2.75}$$

where $M_{\rm g}$ is the mass scale of gravity.

The curvature squared term, when it is dominant, acts as an effective cosmological constant and was naturally proposed by Starobinsky to be a candidate for inflation. The associated potential is written as

$$V(\phi) = \frac{M_{\rm g}^2 \mu^2}{8}\left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\phi}{M_{\rm g}}}\right)^2, \tag{2.76}$$

and the slow-roll parameters are

$$\epsilon = \frac{4}{3}\left(1 - e^{\sqrt{2/3}x}\right)^{-2}, \quad \eta = \frac{8}{3}\left(1 - e^{\sqrt{2/3}x}\right)^{-2} - \frac{1}{3}\left[\sinh\left(\frac{x}{\sqrt{6}}\right)\right]^{-2}, \tag{2.77}$$

where $x \equiv \phi/M_{\rm g}$. Fortunately, the primordial parameters can be expressed (see, for instance (48) in Planck Collaboration et al. 2020) as a function of the number of e-fold

$$n_s - 1 \simeq -\frac{2}{N}, \quad r \simeq \frac{12}{N^2}. \tag{2.78}$$

► **Multi-field inflation** The above models are all single-field models for which the slow-roll approximation is valid, but we can easily extend the slow-roll approximation to multi-field inflation (Gong 2017). For instance, two-field inflation can be described by the Lagrangian

$$\mathcal{L}_{\phi,\sigma} = -\frac{1}{2}g^{\mu\nu}\left(\partial_\mu\phi\partial_\nu\phi - \partial_\mu\sigma\partial_\nu\sigma - C_{\phi\sigma}\partial_\mu\phi\partial_\nu\sigma\right) - V(\phi,\sigma), \tag{2.79}$$

where V is any potential which depends on ϕ and σ and $C_{\phi\sigma}$ is a coupling constant. The equations of motion are simply obtained by computing the three Euler-Lagrange equations: $\partial\mathcal{S}/\partial g_{\mu\nu}$, $\partial\mathcal{S}/\partial\phi$, $\partial\mathcal{S}/\partial\sigma$.

2.1.6 Constraining inflation models

To constrain inflation, particularly single-field slow-roll inflation, the most effective approach currently is to measure the parameters n_s and r that characterize the primordial fluctuations. It is worth noting that inflation also predicts the presence of primordial gravitational waves, which have yet to be observed. These gravitational waves leave an imprint in the B-modes of the CMB polarization and their detection will support the inflationary paradigm. Measuring these B-modes represents one of the most significant challenges in the coming decade, and several dedicated instruments are currently designed such as CMB-S4 (Abazajian et al. 2022) and LiteBIRD (Hazumi et al. 2020).

By accurately measuring the values of n_s and r, we can place constraints on different inflationary models and narrow down the range of possible scenarios, described in the slow-roll single field case by the choice of one potential V. The current observational data, mainly driven by the Planck measurement (Planck Collaboration et al. 2020), have already provided valuable insights and placed constraints on the inflationary parameter space as shown in Fig. 2.5, where popular models of inflation are compared to the data. Some of these models are presented in Section 2.1.5. In particular, our toy model described by (2.33) is shown in yellow.

The Planck18 observations give

$$\begin{aligned} n_s &= 0.9649 \quad \pm 0.0042 \quad (68\%\text{CL}) \\ \alpha_s &= -0.0045 \quad \pm 0.0067 \quad (68\%\text{CL}) \end{aligned}, \tag{2.80}$$

and combining with the BICEP2/Keck Array data[11] (BK15, Keck Array and BICEP2 Collaborations et al. 2018),

$$r(k_{\rm pivot} = 0.002\ \text{Mpc}^{-1}) \equiv r_{0.002} < 0.056 \text{ at } 95\%\ \text{CL}.$$

[11] http://bicepkeck.org/index.html

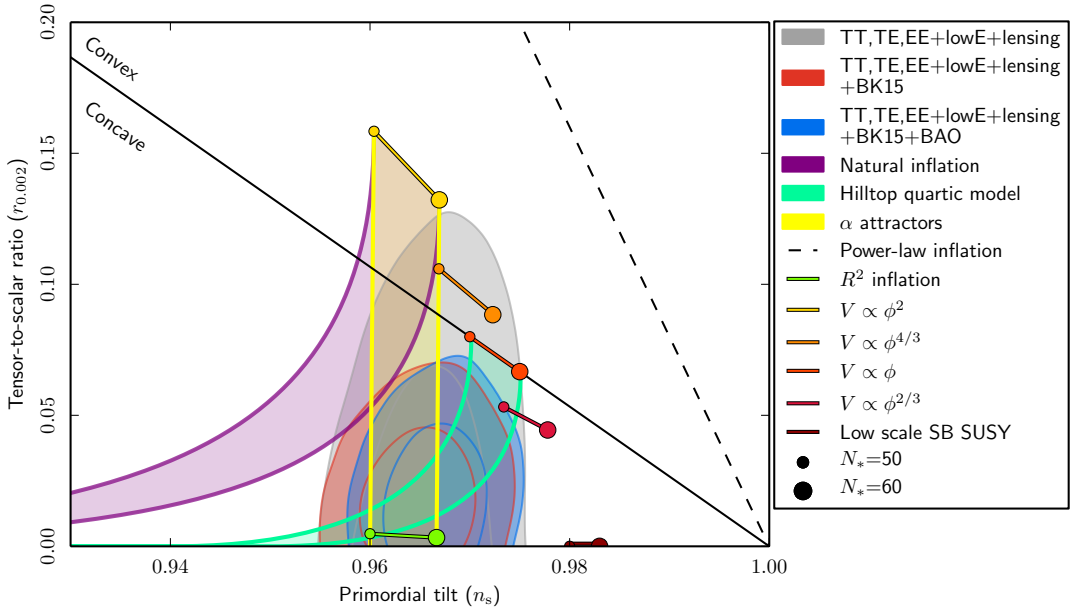


Figure 2.5: Planck18 constraints in the n_s, r plane at $k = 0.002$ Mpc^{-1} with 68% and 95% confidence level (Planck Collaboration et al. 2020). Gray contour gives the constraints using only Planck data and the color ones are the improvement by adding external data. Popular models of inflation are marked on this plane for two e-fold values ($N = 50, 60$). Some of these models are presented in Section 2.1.5.

This result is significantly improved by the new data release of BICEP/Keck (BK18, Keck Array and BICEP2 Collaborations et al. 2021). These constraints can then discriminate the different inflation models introduced in Section 2.1.5:

- Toy model *i.e.* ϕ^2 inflation (yellow): the simplest slow-roll single field inflation generated by a quadratic potential is already excluded at least at 1σ by the low value of r.
- Power-law inflation (black dashed line): these models do not have free parameter in the n_s, r plan and are likely excluded by the Planck 18 measurement.
- Natural inflation (purple): this model, while still in line with the BK15 data, is beginning to be ruled out in the light of the latest BK18 data.
- R^2 inflation (green): this model seems attractive since it is in agreement with the Planck18 measurement at 95% CL for $49 < N < 59$. In particular, it predicts a very low amount of primordial gravitational waves.

Of course, other single field models are also still in complete agreement with the data such as those shown in Fig. 2.5: Hilltop quartic model (Dimopoulos 2020) or α-attractors (Kallosh and Linde 2021).

2.2 Primordial non-gaussianity

As described above, the inflationary paradigm is a very natural way to explain the primordial fluctuations of the Universe. It is therefore reasonable to constrain inflationary models by the shape of these fluctuations. In particular, the prediction of primordial gravitational waves, that have not yet been detected, is a crucial constraint, since their non-detection would not fit the

inflation paradigm. However, this measurement based on the B-modes of the CMB polarization is hampered by the presence of galactic foregrounds. Besides, several inflation models can predict similar forms of primordial fluctuations. One can therefore look for other unique signatures of inflation in the primordial fluctuations that would not have faded.

The simplest models of inflation, single field slow-roll inflation, are expected to generate quasi-Gaussian primordial fluctuations, as we know from the study of the Klein-Gordon equation in quantum field theory (2.51). According to Wick's theorem, the primordial fluctuations are then completely described by the 2-point correlation function. Hence, the detection of primordial non-Gaussianity (PNG) *i.e.* the presence of higher-order correlations in primordial fluctuations is a *smooking-gun* against the single field slow-roll paradigm.

The non-gaussianity (NG) can come from different phenomena:

- *Primordial non-gaussianity (PNG)*: Some mechanism during the inflation can generate non-gaussianity during the early Universe in the primordial curvature perturbation $\mathcal{R}$. This is what we want to probe, to discriminate the different inflation scenarios.
- *Second-order non-gaussianity*: As explained in Section 1.2.3, the matter perturbed density field is connected to the primordial one thanks to the transfer function T_m which is currently computed in `CLASS` or `CAMB` at the first order in the perturbation theory. Some non-gaussianity can be generated by non-linear interaction (second-order perturbations) in the Einstein equations (1.60).
- *Secondary non-gaussianity*: Late time in the Universe, the perturbations become too large to be described by the linear theory (1.102) and one needs to resolve the complete equations (A.9). As explained in Appendix 3, gravity couples the modes k creating some non-gaussianity and in particular non-zero bispectrum (A.17).
- *Foreground non-gaussianity*: Non-gaussianity can be created by observational effect, in particular for the CMB observation, the galactic foregrounds generate some non-gaussianity.

The measurement of the primordial non-gaussianity requires therefore a prescient knowledge of the other forms of non-gaussianity. In particular, the most important contamination comes from the secondary non-gaussianity generated by gravity at small scales. This is all the more important for statistics that are measured at low redshift. Thus, the theory to model this effect needs to be carefully tested with realistic simulations to disentangle late-time evolution with real primordial non-gaussianity, and this is one of the major challenges for measuring PNG with galaxy bispectrum.

2.2.1 Shape of non-gaussianity

The bispectrum is the first order which describes the non-gaussianity. In the Gaussian case, it is expected to be zero. Therefore, the first step in probing primordial non-gaussianity is to measure it.

▶ **Primordial bispectrum** Following Section 1.2.1.2, the primordial bispectrum is

$$\left\langle \overline{\mathcal{R}}\left(\mathbf{k}_1\right)\mathcal{R}\left(\mathbf{k}_2\right)\mathcal{R}\left(\mathbf{k}_3\right)\right\rangle = (2\pi)^3\delta^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3\right)B_{\mathcal{R}}\left(k_1,k_2,k_3\right), \tag{2.81}$$

where k_1, k_2, k_3 form a triangle. The search for primordial non-gaussianity then starts with the exploration of the bispectrum. Note that even if we do not measure any bispectrum, that does not mean there is no non-gaussianity, it may be hidden in the high-order correlation functions.

A crucial point is that the different scenarios of inflation do not generate the same shape of bispectrum *i.e.* the same pattern in the triangular configuration, which makes it possible to discriminate between the different competitive extensions of single-field, slow-rolling inflation. For instance,

- *Squeezed Triangle* ($k_1 \ll k_2 \simeq k_3$):, also called local useful configuration since it enables us to discriminate between single and multi-field inflation, as we will see later (Choi and Byrnes 2010).
- *Equilateral Triangle* ($k_1 = k_2 = k_3$): these shapes can be generated with high-derivative interaction and complex speed of sound as in DBI inflation (Chen 2005) or in Ghost inflation (Izumi and Mukohyama 2010).
- *Folded Triangle* ($k_1 = 2k_2 = 2k_3$): models with non-standard vacuum initial state $|0\rangle$ create non-gaussianity in the folded configuration (Holman and Tolley 2008).

To probe these different configurations, it is useful to introduce some bispectrum templates that mimic the expected behavior of primordial non-gaussianity generated from different inflation scenarios. These templates are just a way to parametrize the amount of non-gaussianity and one can imagine as many parametrizations as one wants. Note that the theory at first order and the observations are in favor of almost Gaussian distribution. We therefore expect only a small deviation from gaussianity.

► **Change of convention** In the following, to match the *mainstream* definition of primordial non-gaussianity, we change our Newtonian gauge convention (1.58) for the one from Dodelson and Schmidt (2020):

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Psi)\mathrm{d}\eta^2 + (1+2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{2.82}$$

This **implies** the following changes:

$$\Phi \to \Psi \text{ and } \Psi \to -\Phi,$$

such that in absence of anisotropic stress ($\Pi = 0$)

$$\Phi = \Psi \to \Psi = -\Phi, \text{ and, } \Phi = -3/5\mathcal{R} \to \Phi = 3/5\mathcal{R} \quad (k \ll \mathcal{H}).$$

In particular, with this new convention the Poisson equation (1.59) becomes

$$\Delta\Phi = -4\pi G a^2 \bar{\rho}\delta, \tag{2.83}$$

and in Fourier space:

$$k^2\Phi = 4\pi G a^2 \bar{\rho}\delta. \tag{2.84}$$

Remark: As described by (2.83), the potential Φ has the opposite sign to the usual Newtonian gravitational potential (well described by Ψ in our notation (1.58)) *i.e.* is positive. Thus, a positive increase of Φ, see (2.86), represents an increase of the gravity!

► **Local template** Popularized by Komatsu and Spergel (2001) (similar parametrization were already introduced by, for instance, Salopek and Bond 1990, Gangui et al. 1994), local

template[12] was one of the first phenomenological parametrizations of primordial non-gaussianity using a non-linear correction of the comoving curvature perturbation $\mathcal{R}$:

$$\mathcal{R}(\mathbf{x}) = \mathcal{R}_g(\mathbf{x}) + \frac{3}{5} f_{\rm NL}^{\rm loc} \left[\mathcal{R}_g(\mathbf{x})^2 - \left\langle \mathcal{R}_g(\mathbf{x})^2 \right\rangle \right], \tag{2.85}$$

where $\mathcal{R}_g$ is the gaussian part of $\mathcal{R}$. The factor 3/5 is generally used since the local template was originally introduced in terms of Newtonian potential Φ that are related to $\mathcal{R}$ in the matter-dominated area via $\Phi = 3/5\mathcal{R}$. With this definition, we have

$$\Phi(\mathbf{x}) = \Phi_g(\mathbf{x}) + f_{\rm NL}^{\rm loc} \times \left[\Phi_g(\mathbf{x})^2 - \langle \Phi_g(\mathbf{x}) \rangle^2 \right], \tag{2.86}$$

where $f_{\rm NL}^{\rm loc}$ tracks the amount of local primordial non-gaussianity in the matter-dominated area.

Depending on how the measurement is made, one can also get back to the value of Φ well before the recombination in the radiation-dominated area such that the factor 3/5 is not the correct one to get (2.86), and one could prefer to use 2/3. It is solely a problem of definition of $f_{\rm NL}^{\rm loc}$ and in this dissertation, we always chose $f_{\rm NL}^{\rm loc}$ such that (2.86) is true in the matter-dominated area.

The Wick theorem ensures that the bispectrum reads as[13]

$$\begin{aligned} B_{\mathcal{R}}^{\rm loc}(k_1, k_2, k_3) &= \frac{6}{5} f_{\rm NL}^{\rm loc} \times \left[P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_2) + 2 \text{ perm. } \right] \\ &= \frac{6}{5} f_{\rm NL}^{\rm loc} \times \frac{A_{\mathcal{R}}^2}{(k_1 k_2 k_3)^2} \underbrace{\left[\frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_1 k_3} + \frac{k_3^2}{k_1 k_2} \right]}_{= 3 \times \mathcal{S}^{\rm loc}(k_1, k_2, k_3)}, \end{aligned} \tag{2.87}$$

where we assumed a scale invariant power spectrum $P_{\mathcal{R}}(k) = A_{\mathcal{R}}/k^3$. The computation can be easily extended for non-scale invariant power spectrum (2.62).

Finally, in the squeezed configuration ($k_1 \ll k_2 \simeq k_3$), the bispectrum becomes

$$\lim_{k_1 \to 0} B_{\mathcal{R}}^{\rm loc}(k_1, k_2, k_3) = \frac{12}{5} f_{\rm NL}^{\rm loc} \times P_{\mathcal{R}}(k_1) P_{\mathcal{R}}(k_3). \tag{2.88}$$

Remark: Current observations show a very low amount of primordial non-gaussianity, the total "non-gaussian" field Φ can be therefore expanded around gaussian solution Φ_g[14]. Note that the amplitude of Φ_g is about 10^{-5}, and the Taylor expansion converges quickly. We can always perform this expansion such that the amount of PNG is correctly given by the value from the template.

► **Equilateral template** Inflations with higher-derivation, as DBI inflation (Chen 2005), have a key characteristic that the bispectrum vanishes when one mode is far outside the horizon suggesting that the bispectrum is maximal and peaks for equilateral configurations ($k_1 \approx k_2 \approx k_3$). One can then search for these signatures using the following template

$$B_{\mathcal{R}}^{\rm equil} = \frac{18}{5} f_{\rm NL}^{\rm equil} \times \frac{A_{\mathcal{R}}^2}{(k_1 k_2 k_3)^2} \underbrace{\left[\left(\frac{k_1}{k_2} + 5 \text{ perm.} \right) - \left(\frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) - 2 \right]}_{= \mathcal{S}^{\rm equil}(k_1, k_2, k_3)}. \tag{2.89}$$

[12] This template is called local since it is defined locally in real space.

[13] The Fourier transform of the non gaussian part of (2.86) can be written as $\int \frac{d^3\mathbf{k}}{(2\pi)^3} \Phi_g(\mathbf{k} + \mathbf{k}') \Phi_g(\mathbf{k}') - (2\pi)^3 \delta_D(\mathbf{k}) \left\langle \Phi_g^2 \right\rangle$.

[14] One can expand $\Phi = f(\Phi_g)$

► **Orthogonal template** One can define a last template orthogonal to the local and equilateral templates, $\mathcal{S}_{\rm ortho}\cdot\mathcal{S}_{\rm loc}=\mathcal{S}_{\rm ortho}\cdot\mathcal{S}_{\rm equil}\equiv 0$. In particular, this shape of the template probes signatures from inflation models with higher-derivative interactions. The template is

$$B_{\mathcal{R}}^{\rm ortho}=\frac{18}{5}f_{\rm NL}^{\rm ortho}\times\frac{A_{\mathcal{R}}^2}{(k_1k_2k_3)^2}\underbrace{\left[-3\left(\frac{k_1^2}{k_2k_3}+2\text{ perm.}\right)+3\left(\frac{k_1}{k_2}+5\text{ perm.}\right)-8\right]}_{=\mathcal{S}^{\rm ortho}(k_1,k_2,k_3)}. \tag{2.90}$$

The different shapes $\mathcal{S}^{\rm loc}$, $\mathcal{S}^{\rm equil}$ and $\mathcal{S}^{\rm ortho}$ are displayed in Fig. 2.6. For visualization, we have plotted $S(1,x_2,x_3)$ with $x_2=k_2/k_1$, $x_3=k_3/k_1$.

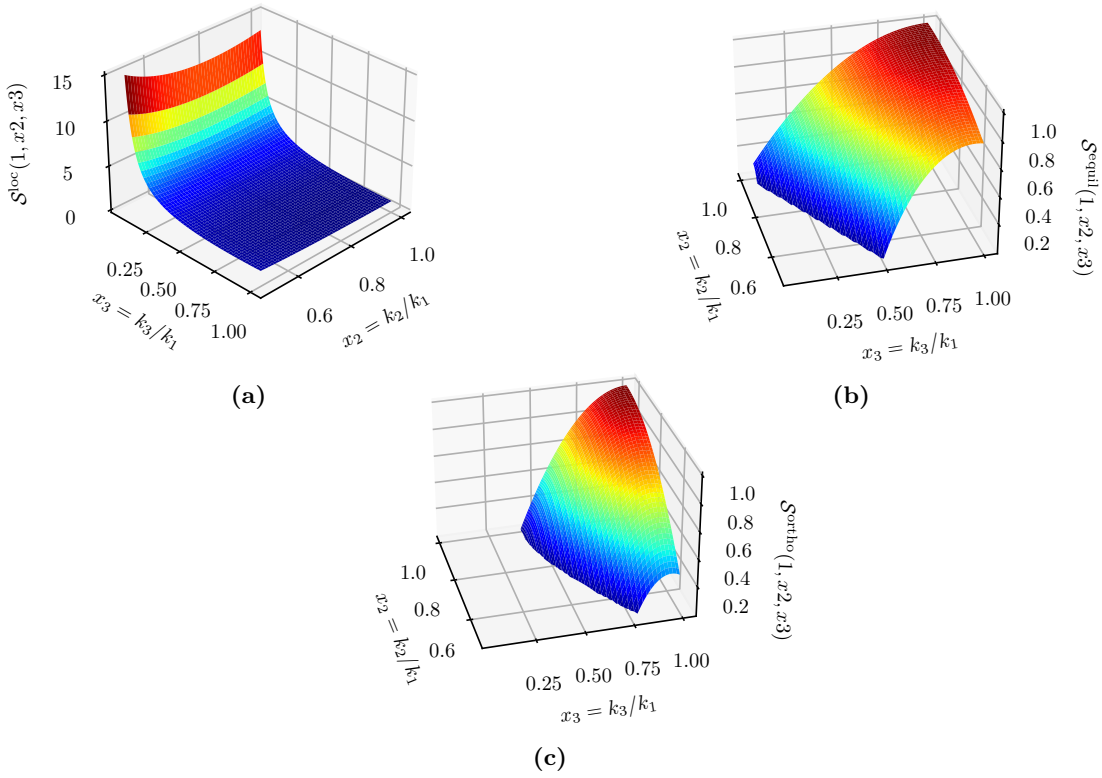


Figure 2.6: Visualization of the (a) local (2.87) (b) equilateral (2.89) (c) orthogonal (2.90) templates. We have plotted here, for simplicity, $S(1,x_2,x_3)$ with $x_2=k_2/k_1$, $x_3=k_3/k_1$. The local template peaks in squeeze configuration ($x_3\approx 0$), while the equilateral template for $x_2\approx x_3\approx 1$.

2.2.2 Consistency Relation

The generation of primordial non-gaussianity during inflation is well beyond the scope of this dissertation and will not be discussed here. We only mention the fact that this can be done through two different mechanisms

- *Before horizon exit*: Quantum-mechanical effects can generate non-gaussianity. This study can be performed via the *Schwinger-Keldysh in-in formalism* (Schwinger 1961), see Maldacena (2003), Weinberg (2005) for a description of the *in-in* formalism applied to inflation.

- *After horizon exit*: Classical non-linear evolution can generate non-gaussianity. This can be computed with the δN formalism (see, for instance, Sugiyama et al. 2013).

However, without digging into these details, one can derive a very powerful relation for any single field inflation model, known as the *consistency relation* (Creminelli and Zaldarriaga 2004). Indeed, in squeezed configuration, one long-wavelength mode $k_1 = k_l$ is correlated with two short-wavelength modes $k_2 \simeq k_3 = k_s$ such that $k_l \ll k_s$ and

$$\langle \mathcal{R}_{\mathbf{k_1}} \mathcal{R}_{\mathbf{k_2}} \mathcal{R}_{\mathbf{k_3}} \rangle \longrightarrow \langle \mathcal{R}^2_{\mathbf{k_s}} \mathcal{R}_{\mathbf{k_l}} \rangle \tag{2.91}$$

Since the modes with longer wavelength freeze earlier (they cross the horizon earlier), the mode k_l will already be frozen when the two shorter modes freeze, and it will act as a background field. Hence, the bispectrum can be estimated to

$$\langle \mathcal{R}^2_{\mathbf{k_s}} \mathcal{R}_{\mathbf{k_l}} \rangle \approx \langle \langle \mathcal{R}^2_{\mathbf{k_s}} \rangle \Big|_{\mathcal{R}_{\mathbf{k_l}}} \mathcal{R}_{\mathbf{k_l}} \rangle. \tag{2.92}$$

Using the ADM formalism (Arnowitt et al. 2008) to correctly add the dependence of the comoving curvature perturbation, the metric once the mode k_l is frozen can be written as

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t) e^{2\mathcal{R}(x)} \mathrm{d}\mathbf{x}^2, \tag{2.93}$$

where the time dependence of $\mathcal{R}$ was removed since it is constant[15] for super-Hubble modes.

Following Creminelli et al. (2011), the computation will be easier in real space, since the comoving curvature perturbation will be absorbed in the line element by a simple local rescaling $\mathrm{d}\mathbf{x}' = e^{\overline{\mathcal{R}}(x)} \mathrm{d}\mathbf{x}$ where $\overline{\mathcal{R}}(x)$ is the background field induced by the frozen $\mathcal{R}_{\mathbf{k}_1}$. Under the effect of the background field,

$$\Delta\mathbf{x}' \simeq e^{\overline{\mathcal{R}}} \Delta\mathbf{x} \approx \Delta\mathbf{x} + \overline{\mathcal{R}} \Delta\mathbf{x}, \tag{2.94}$$

where $\Delta\mathbf{x} = \mathbf{x_3} - \mathbf{x_2}$ and $\overline{\mathcal{R}}$ is evaluated implicitly in $(\mathbf{x}_2 + \mathbf{x}_3)/2$.

Hence, the 2-point correlation of the short modes under a background field can be evaluated via a first-order expansion

$$\begin{aligned} \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle|_{\overline{\mathcal{R}}} &= \langle \mathcal{R}\left(\mathbf{x}'_2\right) \mathcal{R}\left(\mathbf{x}'_3\right) \rangle \\ &= \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle|_0 + \overline{\mathcal{R}} \left[\Delta\mathbf{x} \cdot \boldsymbol{\nabla} \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle \right] \Big|_0 , \end{aligned} \tag{2.95}$$

where $|_0$ denotes the quantity without background field *i.e.* $\overline{\mathcal{R}} \to 0$. The squeezed 3-point correlation function is then, from (2.92),

$$\begin{aligned} \left\langle \overline{\mathcal{R}}(\mathbf{x}_1)\mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \right\rangle &= \left\langle \overline{\mathcal{R}}(\mathbf{x}_1) \right\rangle \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle \Big|_0 + \langle \overline{\mathcal{R}}(\mathbf{x}_1) \overline{\mathcal{R}} \rangle \left[\Delta\mathbf{x} \cdot \boldsymbol{\nabla} \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle \right] \Big|_0 \\ &= \langle \overline{\mathcal{R}}(\mathbf{x}_1) \overline{\mathcal{R}} \rangle \left[\Delta\mathbf{x} \cdot \boldsymbol{\nabla} \langle \mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3) \rangle \right] \Big|_0 \end{aligned} . \tag{2.96}$$

The first term vanishes since it is the 1-point correlation function. To find the bispectrum, one needs to re-write the above equation as an integration on $\mathbf{k}$, the left term can be written as

$$\left\langle \overline{\mathcal{R}}(\mathbf{x}_1) \overline{\mathcal{R}} \right\rangle = \int_{\mathbf{k}_L} P(k_L) e^{i\mathbf{k}_L \cdot (\mathbf{x}_1 - (\mathbf{x}_2 + \mathbf{x}_3)/2)}, \tag{2.97}$$

[15] $\mathcal{R}$ is constant outside the horizon only for adiabatic perturbations. Fortunately, we are considering here single field inflation that generates only adiabatic perturbations.

and the right one as

$$\begin{aligned}\Delta\mathbf{x}\cdot\boldsymbol{\nabla}\langle\mathcal{R}(\mathbf{x}_2)\mathcal{R}(\mathbf{x}_3)\rangle &= \int_{\mathbf{k}_S} P\left(k_S\right)\left(\mathbf{k}_S\cdot\partial_{\mathbf{k}_S}\right)e^{i\mathbf{k}_S\cdot\Delta\mathbf{x}}\\ &= -\int_{\mathbf{k}_S}\partial_{\mathbf{k}_S}\cdot\left[\mathbf{k}_S P\left(k_S\right)\right]e^{i\mathbf{k}_S\cdot\Delta\mathbf{x}}\\ &= -\int_{\mathbf{k}_S} P\left(k_S\right)\frac{\mathrm{d}\ln\left(k_S^3 P\left(k_S\right)\right)}{\mathrm{d}\ln k_S}e^{i\mathbf{k}_S\cdot\Delta\mathbf{x}}\end{aligned}\quad , \tag{2.98}$$

where we have used the integration by parts for the second equality[16] and for the last one can be verified by carefully expanding the two terms[17].

Finally, using the two above results and adding the fact that $\int_{\mathbf{k}_1}(2\pi)^3\delta_{\mathrm{D}}^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3\right)=1$, the 3-point correlation function can be written as

$$\begin{aligned}\left\langle\overline{\mathcal{R}}\left(\mathbf{x}_1\right)\mathcal{R}\left(\mathbf{x}_2\right)\mathcal{R}\left(\mathbf{x}_3\right)\right\rangle = -\int_{\mathbf{k}_1}\int_{\mathbf{k}_2}\int_{\mathbf{k}_3} e^{-i\left(\mathbf{k}_1\cdot\mathbf{x}_1+\mathbf{k}_2\cdot\mathbf{x}_2+\mathbf{k}_3\cdot\mathbf{x}_3\right)}\\ \times(2\pi)^3\delta^{(3)}\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3\right)P\left(k_1\right)P\left(k_2\right)\frac{\mathrm{d}\ln\left(k_2^3P\left(k_2\right)\right)}{\mathrm{d}\ln k_2}\,,\end{aligned} \tag{2.99}$$

leading to the consistency relation:

$$\lim_{k_1\to 0}\left\langle\mathcal{R}_{\mathbf{k}_1}\mathcal{R}_{\mathbf{k}_2}\mathcal{R}_{\mathbf{k}_3}\right\rangle = (2\pi)^3\delta\left(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3\right)\left(1-n_{\mathrm{s}}\right)P_{\mathcal{R}}\left(k_1\right)P_{\mathcal{R}}\left(k_3\right), \tag{2.100}$$

where we use the primordial power spectrum generated by single field slow-roll inflation model (2.61).

Matching (2.100) to (2.88) leads to

$$f_{\mathrm{NL}}^{\mathrm{loc}} = \frac{5}{12}(1-n_s). \tag{2.101}$$

The current observations give $n_s = 0.9649\pm 0.0042$ *i.e.* $f_{\mathrm{NL}}^{\mathrm{loc}}\sim 10^{-2}$ for the single field inflation model.

Remark: The consistency relation (2.100) is, independent of the shape of the potential ($V(\phi)$), of the kinetic term ($\partial_\mu\phi$), and of the initial vacuum state. The detection of primordial non-gaussianity in the squeezed limit thus has the authority to rule out all the single field inflation! That is why the measurement of local shape has gained more and more interest in recent years! The aim of upcoming observations is to achieve the sensitivity to detect $f_{\mathrm{NL}}^{\mathrm{loc}}\sim\mathcal{O}(5)$ with this method in a very near future (Laureijs et al. 2011, DESI Collaboration et al. 2016b) and then to reach $f_{\mathrm{NL}}^{\mathrm{loc}}<\mathcal{O}(1)$.

2.2.3 Scale-dependent bias

As seen above, inflation can imprint specific signatures into cosmological perturbations and thus into the matter distribution through primordial non-gaussianity.

[16] The integrated product term vanishes since the P is 0 at both terminals.

[17] We have used the fact that $\partial_{\mathbf{k}}\cdot\mathbf{k}=3$, derivate the second term of the product using spherical coordinates, and $\mathrm{d}/\mathrm{d}\ln(k) = \left(\mathrm{d}\ln(k)/\mathrm{d}k\right)^{-1}\mathrm{d}/\mathrm{d}k$

► **Probe PNG with LSS** These signatures can be observed directly in the bispectrum as has already been done with the CMB temperature and polarization distributions (Planck Collaboration et al. 2020). In particular, they found,

$$f_{\rm NL}^{\rm loc} = -0.9 \pm 5.1; \quad f_{\rm NL}^{\rm equil} = -26 \pm 47; \quad f_{\rm NL}^{\rm ortho} = -38 \pm 24, \tag{2.102}$$

at 68% confidence level. These are the best competitive measurement so far. Unfortunately, the Planck measurements are already close to the cosmic variance limit[18], and the next generation of the CMB experiment, as CMB-S4, will only improve the current constraints by a factor ~ 2 (Abazajian et al. 2016).

To circumvent the cosmic variance limit of CMB observations, one possibility would be to use the enormous statistical power in the 3D galaxy clustering, probing a large volume of the Universe. The direct measurement of the bispectrum will not be used in this dissertation since is complicated by the important presence of secondary non-Gaussianity, see Appendix 3, and requires a theoretical and technical development that is ongoing (Moradinezhad Dizgah et al. 2021), much more complex than the simple BAO/RSD measurement, see Section 1.3.2.4.

Remark: This dissertation focuses on the local primordial non-gaussianity constraint through the measurement of $f_{\rm NL}^{\rm loc}$ and in the rest of this dissertation, we will not distinguish $f_{\rm NL}$ and $f_{\rm NL}^{\rm loc}$.

► **PNG with power spectrum** A promising approach is through the unexpected tiny imprint left at large scales on the matter power spectrum by local PNG, known as the *scale-dependent bias*. This relation, firstly discovered by Dalal et al. (2008) and later confirmed by (Desjacques et al. 2009, Grossi et al. 2009, Pillepich et al. 2010), has radically changed the search for primordial non-gaussianity over the last decade, since it allows the measurement of PNG only with the power spectrum, and at scales where the linear theory holds.

In Section 1.3.1.1, the galaxy bias has been assumed to depend only on the field $\delta(\mathbf{x})$ at linear order. However, the presence of local PNG can impact the formation of galaxies. Indeed, a positive value of $f_{\rm NL}^{\rm loc}$ in (2.86) will increase the primordial gravitational field where the density field is the densest such that the matter will aggregate in primordial potential wells forming denser regions more quickly. It acts as a local rescaling of the critical threshold above which a galaxy can form; see next paragraph. Remember, here, Φ is positive *i.e.* has the opposite sign to the usual Newtonian gravitational potential. And the exact opposite for a negative value $f_{\rm NL}^{\rm loc}$ which will slow down the aggregation of matter in the primordial potential wells.

Fig. 2.7a shows simulations with similar initial conditions but with different values of $f_{\rm NL}^{\rm loc}$: -5000, -500, 0, 500, 5000 from top to bottom. The impact of $f_{\rm NL}^{\rm loc}$ is clearly visible by eye-matching the large scale structures.

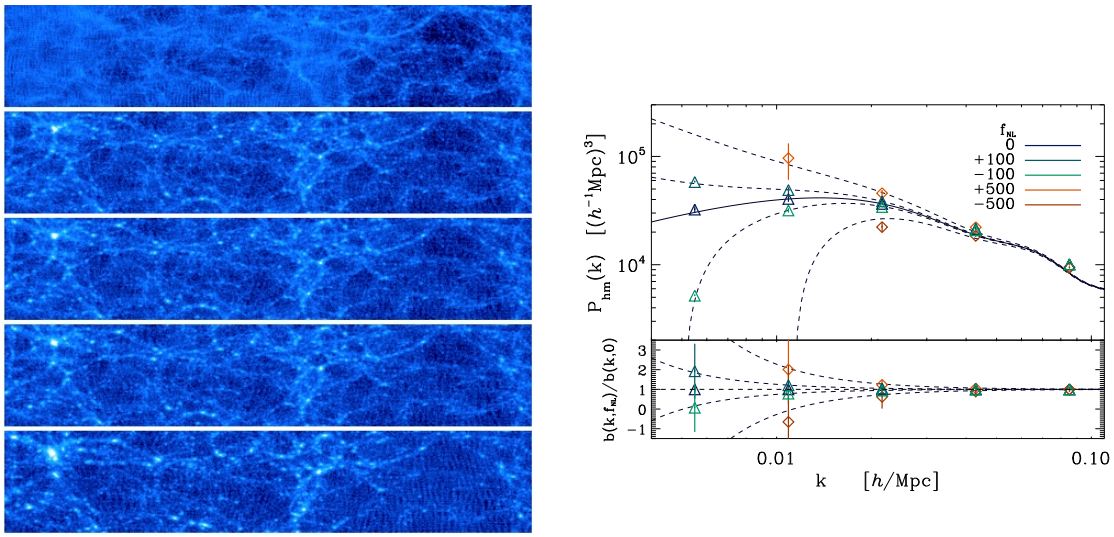
Hence, we need to complete the galaxy bias (1.111) (see Desjacques et al. 2018, for a rigorous derivation) with an additional term taking into account the local modifications implied by the primordial gravitational field,

$$\delta_g(\mathbf{x}) = b_1\delta(\mathbf{x}) + b_\Phi f_{\rm NL}^{\rm loc}\Phi_{prim}(\mathbf{x}) + \epsilon, \tag{2.103}$$

where b_Φ is a new free parameter giving the sensibility of a specific type of galaxy to the local rescaling of the critical threshold induced by the presence of local PNG.

The primordial gravitational field Φ_{prim} can be related to the matter density perturbation δ through a transfer function (1.77) such that $\delta(k) = \alpha(k)\Phi_{prim}(k)$. Note that to match the

[18] See Section 4.1.2.



(a) **(b)**

Figure 2.7: These figures are from Dalal et al. (2008), who first noticed the impact of local PNG on the halos power spectrum. (a) Simulation at $z = 0$ with similar initial conditions but with $f_{\rm NL}^{\rm loc} = -5000, -500, 0, +500, +5000$ from top to bottom. Each slice is 375 h^{-1} Mpc wide and 80 h^{-1} Mpc high and deep. The impact of positive or negative value of $f_{\rm NL}^{\rm loc}$ can be visualized by eye-matching the large scale structures. In the case of positive value, primordial over-dense regions are more clustered. (b) The upper panel shows the cross-power spectra between halos and dark matter. The points are measurements from simulations with several values of $f_{\rm NL}^{\rm loc}$ at $z = 1$ for halos of mass $1.6\times10^{13}M_\odot < M < 3.2\times10^{13}M_\odot$. The black and dashed lines are the theoretical prediction given by (2.104). The bottom panel shows the ratio $b(k, f_{\rm NL}^{\rm loc})/b(k, f_{\rm NL}^{\rm loc} = 0)$. The measurements with $f_{\rm NL}^{\rm loc} = -500, 500$ are from smaller simulations and then have bigger statistical errors. The scale-dependent bias strongly modifies the value of the power spectrum at large scales *i.e.* at small k. The expected effect is of course much smaller than the one shown here.

definition (2.86), Φ_{prim} is normalized in the matter-dominated area such that $\Phi = 3/5\mathcal{R}$. The Fourier transform of the field δ_g is

$$\delta_g(\mathbf{k}) = \left(b_1 + \frac{b_\Phi}{\alpha(k)} f_{\rm NL}^{\rm loc}\right)\delta(\mathbf{k}), \tag{2.104}$$

where the stochastic contribution was dropped for convenience.

Finally, in the presence of local PNG, the galaxy power spectrum in linear order is

$$P(k,z) = \left(b_1(z) + \frac{b_\Phi(z)}{\alpha(k,z)} f_{\rm NL}^{\rm loc}\right)^2 \times P_{lin}(k,z), \tag{2.105}$$

where $\alpha(k,z)$ can be calculated in two different ways:

- One can directly compute the transfer function from `CLASS`[19]

$$\alpha(k,z) = T_{\Phi\to\delta}(k,z) = \sqrt{\frac{P_\delta(k,z)}{P_{\Phi_{prim}}(k)}}, \tag{2.106}$$

[19]The different convention used in `CLASS` ($\Phi \to -\Phi$) is not a problem since we are computing the transfer function as the square root of the ratio of power spectra! One just needs to check the expected sign of α that should be positive.

and since Φ_{prim} is normalized in the matter-dominated area such that $\Phi = 3/5\mathcal{R}$, and with (1.82), the primordial power spectrum can be written as

$$P_{\Phi_{prim}}(k) = \frac{9}{25}\frac{2\pi^2}{k^3}A_s\left(\frac{k}{k_{\text{pivot}}}\right)^{n_s-1}. \tag{2.107}$$

- Or as in Slosar et al. (2008), one can link δ to Φ in matter dominated area through the Poisson equation (2.84):

$$\alpha(k,z) = \frac{2}{3}\frac{k^2T(k)D(z)}{\Omega_m}\frac{c^2}{H_0^2}, \tag{2.108}$$

where c is the speed of light, H_0 is the present-day Hubble parameter, Ω_m is the matter density parameter at $z = 0$, and $T(k)D(z)$ is the transfer function for the gravitational potential $T_{\Phi\to\Phi}(k,z)$ with $D(z)$ the linear growth factor normalized to be $(1+z)^{-1}$ in the matter-dominated area.

In the following, we will prefer to use the direct computation with the ratio power spectrum given by `CLASS`.

The scale-dependent bias leaves a unique imprint on large scales ($k < 0.001$ Mpc^{-1} h) in the power spectrum as shown in Figure 2.7b. Although this relationship is theoretically very promising, since it involves scales where the theory is linear, it is, unfortunately, subject to very significant systematic effects, known as *imaging systematic effects*, making them tricky to exploit, as we shall see in this dissertation.

► **Expecting value of b_Φ** The scale-dependent bias (2.104) introduce a new free parameter, b_ϕ, that can be derived in the peak-background formalism (Cole and Kaiser 1989). We follow Slosar et al. (2008)[20].

The density field can be split into uncorrelated long (δ_l) and short (δ_s) wavelength perturbations

$$\rho(\mathbf{x}) = \bar{\rho}\,(1+\delta_l+\delta_s)\,. \tag{2.109}$$

Similarly, The gaussian part of the primordial gravitational field can be split as $\Phi_g = \phi_l + \phi_s$, such that

$$\Phi = \phi_l + f_{\text{NL}}\phi_l^2 + (1+2f_{\text{NL}}\phi_l)\,\phi_s + f_{\text{NL}}\phi_s^2 + \text{ const.} \tag{2.110}$$

The contrast density field computed from a sufficiently average local density depends only on the large scale modes *i.e.* on δ_l, and the transfer function (2.106) gives $\delta_l(k) = \alpha(k)\Phi(k)$. Hence, with (2.110),

$$\delta_l(k) = \alpha(k)\phi_l(k), \tag{2.111}$$

since $f_{\text{NL}}\phi_l^2$ is secondary order in ϕ_l, $(1+2f_{\text{NL}}\phi_l)\phi_s$ has short wavelength contribution and $f_{\text{NL}}\phi_s^2$ adds a small constant contribution at large scales. Note that ϕ_l and ϕ_s are uncorrelated such that $\phi_l \times \phi_s$ is not a second order term.

Within a large scale over-density, the short wavelength modes of the contrast density field are

$$\delta_s = \alpha\left[X_1\phi_s + X_2\phi_s^2\right], \tag{2.112}$$

[20] A complete derivation can be found in Section 7 of Desjacques et al. (2018).

with $X_1 = 1 + 2f_{\rm NL}\phi_l$, $X_2 = f_{\rm NL}$. The local Lagrangian number density of galaxies *i.e.* the number density of haloes per unit halo mass at a position $\mathbf{x}$ will then depends not just on δ_l but also on X_1, X_2: $n(\delta_l, X_1, X_2)$ such that the Lagrangian bias reads as

$$\begin{aligned} b_L(M,k) &\equiv \bar{n}^{-1}\frac{\mathrm{d}n}{\mathrm{d}\delta_l} \\ &= \bar{n}^{-1}\left[\frac{\partial n}{\partial\delta_l(\mathbf{x})} + 2f_{\rm NL}\frac{\mathrm{d}\phi_l(k)}{\mathrm{d}\delta_l(k)}\frac{\partial n}{\partial X_1}\right], \\ &= b_L^g + 2f_{\rm NL}\frac{\mathrm{d}\phi_l(k)}{\mathrm{d}\delta_l(k)}\frac{1}{\bar{n}}\frac{\partial n}{\partial X_1} \end{aligned} \tag{2.113}$$

where we have used the definition of the Lagrangian bias in the Gaussian case:

$$b_L^g = \bar{n}^{-1}\,\partial n/\partial\delta_l(\mathbf{x})\,. \tag{2.114}$$

Non-gaussianity acts as a local rescaling of the amplitude of small-scale perturbations (2.112) and it can be parametrized by σ_8 (1.137) such that

$$\sigma_8^{\rm local}(\mathbf{x}) = \sigma_8 X_1(\mathbf{x}) \quad \Rightarrow \quad \delta\sigma_8^{\rm local} = \sigma_8\delta X_1 \tag{2.115}$$

Since the Lagrangian bias is related to the Eulerian bias through $b_1 = b_L + 1$ (Matsubara 2008), one can identify with (2.113) the value of b_Φ:

$$b_\Phi = 2\frac{\partial\ln n}{\partial\ln\sigma_8}, \tag{2.116}$$

where we have dropped the local label.

► **Universal mass function** Haloes from dark matter simulations respect the *universal mass function* (Manera et al. 2010, Hoffmann et al. 2015) *i.e.* the mean abundance of haloes per unit halo mass $\bar{n}$ can be written only as a function of the significance $\nu(M)$[21]

$$\bar{n}(M) = \bar{n}(M,\nu) = \frac{\bar{\rho}_m}{M}\nu f(\nu)\frac{\mathrm{d}\ln\nu}{\mathrm{d}\ln M}, \tag{2.117}$$

where the significance is defined as $\nu = \delta_c/\sigma(M)$ with $\delta_c = 1.686$, the spherical collapse linear over-density (see, for instance, Mo and White 1996) and $\sigma(M)$ is the variance of the linear matter density field smoothed on the Lagrangian radius R that is related to M through $M = (4\pi/3)\bar{\rho}_m R^3(M)$. $f(\nu)$ is the fraction of mass that collapses into haloes of significance between ν and $\nu + \mathrm{d}\nu$.

Applying (2.114) and (2.116), the Lagrangian bias and the new bias b_Φ can be computed explicitly

$$\begin{aligned} b_L &= \bar{n}^{-1}\frac{\partial\bar{n}}{\partial\delta_l(\mathbf{x})} = -\frac{1}{\sigma}\frac{1}{\nu f(\nu)}\frac{\mathrm{d}\left[\nu f(\nu)\right]}{\mathrm{d}\nu} \\ b_\Phi &= 2\frac{\partial\ln\bar{n}}{\partial\ln\sigma_8} = -2\frac{\delta_{cr}}{\sigma}\frac{1}{\nu f(\nu)}\frac{d\left[\nu f(\nu)\right]}{d\nu} \end{aligned}. \tag{2.118}$$

Finally, the new free parameter b_Φ can be related to known parameters

$$b_\Phi = 2\delta_c b_L = 2\delta_c(b_1 - 1), \tag{2.119}$$

[21] The Jacobian is $\left|\frac{\mathrm{d}\ln\sigma(M)}{\mathrm{d}\ln M}\right| = \frac{\mathrm{d}\ln\nu(M)}{\mathrm{d}\ln M}$.

this expression of b_Φ is known as the *universal mass relation*.

Remark: Unfortunately with galaxy surveys, we are not probing directly the distribution of dark matter haloes but much more complex structure formations which are not entirely specified by the halo mass M. The universal mass relation is a reasonable starting point, but it needs to be refined according to the type of the observed objects. For quasars, for instance, Slosar et al. (2008) uses the fact that the quasars formation is thought to be triggered by mergers of halos such that (2.119) has to be corrected by $b_\phi = 2\delta_c(b_1 - 1 - \delta_c^{-1})$ and was later verified by Reid et al. (2010) with dark matter simulations. Hence, the bias is generally described by a parameter p such that

$$b_\phi = 2\delta_c(b_1 - p), \tag{2.120}$$

with $p = 1$ for dark matter haloes and $p = 1 + \delta_c^{-1} \simeq 1.6$ for recent mergers as quasars. More recent studies (Barreira 2020; 2022b) based on hydrodynamic simulations which emulate galaxy formation more accurately than the simple dark matter simulations, have shown that this relationship is not necessarily verified by all types of tracers, particularly those selected by their magnitude. This complicates the measurement of PNG, since the parameter b_ϕ cannot be fixed and is completely degenerated with $f_{\rm NL}^{\rm loc}$. This degeneracy can be removed either by fixing the value of b_ϕ or by using the galaxy bispectrum (see, for instance, Moradinezhad Dizgah et al. 2021).

In the following, we will fix the value of b_ϕ with (2.120) as in Slosar et al. (2008), Castorina et al. (2019), Mueller et al. (2022), knowing that the value of b_ϕ will greatly influence the constraints obtained on $f_{\rm NL}^{\rm loc}$.

► **Improvement** Two other effects due to the presence of local PNG have been neglected in (2.104). They become non-negligible at small scales and can therefore be omitted if only large scales are taken into account. The complete additional bias due to local PNG is (Desjacques et al. 2009)

$$\Delta b\,(k, f_{\rm NL}) = \Delta b\,(k, f_{\rm NL}) + \Delta b_{\rm I}\,(f_{\rm NL}) + b_1(M)\beta_{\rm m}\,(k, f_{\rm NL})\,. \tag{2.121}$$

The different contributions are

- Δb is the scale-dependent bias correction defined in (2.104).
- Δb_I is a scale-independent correction noted but neglected in Slosar et al. (2008). This correction arises from the change in the mean number density of haloes due to the presence of the local PNG (there are more massive haloes for positive $f_{\rm NL}$) and it is given by

  $$\Delta b_I\,(f_{\rm NL}) = -\frac{1}{\sigma(M)}\frac{\partial}{\partial\nu}\ln\left[\frac{f\,(\nu, f_{\rm NL})}{f(\nu, 0)}\right]$$

  Note that the correction has an opposite sign to $f_{\rm NL}$ since the bias decreases (*resp.* increases) when the mass function is magnified (*resp.* suppressed).
- β_m quantifies the modification of the linear matter power spectrum in the presence of local PNG, since a positive value of $f_{\rm NL}$ tends to increase the power at small scales. Albeit relatively small, it is worth including if small scales are used,

  $$\beta_m\,(k, f_{\rm NL}) = \frac{P_{\rm mm}\,(k, f_{\rm NL}) - P_{\rm mm}(k, 0)}{P_{\rm mm}(k, 0)}.$$

Here, only a small value of $f_{\rm NL}$ is considered, otherwise one needs to consider also higher-order terms. The impact of each term is shown in Fig. 2.8. In particular, the scale-independent

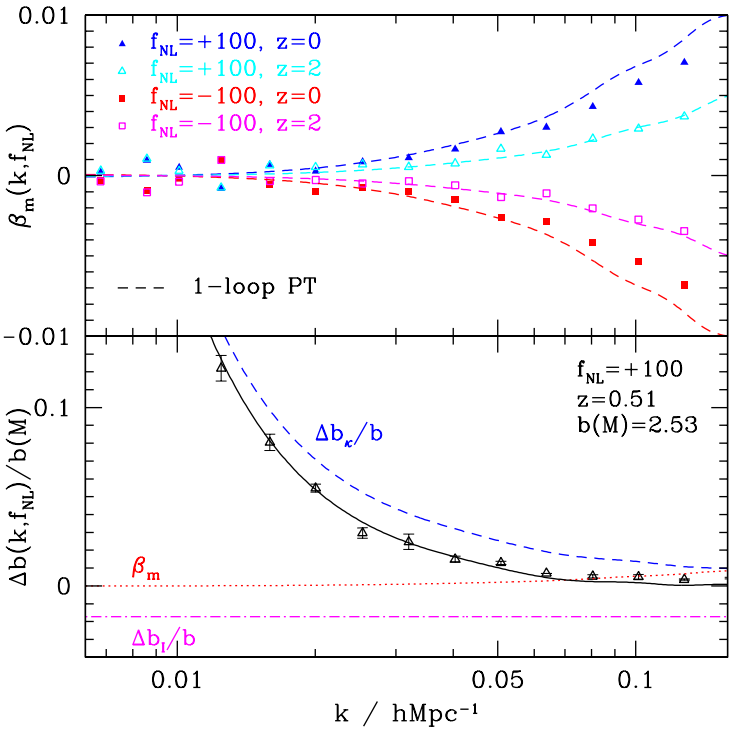


Figure 2.8: Validation of the bias in presence of local PNG in Desjacques et al. (2009). The top panel shows the relative modification β_m of the linear matter power spectrum in the presence of local PNG for $z = 0, 2$ and $f_{\rm NL} = -100, 100$. The theory is the linear order in the perturbative expansion. The different contributions to the non-gaussian bias are shown on the bottom panel at $z = 0.5$. The description given by (2.121) is accurate up to $k \sim 0.05\ h\mathrm{Mpc}^{-1}$.

correction can be absorbed into the linear bias and the modification of the linear power spectrum becomes relevant at small scales ($k > 0.05\ h\mathrm{Mpc}^{-1}$).

Remark: The scale-dependent bias relation was extensively tested with dark matter simulations over the past decade. For instance, Biagetti et al. (2017) investigate the impact of the algorithm to extract haloes from dark matter simulation. They found, in particular, a discrepancy between simulations and (2.121) with high mass haloes, in a similar range than for the quasars. The difference can be explained by the choice of δ_c which was motivated by the spherical collapse approximation. However, there is some evidence (More et al. 2011) to suggest that the Friend-of-Friend with linking length $\lambda = 0.2$[22] are not extracting spherical haloes. From Biagetti et al. (2017), δ_c has to be corrected by a factor $\sqrt{q} = 0.85$ such that $\delta_c \to \sqrt{q}\delta_c$. Hence, there is another freedom in the choice of δ_c since dark matter haloes never precisely collapse as an isolated spherical perturbation, and might be better described by elliptical collapse.

2.2.4 Current constraints

The hunt for primordial non-gaussianity began in the early 2000s with the WMAP satellite (Komatsu et al. 2003, Spergel et al. 2007, Komatsu et al. 2009, Komatsu 2010) and is still today a crucial aspect of cosmology. The different constraints are displayed in Fig. 2.9. Subsequently, the discovery of scale-dependent bias (2.105) allowed these measurements to be made with galaxy surveys, in particular with SDSS (Slosar et al. 2008, Xia et al. 2011, Ross et al. 2013, Castorina

[22] See Section 5.2.1 for further explanation.

et al. 2019, Mueller et al. 2022). It is worth noting that the constraints obtained with SDSS are of the same order of magnitude as those obtained with WMAP data.

Today, the best constraints come from WMAP's successor, the Planck satellite (Planck Collaboration et al. 2014; 2016; 2020). Unfortunately, the new CMB measurement experiments will only be able to reduce the constraints by a factor of 2 (Abazajian et al. 2016). To circumvent the cosmic variance limit of CMB observations, new large scale galaxy surveys as DESI (what I'm doing in this dissertation!) or Euclid are expected to measure $f_{\rm NL}^{\rm loc}$ with a precision of the order of five.

Another promising approach that has gained prominence in recent years is the cross-correlation between photometric surveys of galaxies and CMB lensing (Giannantonio et al. 2014, McCarthy et al. 2022, Krolewski et al. 2023). This method is particularly attractive as it avoids large scale systematic effects that complicate measurements using galaxy power spectra.

Furthermore, the recent utilization of the galaxy bispectrum has proven to be a powerful tool in constraining different forms of non-Gaussianity (D'Amico et al. 2022, Cabass et al. 2022). This method, based on the effective theory of large scale structures (see, for instance, Carrasco et al. 2012), offers a decisive solution for breaking the degeneracy between b_Φ and $f_{\rm NL}^{\rm loc}$, and could reduce the error by a factor 2 or 3 (Moradinezhad Dizgah et al. 2021). However, it introduces additional complexity to the theoretical framework.

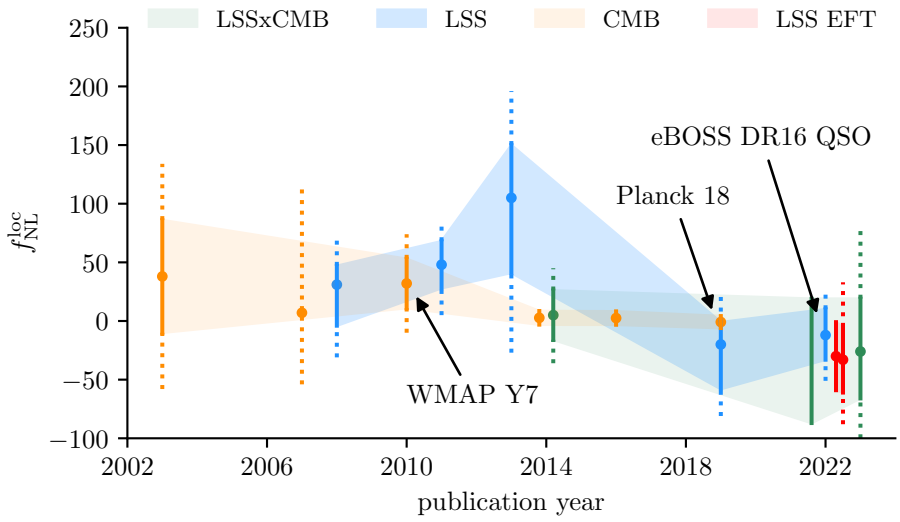


Figure 2.9: Compilation of local primordial non-gaussianity measurements. Dashed lines correspond to the 95% confidence level, while the full ones to the 68% confidence level. Shaded areas show the evolution of the 68% constraints. References to these measurements are given in the text. Inspired from Mueller et al. (2022).

Chapter 3

Galaxy survey

Contents

Il éprouvait le sentiment radieux de s'être une fois de plus emparé d'un fragment du monde ; d'avoir découpé avec son scalpel imaginaire une mince bande de tissu dans la toile infinie de l'univers.

Milan Kundera, Nesnesitelná lehkost bytí (1982)

Mapping the Universe requires acquiring the 3-dimensional coordinates of galaxies that surround us. We first use a photometric survey to detect and locate galaxies in the sky. Then, we use spectroscopy to determine their redshifts (1.24), which serve as proxies for their distances (1.26).

A photometric survey captures images of the sky with or without optical filters. Longer exposure times allow for the detection of fainter objects. The position of each object in the sky is generally given in equatorial coordinates, represented as (R.A., Dec.). The *Declination* (Dec.) is analogous to terrestrial latitude, while the *Right Ascension* (R.A.) is comparable to terrestrial longitude.

While photometric measurements are relatively quick as they only depend on the exposure time, spectroscopic measurements take significantly longer. This is because the light of each

object spread over a CCD needs to be collected in a fiber and then analyzed individually with a spectrograph. The following section discusses the evolution of spectroscopic surveys over the past four decades, while the second section provides a more detailed description of the *Dark Energy Spectroscopic Instrument* (DESI), a new generation spectroscopic survey.

3.1 History of LSS survey

The use of spectroscopy as a tool to measure distances and map the Universe started in the 1970s, but it gained significant momentum in the 2000s with the race to discover the BAO signal. Fig. 3.1 shows the evolution of the number of objects mapped by spectroscopy in the main dedicated survey as a function of time. It is well-known as the Moore law for the galaxy surveys where the number of observed galaxies increases by an order of magnitude every ten years. The color represents the mean redshift of the sample. Galaxy surveys have undergone two major revolutions: multi-fiber observation (from stage II to III) but each fiber was positioned by hand on the focal plane, and the creation of robotic positioners (from stage III to IV).

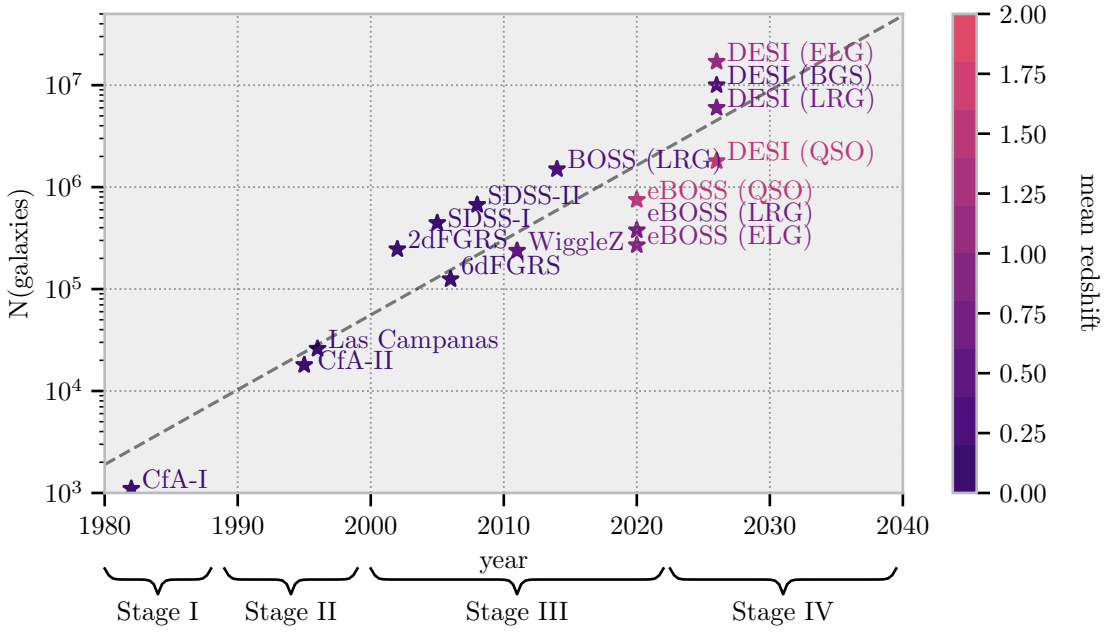


Figure 3.1: Moore law for the galaxy surveys. The number of observed objects increases by an order of magnitude every ten years thanks to technical progress. DESI will be the first spectroscopic survey of the fourth generation to be on sky.

3.1.1 Stage I / II: Knockin' on heaven's door

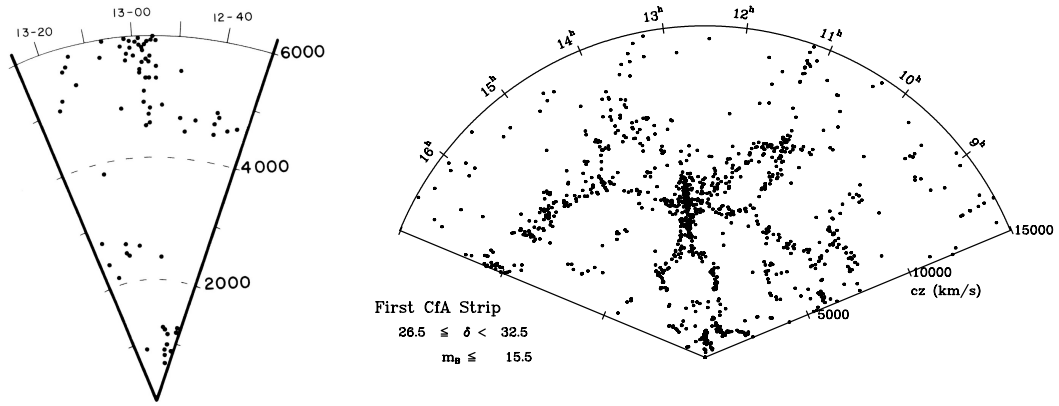
Redshift survey history began during the 1970s when high-voltage image intensifier devices replaced the photographic plates as the primary detector in telescopes. A redshift accuracy of ± 100 km s^{-1} for a galaxy of magnitude $m \sim 15$ is then reached in only 15 min compared to 2.5 hours with the previous devices.

► **First spectroscopic map** The usual representation of a galaxy survey showing data in polar coordinates, where the right ascension is used as the angular coordinate and the redshift

as the radial coordinate, appears with the Coma cluster redshift survey in Tifft and Gregory (1976). It is displayed in Fig. 3.2a. See Thompson and Gregory (2011) for useful historical references. At that time, the number of observations was too small to describe the structure of the cosmic web, but this paved the way for the study of the galaxy distribution today.

► **CfA Redshift Survey** The Center for Astrophysics (CfA) conducted the first attempt, the CfA Redshift Survey[1], to map the large-scale structures of the Universe. They achieved two surveys: CfA from 1977 to 1982 (Davis et al. 1982, Huchra et al. 1983) and CfA2 from 1985 to 1995 (Huchra et al. 1999), collecting the redshifts of 18,000 bright galaxies in the Northern sky. Data collected during these two surveys showed that galaxies were not evenly distributed in the sky but clustered. Fig. 3.2b shows the first slice of the CfA2 Survey (de Lapparent et al. 1986) which is a set of 1100 galaxies in a strip on the sky six degrees wide and about 130 degrees long. Clusters and voids can clearly be identified in this figure.

CfA2 discovered also one of the largest known structures in the observable Universe. The first Great Wall is at $z \sim 0.03$, 61 Mpc in width and about 5 Mpc in thickness. This supercluster is too large to be produced by gravitational collapse since the beginning of the Universe and is explained as artifacts of quantum fluctuations in the inflationary epoch of the Universe.



(a) **(b)**

Figure 3.2: (a) Redshift survey of the Coma cluster by Tifft and Gregory (1976). It is the first use of a cone diagram in the Astrophysical Journal. (b) Diagram showing the first CfA strip observed in 1985 by de Lapparent et al. (1986). A cluster of galaxies with the shape of a man surrounded by voids can be surprisingly identified.

3.1.2 Stage III: Multi-fiber survey

The first revolution in the spectroscopic survey is the emergence of multi-fiber surveys (a few hundred fibers) which increases drastically the number of objects which can be observed in a single night.

[1] http://tdc-www.harvard.edu/zcat/.

► **Two-degree-Field Galaxy Redshift Survey** The Australian Astronomical Observatory (AAO) performed from 1997 to 2002 the Two-degree-Field Galaxy Redshift Survey (2dFGRS[2], Colless et al. 2001) with the 3.9 m Anglo-Australian telescope. It was the largest redshift survey at that time collecting 232,155 galaxies on an area of about 1500 square degrees with a depth up to $z \simeq 0.2$. This survey allowed to accurately measure the density parameters of non-relativistic matter and, with the help of SDSS (see next paragraph), to detect the Baryon acoustic oscillations in the large-scale structures (Percival et al. 2007). Then, the AAO conducted two other smaller surveys: 6dFGS (Jones et al. 2009) and WiggleZ (Drinkwater et al. 2010).

► **Sloan Digital Sky Survey** The Sloan Digital Sky Survey (SDSS)[3] is the major astrophysical survey of the past decade starting in 2000. In particular, the program Baryon Oscillation Spectroscopic Survey (BOSS) and the extended Baryon Oscillation Spectroscopic Survey (eBOSS) produced the largest 3D map of our universe in 2020 (Ahumada et al. 2020). Previously to these two dedicated programs, the BAO peak was first detected by Eisenstein et al. (2005) in the first galaxy survey performed by SDSS. Fig. 3.3 shows the 2.3 million objects observed over 20 years divided into different target classes[4]. The associated observed BAO peaks are plotted on the right and were used in Fig. 1.19c.

Dedicated analysis are available for each tracer: Ross et al. (2015), Howlett et al. (2015) for Main galaxy sample, Alam et al. (2017) for BOSS galaxies, Gil-Maŕin et al. (2020), Wang et al. (2020), Bautista et al. (2021) for luminous red galaxies (LRG), Tamone et al. (2020), Raichoor et al. (2021), De Mattia et al. (2021) for emission line galaxies (ELG), Hou et al. (2021), Neveux et al. (2020) for quasars (QSO) and du Mas des Bourboux et al. (2020) for Lyα forest. Each tracer will be discussed in more detail in Section 3.2.3.2.

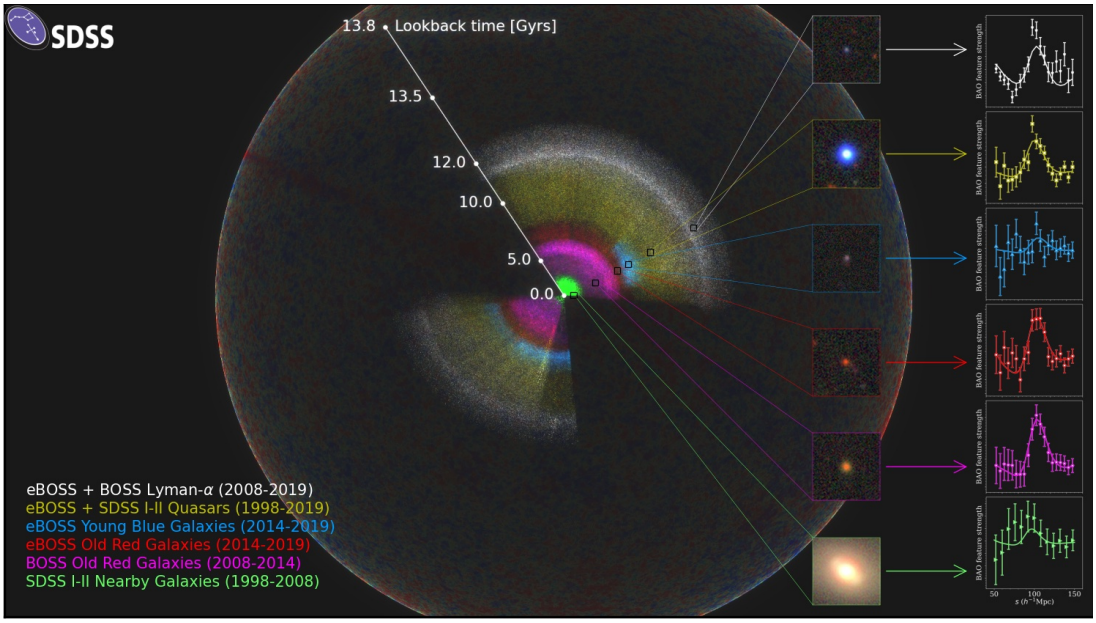


Figure 3.3: SDSS map from DR16 release after two decades of observation. Each color represents a specific target class and the associated measured BAO peak at different redshifts is plotted on the right. Credit: Anand Raichoor, Ashley Ross and the SDSS Collaboration

[2] http://www.2dfgrs.net/
[3] https://www.sdss.org/
[4] For a short walk inside: https://www.youtube.com/watch?v=KJJXbcf8kxA&t=3s.

3.1.3 Stage IV: Mapping efficiently the sky and beyond

Apart from the size of the telescope which limits the amount of light and thus the number of possible fibers, the major limitation of the BOSS/eBOSS programs was the manual positioning of fibers during each exposure, which took about 15 minutes to change the cartridge containing the optical fibers. Each of these cartridges was prepared during the day, taking 40 minutes for each. However, the fourth-generation surveys have revolutionized fiber positioning by automating the process. For instance, DESI, see Section 3.2.1.3, puts its fibers automatically in the sky in less than two minutes. To achieve the ten-fold increase in survey capabilities predicted by Moore's Law (Fig. 3.1), DESI will use 5000 fibers compared to 1000 for BOSS/eBOSS.

DESI is not the only 4th generation spectroscopic survey to emerge. For instance, the European Euclid satellite (Scaramella et al. 2022) which will operate in slitless mode, or the 4-meter Multi-Object Spectroscopic Telescope (4MOST[5], de Jong et al. 2022) which will be hosted by the VISTA telescope in the coming years.

However, the depth of the photometric surveys and the availability of targets remain a crucial limitation of the upcoming spectroscopic surveys. Hence, to probe deeper into the Universe and improve our statistical measurements, the next generation of surveys, following DESI, will require much deeper photometry to observe more distant objects. Fortunately, the Vera C. Rubin Observatory, formerly known as the Large Synoptic Survey Telescope (LSST[6], Ivezić et al. 2019), will provide an unprecedented quality of photometry.

3.2 Dark Energy Spectroscopic Instrument (DESI)

3.2.1 Overview

3.2.1.1 Context

The Dark Energy Spectroscopic Instrument (DESI)[7] is an international scientific collaboration to study dark energy through the measurement of the expansion rate of the Universe. The instrument is hosted by the Mayall telescope on top of Kitt Peak at an elevation of 21000 m. Kitt Peak National Observatory is located in the Sonoran Desert near Tuscon, Arizona, USA. The Mayall telescope is a 4-meter telescope and was built in 1973. Figure 3.4 shows the instrument with the 5000 fibers installed on the Mayall telescope.

DESI is capable of taking 5,000 simultaneous spectra over a wavelength range from 360 nm to 980 nm. The nominal survey will be conducted over 5 years and will cover an area of 14,000 deg^2 *i.e.* 1/3 of all the sky area. DESI will then observe 40 million objects covering a redshift range of 0.05 to 4, classified between bright galaxy sample (BGS), luminous red galaxies (LRGs), emission line galaxies (ELGs) and quasars (QSOs), see Section 3.2.3.2. DESI is the first stage-IV galaxy survey on sky. For comparison, DESI will collect in six months the same amount of spectra as during all the BOSS/eBOSS programs.

The R&D phase for DESI started in December 2012 and was managed by the Lawrence Berkeley National Laboratory. The instrument construction started in 2016 and the first light of the new corrector system was obtained on the night of April 1, 2019, and on the night of October 22, 2019, for the entire instrument. The commissioning phase was completed in March 2020. After a long break due to the COVID-19 pandemic, DESI had its *survey validation* during

[5] https://www.4most.eu/cms/home/
[6] https://www.lsst.org/
[7] https://www.desi.lbl.gov/

the winter of 2021 and the *nominal survey* started in May 2021. For more detailed information, see the DESI Science final design report DESI Collaboration et al. (2016a;b).

Figure 3.4: The Mayall Telescope at Kitt Peak National Observatory with DESI installed. Photo from DESI collaboration.

3.2.1.2 Science motivation

The main purpose of DESI is to constrain the expansion of the universe (1.34) and thus to constrain the amount of dark energy and its equation of state (1.30). Like BOSS/eBOSS, DESI will observe the distribution of matter in the Universe in which one will be able to measure the acoustic oscillation of baryons and the redshift space distortion as explained in Section 1.3.2.4.

By increasing by a factor of 10 the number of observed galaxies and quasars, DESI will increase the constraining power from the galaxy survey on the ΛCDM model. Fig. 3.5 gives the forecast on the BAO and RSD parameters based on the survey validation of DESI DESI Collaboration et al. (2023a). With this large number of points measured over a wide range of redshifts, and with the help of the BBN measurement to break the degeneracy of r_d (see Section 1.3.2.3), DESI should have competitive constraints on ΛCDM parameters compared to Planck (Planck Collaboration et al. 2020) alone.

In addition to the BAO/RSD parameters, the increase in statistics will allow DESI to strongly improve the measurement of the entire shape of the power spectrum pioneered by SDSS, and then to improve the constraints on several other effects such as the neutrino mass, modified gravity or primordial non-gaussianity.

This dissertation is focused on the measurement of the latter one, in order to constrain the different scenarios of inflation. The large volume probed by DESI, especially by the quasar sample, is an excellent opportunity to improve the previous measurement performed in eBOSS and reach similar sensitivity to Planck.

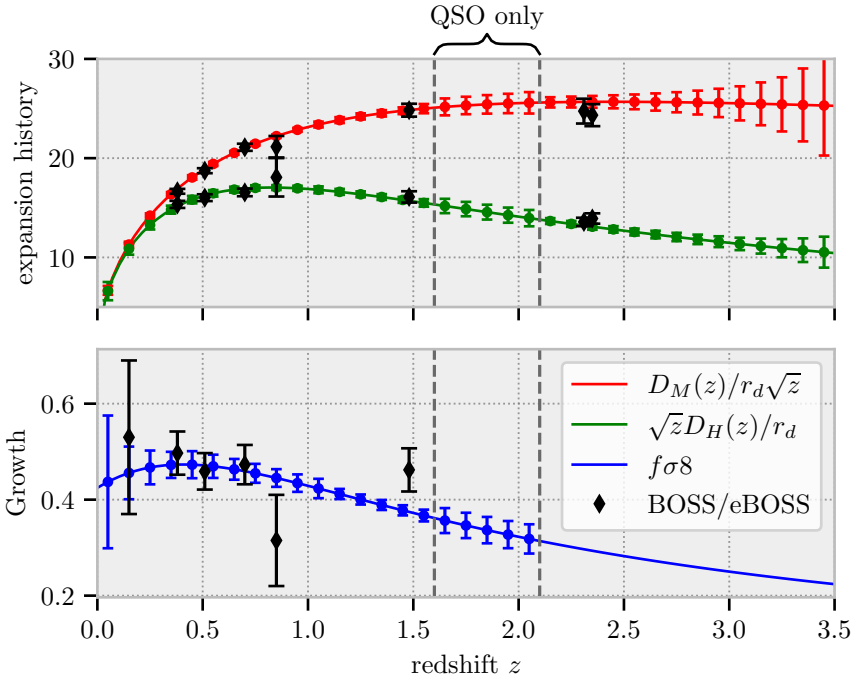


Figure 3.5: Same as Fig. 1.19c. BOSS/eBOSS points are from Alam et al. (2021) and references therein. Errors for DESI are from DESI Collaboration et al. (2023a). At low redshift, the errors are larger because the volume probed is smaller (1.27). Above $z > 2.1$, the points come from Lyα forest study and cannot provide a constraint, for the moment, on the growth rate. In DESI, QSO is the unique tracer available to probe the Universe above $z = 1.6$ either with QSOs as a direct tracer or with the Ly-α forest contained in their spectra.

3.2.1.3 Telescope & Instrument

DESI is mounted on the Mayall telescope located at Kitt Peak in Arizona (USA). It uses the existing primary mirror of the 50 years old Mayall telescope with a new prime focus corrector, specifically developed for DESI, which allows the use of 5,000 fibers over 8 deg^2. Figure 3.6 shows the position of the relevant part of DESI installed on the Mayall. A full description can be found in the instrument design report (DESI Collaboration et al. 2016b) or in the instrument overview (DESI Collaboration et al. 2022). Briefly, the DESI experiment consists of:

- *Telescope & Prime focus corrector*: As shown on Fig. 3.6, the Mayall telescope is a reflective telescope. To have a large field of view and collect the light of 5000 objects simultaneously, the prime focus corrector and the focal plane system were installed at the top of the telescope. It required the modification of the telescope by adding a large metallic ring

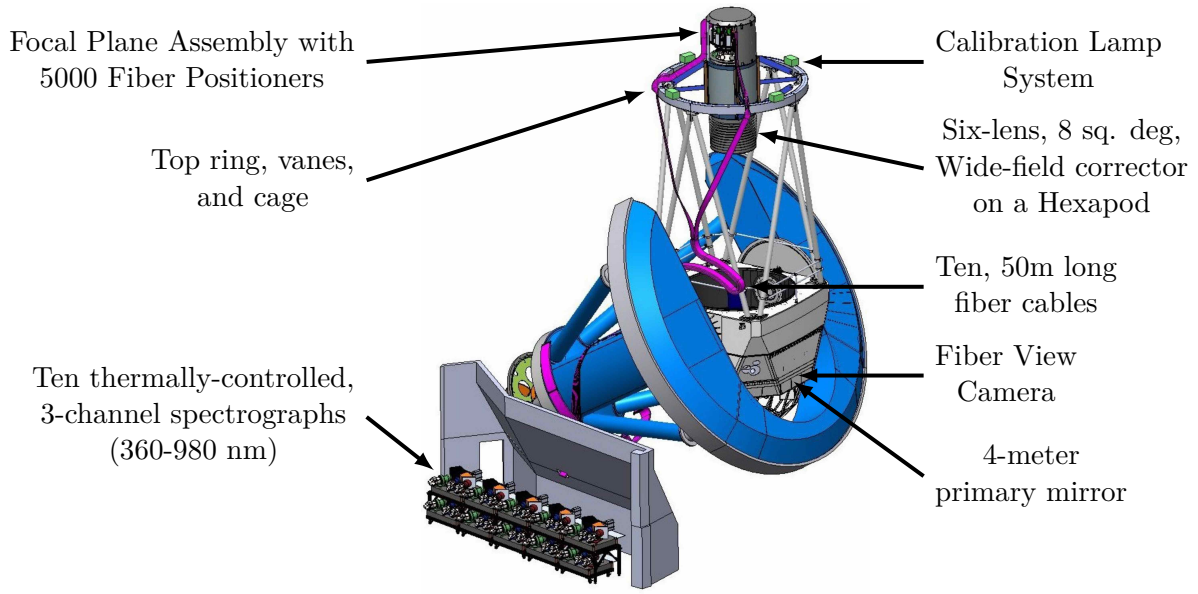


Figure 3.6: 3D CAD model of DESI installed on the Mayall Telescope. The most relevant parts of the instrument are annotated.

supported by long arms. Finally, the light is reflected by the 4-meter diameter primary mirror and goes through the prime focus corrector at the top of the telescope. After passing through the corrector, the light reaches the focal plane system and is collected by the optical fibers.

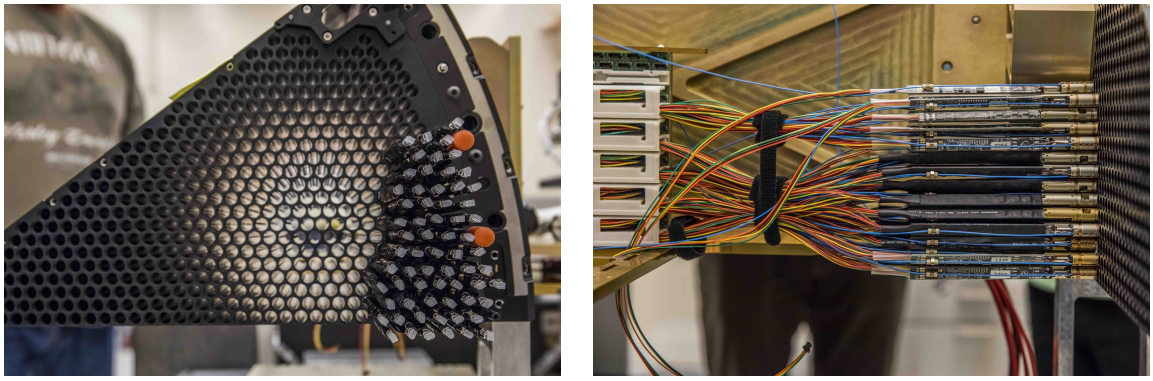
The *prime focus corrector* (Miller et al. 2023) was designed specifically for DESI and consists of six lenses, each with a diameter of approximately 1 meter. The corrector expands the field of view on the focal plane system to approximately height square degrees (DESI Collaboration et al. 2022). The last two lenses in the corrector act as an atmospheric dispersion corrector (ADC), which compensates for the wavelength-dependent atmospheric dispersion up to 60 degrees from the zenith. To align the corrector and the focal plane, the corrector is connected to the metallic ring by a hexapod. Typically, the alignment may be affected by mechanical deformation caused by the weight of the corrector. The hexapod provides up to six degrees of freedom.

- *Focal plane system*: The great advance compared to BOSS/eBOSS, the DESI ancestor, is the use of robotic fiber positioners which allow DESI to position 5,000 fibers in the field of view in less than 2 min instead of spending 15 min to change the cartridge that contains the optical fibers of BOSS/eBOSS. Hence, this step in DESI is done in a similar amount of time to read the CCDs.

  The focal plane system consists of 10 petals of 500 fibers each. Fig. 3.7 shows a petal with a bunch of robotic positioners already installed that independently control one optical fiber each. Robotic positioners are installed along a hexagonal grid with a 10.4 mm separation. Additionally, each petal is equipped with one *Guide, Focus and Alignment camera* (GFA), 10-12 field-illuminated fiducials (FIF) and two GFA illuminated fiducials (GIF). GFAs are used for the guidance, focus and alignment of the telescope by tracking reference stars. The spherical shape ensures a correct focus and alignment of the instrument. Thanks to these cameras, the telescope can follow the rotation of the sky.

  To ensure accurate positioning of the fibers on the focal plane, DESI employs a *Fiber View Camera* (FVC), which is situated at the center of the primary mirror. When fibers are positioned, they are back-illuminated and the FVC, using FIFs and GIFs as a reference, is

able to measure both the position of the fibers and any deformations caused by the prime focus corrector. This image is used to correct the position of the fibers to reach a maximal RMS error of about 5 μm, while the diameter of one fiber is 107 μm that is comparable to twice the diameter of a hair (Silber et al. 2023).



(a) **(b)**

Figure 3.7: (a) One of the 10 petals that will be placed at the focal plane of the telescope. Each of them consists of 500 optical fibers that are controlled individually and simultaneously by robotic arms. Arms with fibers are being installed on the petal. (b) Backside of the petal reveals some arms with their controlled cards and fibers. Photos from Marilyn Chung.

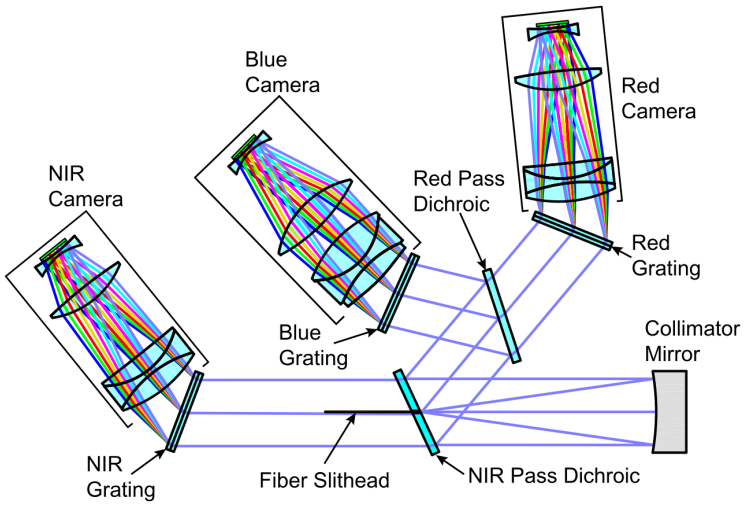
- *Spectrographs*: The 500 fibers of each petal are grouped together in one fiber cable and reach one of the ten spectrographs after running 50 meters of fiber cable. Fig. 3.8a shows the diagram of one spectrograph. The light of each fiber is separated firstly by a NIR pass dichroic reflecting the blue and the red to a red pass dichroic (Edelstein et al. 2018). At the entrance of each arm of the spectrograph, lights from fibers are spread by a volume phase holographic grating providing a high transmission over all the desired wavelengths. Finally, the infrared, red and blue light are collected by CCD sensors of 4096×4096 pixels for blue cameras and 4114×4128 pixels for red and infrared cameras. To reduce the reading noise of the CCDs, they have to work almost in a vacuum at very low temperatures. This is made possible by cryostats provided by the IRFU (CEA Saclay), which allow maintaining the temperature at 163 K (*resp.* 140 K) and 10^{-6} *mbar* for the B channel (*resp.* R and Z). Table 3.1 gives the bandwidth of each channel as well as their resolutions which are obtained by measuring the PSF of a 12 μm diameter light source.

  Fig. 3.8 shows the raw CCD images from spectrograph 6 of exposition 12127 with 1334 s exposure time. The 500 fibers are horizontal, while the increased wavelengths are vertical. Horizontal arcs correspond to the sky emission lines more prominent in the near-infrared channel. The red and near-infrared channels are more sensitive to cosmic rays. These raw CCD images need to go through all the spectroscopic pipeline `desispec`[8] to extract the spectra of the observed objects as detailed in Guy et al. (2023) before measuring their redshifts.

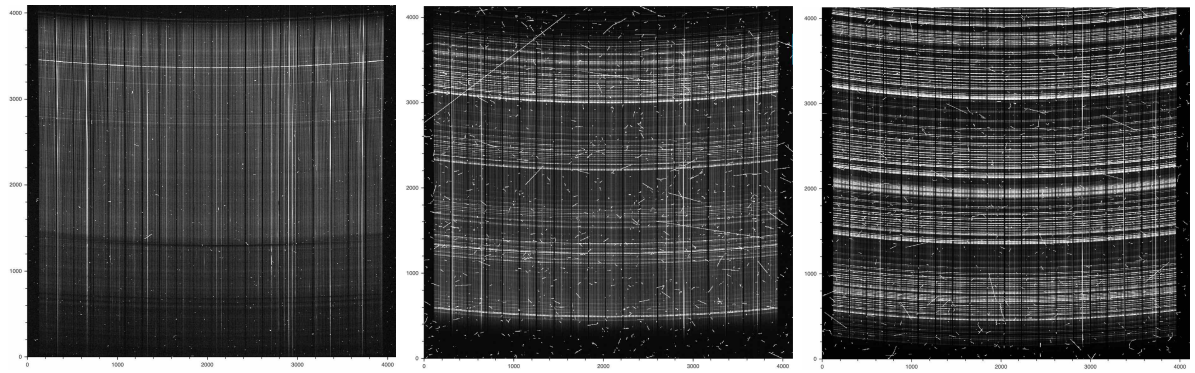
[8] https://github.com/desihub/desispec

Table 3.1: Range and resolution for the three channels (B, R, Z) of the ten DESI spectrographs. From Table 5.1 of DESI Collaboration et al. (2016b).

Channel	Range [Å]	Resolution
Blue (B)	3600 – 5930	2000 – 3200
Red (R)	5600 – 7720	3200 – 4100
Near-infrared (Z)	7470 – 9800	4100 – 5100



(a)



(b) **(c)** **(d)**

Figure 3.8: (a) Diagram of the 3 channels that constitute each of the 10 spectrographs of DESI (DESI Collaboration et al. 2016b) (b) Spectrograph 6 - Blue (B) channel (c) Red (R) channel (d) Near-infrared (Z) channel of the exposition 12127 with 1334 s exposure time and taken during the 02/04/22 night. They are raw CCD images and need to go through the spectroscopic pipeline. x-axis represents 500 fibers, while y-axis represents increasing wavelength. The horizontal arcs correspond to the sky emission lines which contaminate, at fixed wavelengths, the data. The black columns are gaps in each set of 25 fibers. Red and near-infrared channels are more sensitive to cosmic rays.

3.2.2 Observing the sky

3.2.2.1 One night of observation

Before performing one night of observation, observers check the functioning of all cameras and the instrument, and they carry out a nightly calibration of the spectrograph cameras. For this step, the dome is closed, and no external light is in the room. The calibration consists of several

steps:

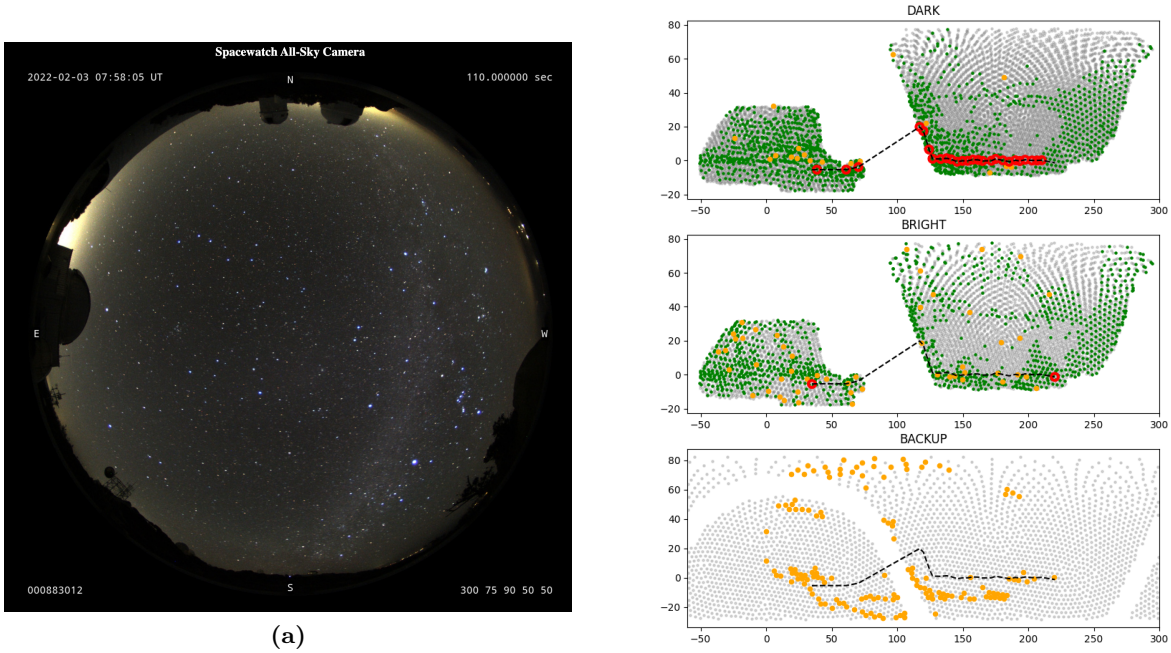
1. *Zero*: to determine the intrinsic readout noise from the bias level of the CCD depending on the position. The bias level will be subtracted from the raw images and the readout noise will be accounted in the error.

2. *Dark*: to measure the dark current (electric noise produced by the CCD) during nominal exposure time that will be subtracted from the raw images.

3. *Flat*: to quantify the response of cameras to light with a smooth spectrum at the pixel level and to normalize the incoming flux during the observation.

4. *Arc*: to quantify the response of cameras to well-defined emission lines (Ar, Cd, Hg, Kr, Ne, and Xe) and to derive the resolution of the cameras for each fiber, as well as, to model the full two-dimensional point spread function (2D PSF) for each image using the spectroperfectionism algorithm (Bolton and Schlegel 2010).

After these calibrations, the observation night can start when the Sun is 12 degrees below the horizon and will stop when the Sun reaches the opposite side, 12 degrees below the horizon. Fig. 3.9a shows a perfectly dark night during my observation shift on 02/03/2022. The yellow lights on the side are from humans: Tucson (east) and Phoenix (north) cities.



(a) **(b)**

Figure 3.9: (a) All-sky camera of a clear night at Kitt Peak National Observatory taken during my observation shift on 02/03/2022. (b) The manual plan for the same night is used only if you do not want to follow the Next Tile Selector (NTS) program. The red points are the location of the observation in this plan.

The DESI sky has been paved by tiles which are circles of 8 deg^2 in which about 5000 objects can be observed. To be able to achieve SDR[9] requirement (DESI Collaboration et al. 2016b), DESI will have to make about 5 passes at each place of the sky (Schlafly et al. 2023).

[9]Science Design report

The targets of each unobserved tile are updated every day via the *Merge Target List* (MTL, Raichoor 2023) to respect a predefined order of priority and then to allow the re-observation of objects such as high-z quasars ($z > 2.0$) for the Ly-α forests. Fig. 3.9b shows the manual plan for the 02/03/2022 night which corresponds to what should be done in the case of optimal weather. To optimize the observation time and avoid bad weather zones, DESI uses a real-time algorithm that determines the best tile to observe via the *Next Tile Selector* (NTS).

3.2.2.2 Survey validation

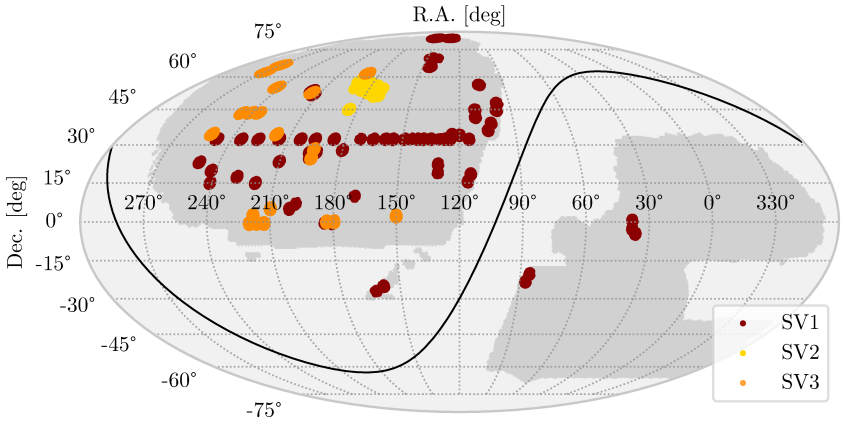


Figure 3.10: Survey Validation map; SV3 covers a wide range of regions in order to validate the main survey which should observe all the grey region above Dec. $> -20°$. Grey is the footprint of the Legacy Imaging Survey. Since DESI is in the Northern Hemisphere, it cannot cover all this area.

From December 2020 to April 2021, the collaboration has performed a survey validation (SV, DESI Collaboration et al. 2023a) to validate the choice made for the target selection, optimize the exposure time, the survey strategy and test the performance of the spectroscopic pipeline. This crucial step, before the main survey, will determine the global strategy for the next 5 years. Survey validation was done in several stages:

- SV1: dedicated observation to test, optimize and validate the target selection. In particular, tiles with important exposure time ($\sim$ 10000 s) were observed to determine the efficiency of the redshift measurement and hence derive the nominal exposure time ($\sim$ 1000 s). From these observations, the collaboration built truth tables by visually inspecting several thousand spectra (Lan et al. 2023, Alexander et al. 2023).

- SV2: optimization of the target selection before SV3 which will be used as validation for the main survey. Actually, small modifications were applied to target selection between SV2 and SV3.

- SV3: dress rehearsal for the main survey with 20% longer exposure time. Final validation and tiny optimization of the target selection. The observations are done in several rosettes which are made of a lot of tiles to observe almost every target in these areas, and thus to fully characterize the final main sample. In particular, these observations are used to derive important parameters and characteristics to tune the simulations for the first cosmological analysis based on the year 1 dataset.

Fig. 3.10 shows the different areas covered by the survey validation. All these observations

were published in the early data release (EDR, DESI Collaboration et al. 2023b) and are publicly available.

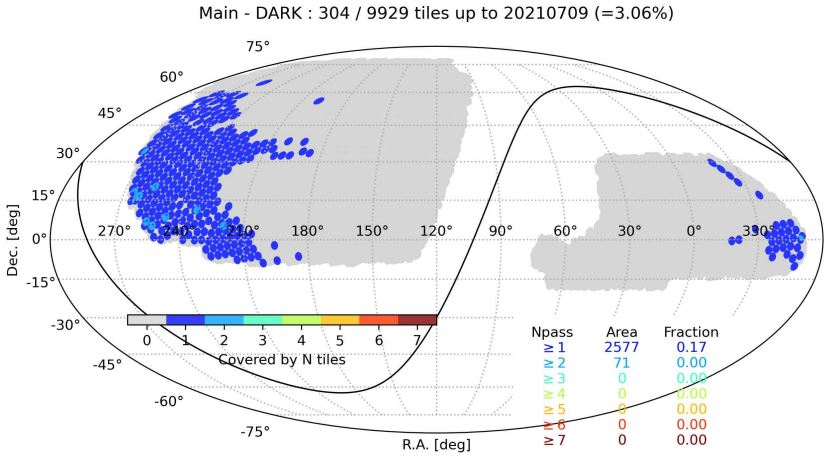
Figure 3.11: View from the Mayall telescope hosting DESI. On Friday morning 17 June 2022, the Contreras Fire reaches the summit of Kitt Peak causing the stop of the observations during the two summer months. DESI will start mapping the sky again in early September 2022. Credit: KPNO/NOIRLab/NSF/AURA.

3.2.2.3 Main survey

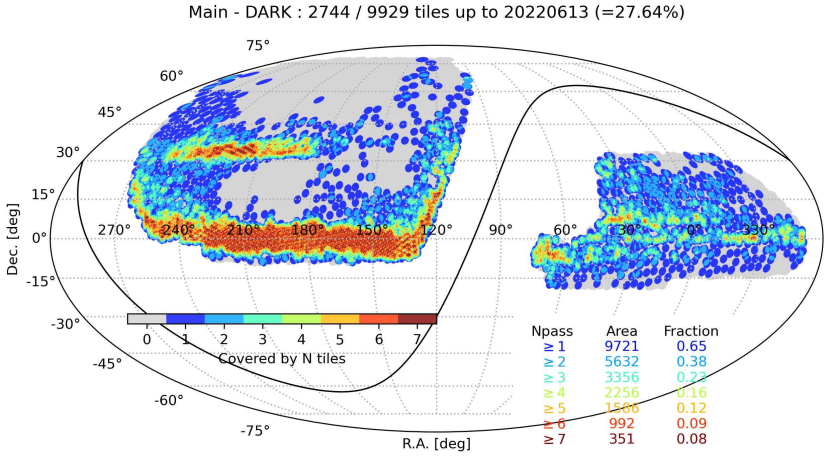
DESI started the main survey in May 2021 for a nominal period of 5 years. After a first shutdown during the summer of 2021 for a hardware upgrade, DESI outperformed the planning during its first year of observation. Unfortunately, DESI was forced to make a second shutdown due to the Contreras fire[10]. Fig. 3.11 is the view from the Mayall telescope when the fire reaches the summit of Kitt Peak. No structural damage other than ash has impacted DESI. The survey restarted in early September 2022 without a cable internet connection and with a fuel generator as a power source.

Fig. 3.12 shows the progression of the main survey for the dark time. Fig.3.12a is after the two months of observation before the first shutdown. Fig. 3.12b is just before the fire. These data will constitute the year 1 (Y1) data release, with which I perform the measurement of local primordial non-gaussianity in this dissertation. Fig. 3.12c is when I finish to write these lines, approximately 2 years after the survey began.

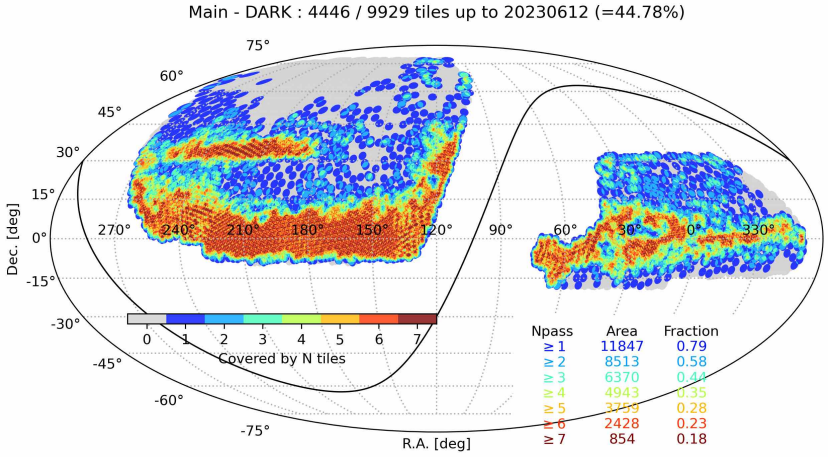
[10] https://www.desi.lbl.gov/2022/06/29/contreras-fire-threatens-desi-and-kitt-peak-national-observatory/



(a)



(b)



(c)

Figure 3.12: Number of visits per sky positions which tracks the main DESI survey (dark time) progress for three different moments: (a) End of the two first months of observations – 2021-07-09 (b) First-year coverage – 2022-06-13 (c) Today – 2023-06-12. After the nominal 5 years of the survey, DESI is expected to cover all the grey area with at least 5 tiles. These figures were generated using a script written by Anand Raichoor.

3.2.3 Mapping the Universe

3.2.3.1 DESI Legacy Imaging Surveys

To select targets for the DESI spectroscopic survey, the *Legacy Imaging Surveys*[11] program was conducted over more than 14,000 deg^2 of the sky from the Northern Hemisphere, in three optical bands : g, r and z. The optical surveys were complemented by two infrared bands from the custom "unWISE" coadds (Meisner et al. 2017) of the all-sky data of the Wide-field Infrared Survey Explorer (WISE) satellite (Wright et al. 2010), namely: $W1$ and $W2$. Table 3.2 recaps the five bands from the Legacy Imaging Surveys. DESI will use the data release 9 (DR9) for the target selection. An overview of the Legacy Imaging Surveys is available in Dey et al. (2019). The optical bands were collected via three independent programs:

- The Beijing-Arizona Sky Survey (BASS, Zou et al. 2019) observed ∼5,100 deg^2 of the North Galactic cap (NGC) in g and r using the 2.3-meter Bok telescope. The area surveyed corresponded to approximately Dec. $>$ 32.375 deg.

- The Mayall z-band Legacy Survey (MzLS, Silva et al. 2016) provided z-band observations over the same footprint as BASS using the 4-meter Mayall telescope. Because the median value of the seeing is significantly better than in the BASS data, the MzLS data are critical for deblending sources and deriving source morphology.

- The Dark Energy Camera Legacy Survey (DECaLS) was performed with the Dark Energy Camera (DECam, Flaugher et al. 2015) on the 4-meter Blanco telescope. DECaLS observed the bulk of the Legacy Imaging Surveys footprint in g, r and z. DECam was initially built to conduct the Dark Energy Survey (DES) and DECaLS expanded the DES area using publicly available DECam time. However, DES imaging is significantly deeper than standard DECaLS observations as it is covered by more exposures (more than 4 in each band).

Table 3.2: Legacy Imaging Surveys bands.

Band	range [Å]	Mean wavelength [Å]
g	3800 – 5657	4770
r	5400 – 7350	6231
z	8250 – 10150	9134
$W1$	26200 – 39500	33680
$W2$	39500 – 53900	46180

Fig. 3.13 shows the effective band passes for the three optical programs BASS/MzLS/DECaLS[12] and for the WISE satellite[13]. Although the effective band passes are quite similar, it will be necessary to take into account these slight differences to be able to compare the magnitudes of the objects measured by the three programs, in particular for the g band between BASS and DECaLZ.

A photometric survey can be described by the quality of their images, *e.g.* with the PSF

[11] https://www.legacysurvey.org/dr9/

[12] Optical throughputs can be found here: https://www.legacysurvey.org/dr6/description/, https://www.legacysurvey.org/dr7/description/

[13] WISE throughputs can be found here: https://www.astro.ucla.edu/~wright/WISE/passbands.html

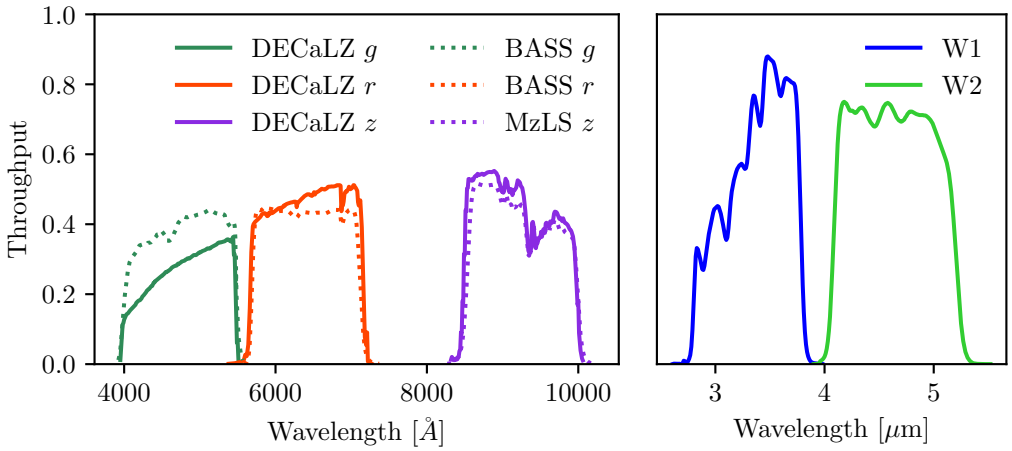


Figure 3.13: The effective band-passes used for the Legacy Imaging Surveys. *Right:* The DECaLS, BASS and MzLS effective filter throughputs for the entire system are shown. These include the transmission of the atmosphere (at a median airmass of 1.0 for BASS and MzLS and 1.4 for DECaLS), the reflectivity and obscuration of the primary mirror, the corrector transmission, and the quantum efficiency of the CCDs. Two different airmasses are used since telescopes are not on the same location. *Left:* The same for WISE. Since WISE is a satellite, the throughputs are not impacted by the transmission of the atmosphere and are therefore higher.

Table 3.3: Median values of the PSF depth and PSF size for the three imaging surveys that together constitute the DR9 Legacy Imaging Surveys. DECaLS is split according to the DES region since the quality of the photometry inside this region is significantly better. For comparison, the corresponding values for the SDSS imaging are displayed in the last row.

	PSF Depth [mag]			PSF Size [arcsec]		
	g	r	z	g	r	z
DECaLS (non DES)	24.7	24.2	23.3	1.51	1.38	1.31
DES	25.2	25.0	23.8	1.42	1.24	1.14
BASS	24.2	23.7		1.89	1.67	
MzLS			23.3			1.24
SDSS	23.13	22.70	20.71	1.44	1.32	1.29

Depth [1/nanomaggies^2] which is the 5-sigma point-source magnitude depth[14], and the PSF size [arcsec] which is the inverse-noise-weighted average of the full width at half maximum of the point spread function, also called the delivered image quality or seeing. A small value of PSF size corresponds to a good image resolution. The median values of the PSF depth and PSF size for each program are given in Table 3.3. For comparison, the corresponding values for the SDSS imaging[15] are displayed in the last row.

Fig. 3.14 shows the PSF Depth r in the Legacy Imaging Surveys and highlights three distinct regions:

1. In blue, the combination of BASS and MzLS covering the northern part (∼5,100 deg^2) of the DESI footprint (designated *North* hereafter).

[14]For a 5σ point source detection limit in band x, $5/\sqrt{x}$ gives the PSF Depth as a flux in nano maggies and $-2.5\left(\log_{10}(5/\sqrt{x}) - 9\right)$ gives the corresponding magnitude (see https://www.legacysurvey.org/dr9/catalogs/).

[15]https://www.sdss4.org/dr16/imaging/other_info/

2. In red, the DES part of DECaLS covering ∼4,600 deg^2 (designated *DES*).

3. In green, the non-DES part of DECaLS covering ∼9,900 deg^2 (designated *South*).

The region around the Large Magellanic Cloud (R.A., Dec. in the ranges $[52°, 100°]$ and $[-70°, -50°]$ respectively) is excluded in the following since it is heavily contaminated by stars. Note that DESI will only observe Dec. $> -20°$ since the Mayall telescope is in the North hemisphere.

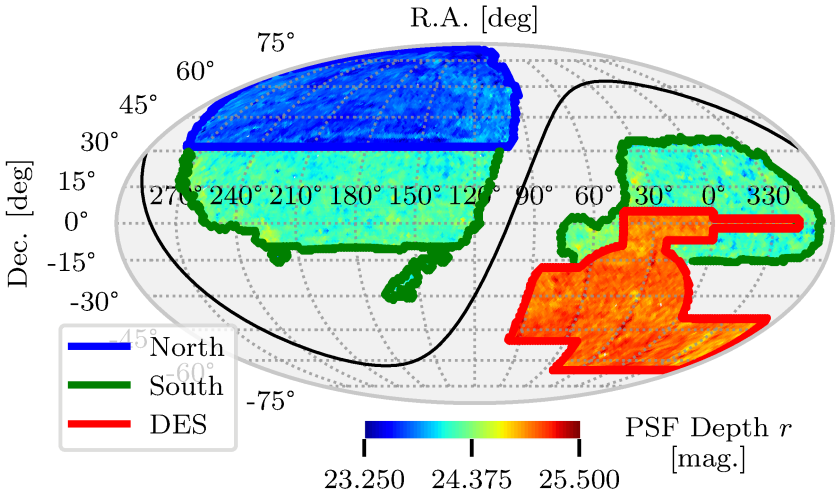


Figure 3.14: Distribution of the PSF Depth r in the DR9 Legacy Imaging Surveys footprint. The solid black line shows the Galactic plane. Three different imaging footprints are highlighted. The designations are as in Chaussidon et al. (2022). The blue region is the combination of BASS and MzLS (designated *North* in the following). The red region is the DES part of DECaLS (designated *DES*). The green region, which excludes the red and the blue regions, is the non-DES part of DECaLS (designated *South*).

3.2.3.2 DESI targets

To map the matter throughout the Universe, DESI will observe different types of sources that depict different epochs of the Universe. Fig. 3.15 shows images from the DR9 of the Legacy Imaging Surveys, which can be explored via the viewer[16]. Each source object in this figure represents as many potential galaxies or stars as DESI can observe. However, even with its 5000 fibers DESI cannot observe all these objects simultaneously. Hence, the first step of a galaxy survey is to select which source objects in the sky will receive an optical fiber when DESI will measure the redshift of these objects tranforming this 2D figure into a 3D map.

Table 3.4: Required number of targets for DESI by the science committee by deg^2

Tracer	density (# deg^{-2})	magnitude cut
BGS	1400	$r < 20.175$
LRG	615	$z < 21.5$
ELG	2400	$g < 23.5$
QSO	310	$r < 23.0$

The target selection is based on the photometric properties of each tracer and to further reduce the contamination of the sample from unwanted stars or galaxies, a limit on source

[16]Enjoy the viewer: https://www.legacysurvey.org/viewer?ra=217.6059&dec=11.9277&layer=ls-dr8&zoom=12. For DESI members, you can even consult directly the spectra of objects already observed by DESI.

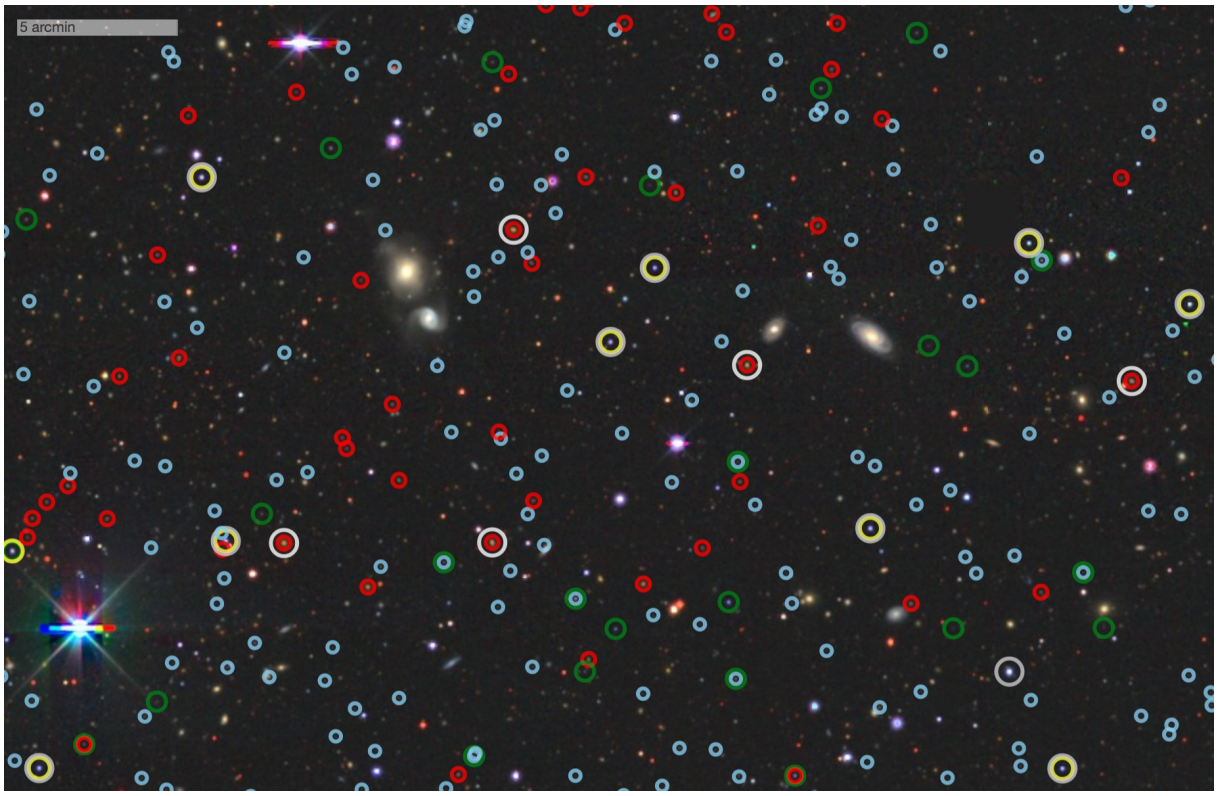


Figure 3.15: Image from the DR9 of the DESI Legacy Imaging Surveys at R.A. $\sim 217°$ and Dec. $\sim 16.2°$. The scale is given by the grey bar in the top left of the figure which corresponds to 5 arcmin. All the light points are potential targets for DESI. The target selection process selects which points will be observed during the survey. Circles are the targets for the MAIN survey. Yellow circles are stars, whites are bright galaxies, reds are luminous red galaxies, blues are emission line galaxies, and greens are quasars.

magnitude is also applied. The target selection was tuned on previous spectroscopic samples as those from the BOSS/eBOSS program, and of course, were intensively tested during the Survey Validation phases emulating a nominal observation with DESI. It will apply to the Legacy Imaging Surveys using the package `desitarget`[17] (Myers et al. 2023). Figure 3.16 gives an overview of the target selection for each tracer and Table 3.4, the target density and the magnitude limit for each tracer. In the following, a quick description of each tracer is given in order of distance from us[18]:

► **Milky Way Survey (MWS)** During the full moon, the brightness of the night sky can make difficult the observation of faint objects. To address this, DESI conducts two independent surveys: one called *bright time*, when the sky is too bright for observing faint targets, and another called *dark time* when the sky is darker and more suited for observing distant objects.

However, the number of bright galaxies, such as those in the Bright Galaxy Sample, is not sufficient for DESI to observe continuously throughout the entire nominal phase. Hence, in parallel, DESI will conduct a star survey of the Milky Way. It is expected to observe about 6.42 million stars including high-completeness samples of white dwarfs, low-mass stars within 100 pc of the Sun, and horizontal branch stars (Prieto et al. 2020, Cooper et al. 2023).

[17] https://github.com/desihub/desitarget

[18] Blog post https://www.desi.lbl.gov/2022/01/12/selecting-targets-for-the-desi-survey/ that I wrote on this topic.

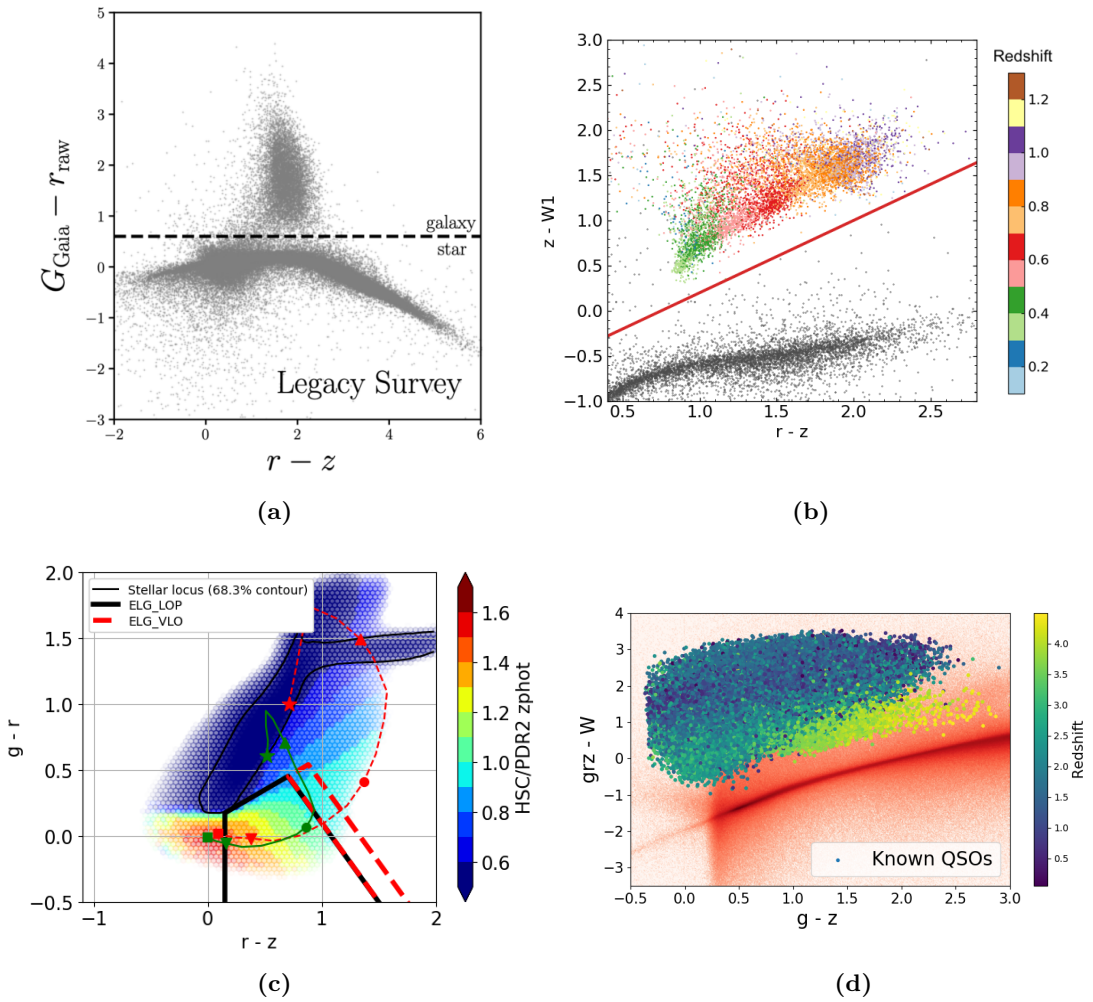


Figure 3.16: Color-color diagrams used for the target selection of the four main DESI tracers. (a) BGS: Star-galaxy separation is performed using $g_{\rm Gaia} - r > 0.6$ cut, see Hahn et al. (2022). (b) LRG: selection in DECaLS footprint. Color points are spectroscopically LRG confirmed and redshifts are from DESI spectroscopy. The stellar rejection cut is displayed in red, with point sources (almost all of which are stars) in gray, see Zhou et al. (2023). (c) ELG: each point is color-coded by the mean photometric redshift from the HSC/DR2. The ELG selection is displayed in black polygon (ELG_LOP) and the associated extension is in red (ELG_VLO). The stellar locus is the small full black line. Green (*resp.* red) track shows the evolution of star-forming (*resp.* passive) galaxy, starting from $z = 2$ (square) to $z = 0.1$ (star) via $z = 1.6$ (downward facing triangle), $z = 1.1$ (circle) and $z = 0.6$ (upward facing triangle), see Raichoor et al. (2023). (d) QSO: red are stars and blue/yellow are spectroscopically classified quasars. Here, grz (*resp.* W) is the magnitude of the weighted flux $flux(grz) = 1/2.3 \times [flux(g) + 0.8 \times flux(r) + 0.5 \times flux(z)]$ (*resp.* $flux(W) = 0.75 \times flux(W_1) + 0.25 \times flux(W_2)$), see Chaussidon et al. (2023).

▶ **Bright Galaxy Sample (BGS)** DESI survey expects to observe 13.8 million bright galaxies with a redshift range $0.05 < z < 0.4$. It is the closest matter tracer. They will be, with the MWS, the targets observed during the bright time.

Figure 3.16a shows the criterion for the BGS selection. The Star–galaxy separation is performed using the $g_{\text{Gaia}} - r$ cut, where g_{Gaia} is the photometry from ESA GAIA mission[19] (Prusti et al. 2016). The criterion exploits the fact that the Gaia magnitude is measured with an aperture of a space-based point-spread function (PSF) while the Legacy Imaging Surveys magnitude captures the light from the entire source. This cut separates point sources (stars) from extended sources (galaxies), and the selection will observe about 850 targets per square degree of the *bright* sample ($r < 19.5$), and with additional cut to ensure a high redshift efficiency, about 550 targets per square degree of the *faint* sample ($19.5 < r < 20.175$) (Ruiz-Macias et al. 2020, Hahn et al. 2022).

▶ **Luminous Red Galaxy (LRG)** From here, the tracers will be observed during the dark time. DESI uses the expected 7.46 million LRGs to probe the Universe in the redshift range $0.4 < z < 1$. These are the most massive galaxies, composed largely of old stars. Their red color makes them easy to select in imaging. In particular, stars are rejected through $z - W1$ vs. $r - z$, the redshift is controlled through $g - r$ vs. $z - W1$ and the density is tuned through $r - W1$ vs. $W1$.

The LRG selection is done by the cut (see red line in Figure 3.16a) in the $(r - z)$–$(z - W1)$ space. This selection uses the W1 infrared band to separate galaxies (color points) from stars (grey points). The different colors show the redshift of the galaxies in the color–color space. This selection will observe about 615 targets per square degree (Zhou et al. 2020; 2023).

▶ **Emission Line Galaxy (ELG)** DESI uses ELG for the redshift range $0.6 < z < 1.6$ and should collect 15.7 million. This is the largest sample in DESI. These are fainter and more distant, but their vigorous star formation and hot young stars create strong emissions in distinct wavelengths that DESI can easily detect.

Figure 3.16c shows the ELG target selection. The color histogram is the redshift distribution of ELGs in the color space. To avoid stellar contamination (black smooth line), ELG selection is based on a cut in $(g - r)$-$(r - z)$ space. Since the objective of the DESI ELG sample is to provide the tightest cosmological constraints over the $0.6 < z < 1.6$, favoring as much as possible the high-z part of the sample ($1.1 < z < 1.6$) where other DESI tracers are the least dense, the ELG target selection is composed of two disjoint subsamples. ELG_LOP (low priority), selected by the black polygon, will be for the $1.1 < z < 1.6$ subsample with a density of about 1940 deg^{-2}, while ELG_VLO (very low priority), selected by the red dashed polygon will be for the $0.6 < z < 1.1$ with a density of about 460 deg^{-2} and have a lower priority during the observation than ELG_LOP (Raichoor et al. 2020; 2023).

▶ **Quasar (QSO)** Then to go further and probe the young Universe, DESI uses quasars. Quasars or quasi-stellar objects (QSO) are forming galaxies where the central supermassive black hole is accreting large amounts of gas that are glowing as it reaches relativistic speeds. Quasars will be used either as direct tracers of the matter or to study Ly-α forest with high-z QSOs such that $z > 2.0$. Ly-α forests are clouds of neutral hydrogen in the line of sight of a quasar. They absorb the continuum light from the quasar that is redshifted to the Ly-α

[19] https://sci.esa.int/web/gaia

wavelength by the Universe expansion. These absorption lines track the matter distribution at high redshift. DESI is expected to observe 2.87 million quasars.

Only point sources are considered during the selection and the target selection aims to separate quasars from stars. However, the separation between these two classes is less obvious than in the previous cases. A more sophisticated selection has to be applied. Quasars are selected via a Random Forest classification instead of the classical color cuts selection done in previous spectroscopic surveys. The idea of the selection is to separate quasars from stars based on the "infrared excess" of QSOs. In Figure 3.16d, the stellar locus is illustrated by the red line and the quasars by the blue/green/yellow points. This selection will observe about 310 targets per square degree (Yèche et al. 2020, Chaussidon et al. 2023). See Section 4.2 for a detailed description.

3.2.3.3 Redshift determination

After observing the targets, the next step is to classify them and extract their redshifts. The classification is necessary since the selection process is not perfect and some unwanted objects, which have no cosmological information, may have been included *e.g.* stars into the quasar sample. To test and improve the accuracy of the automated classification and redshift measurement algorithms, a truth table is constructed by visually inspecting several thousand spectra, see Lan et al. (2023) for galaxies and Alexander et al. (2023) for quasars.

After all the spectroscopic processing (Guy et al. 2023), DESI will use `redrock`[20] (Bailey 2023) to classify and measure the redshifts of spectra by fitting templates from principal component analysis (PCA) to the data. Redrock is inspired by the previous algorithm used in SDSS (Bolton et al. 2012) and was already used during the eBOSS program to improve the redshift estimate for LRG and ELG targets (Ross et al. 2020). Redrock performs a redshift scan and selects the minimum χ^2 as the best fit spectral classification and redshift. For this purpose, it uses four sets of templates, one for stars, one for galaxies and two for quasars: one with $0.05 \leq z < 1.6$ and the second covers $1.4 \leq z < 7.0$ (see Brodzeller et al. 2023, for the QSO template performance). Note that up to *Guadalupe* version, Redrock used only one quasar template.

Figure 3.17 shows typical spectra for the five main target classes observed with DESI. BGS is characterized by various and strong emission lines, LRG by their absorption lines and the global form of their spectra, ELG by the characteristic [OII] doublet emission, and QSO by their broad emission lines. The Ly-α forest is blueward of the Ly-α emission peak. Redrock will simply match these different spectra with the different sets of templates trying different redshifts.

3.2.3.4 Redshift distribution

An important result of the survey validation, see Section 3.2.2.2, is the determination of the redshift distribution of each tracer, since it enables the use of DESI targets in the cross-correlation with the CMB lensing map (see, for instance, Krolewski et al. 2023).

Fig. 3.18 shows the redshift distributions for the 4 main tracer of DESI. These distributions are for the sample that will be used for the clustering analysis, as explained in DESI Collaboration et al. (2023b). Unlike other tracers, QSOs are not very dense but cover a wide redshift range!

[20] https://github.com/desihub/redrock

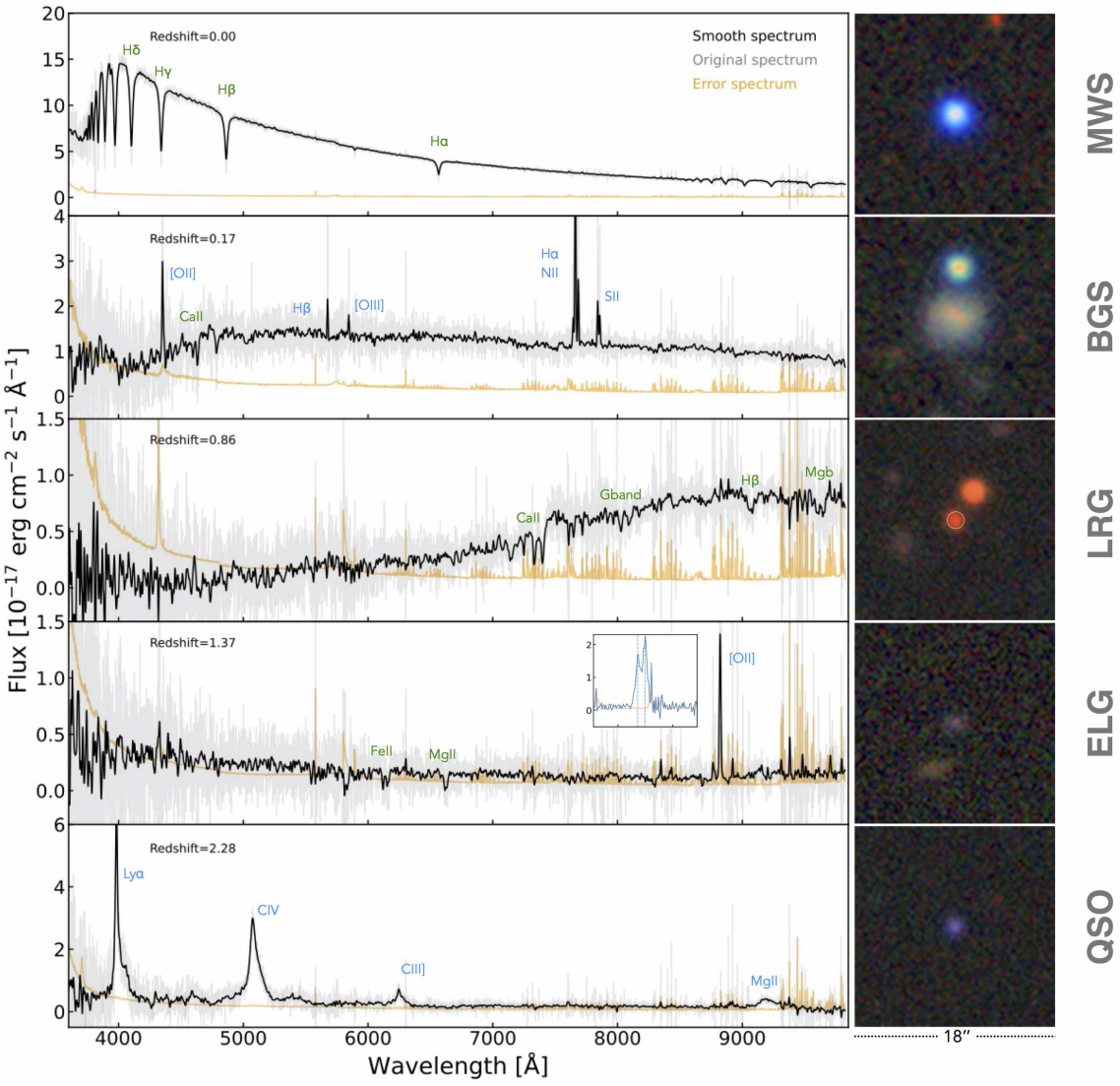


Figure 3.17: Spectra for the DESI five main target classes observed during the survey validation. From top to bottom: stars and bright galaxies are collected during the bright time, luminous red galaxies, emission line galaxies and quasars are collected during the dark time. These spectra were collected during SV1 with a very long exposure time, typically 10 times longer than the nominal exposure time. Credit: Ting-Wen Lan and DESI Collaboration.

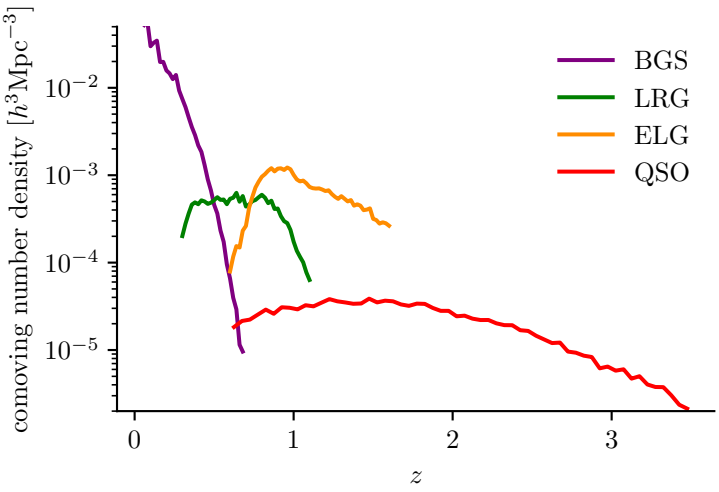


Figure 3.18: $n(z)$ distributions for the 4 main DESI tracer measured during the survey validation. Here, only good objects for the clustering measurement are considered as explained in DESI Collaboration et al. (2023b). The ELG distribution contains the two sub-selection ELG_LOP and ELG_VLO.

Chapter 4

DESI quasar survey

Contents

Une mathématique bleue,
dans cette mer jamais étale
D'où me remonte peu à peu
Cette mémoire des étoiles
Léo Ferré, La Mémoire et la Mer (1970)

In the past two decades, quasars (a.k.a. quasi-stellar objects, or QSOs), have become a key ingredient in our understanding of cosmology and galaxy evolution. Being among the most luminous extragalactic sources, they have become a mainstay of cosmological surveys such as the 2dF Quasar Redshift Survey (2QZ; Croom et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), where they are the privileged targets to study large-scale structures at high redshift.

As part of the third-generation of the Sloan Digital Sky Survey (SDSS-III; Eisenstein et al. 2011), the Baryon Acoustic Oscillation Survey (BOSS; Dawson et al. 2013) measured the spectrum of about 300,000 quasars, 180,000 of which are at $z > 2.15$, to a limiting magnitude of $g \sim 22$. As part of SDSS-IV, the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al. 2016) has observed 350,000 quasars with redshifts of $0.8 < z < 2.2$ to $g \sim 22.5$, in addition to targeting 60,000 new quasars at $z > 2.2$ (Lyke et al. 2020). DESI is aiming to quadruple the number of known quasars and to obtain spectra of nearly three million quasars, reaching limiting magnitudes $r \sim 23$.

DESI will use this sample to measure, see Section 1.3.2.4, the scale of baryon acoustic oscillations (BAO) and the growth of structure through redshift-space distortions (RSD) that was pioneered for the quasars by eBOSS studies (Zarrouk et al. 2018, Hou et al. 2021, Neveux et al. 2020).

Note that the Section 4.2 and 4.3 are from Chaussidon et al. (2023) and Chaussidon et al. (2022).

4.1 Quasi-stellar object

4.1.1 A particular active galactic nuclei

Quasars are a specific class of active galactic nuclei (AGN). The first optical spectrum of active galactic nuclei was collected in 1908 at Lick Observatory by E.A. Fath who noted the presence of strong emission lines in the nebula NGC 1068. Then, these objects were first classified by Carl Seyfert in 1943.

Several decades later, Schmidt (1963) observed for the first time a new type of object that will be called quasars for quasi-stellar objects, since they have the particularity to be point sources, (see Kellermann 2014, for a historical review). Quasars are characterized by the following attributes: they are point sources with radio sources, have time-variable continuum flux, large

UV flux, broad emission lines and large redshifts (see Peterson 1997, for a complete introduction), whereas QSOs stand for similar characteristics than quasars but without radio emission. Fig. 4.1 shows the modern classification of active galactic nuclei. Note that, in the rest of this dissertation, QSOs and quasars will be identified as a single type of active galactic nuclei, since we will not be able to separate these two types.

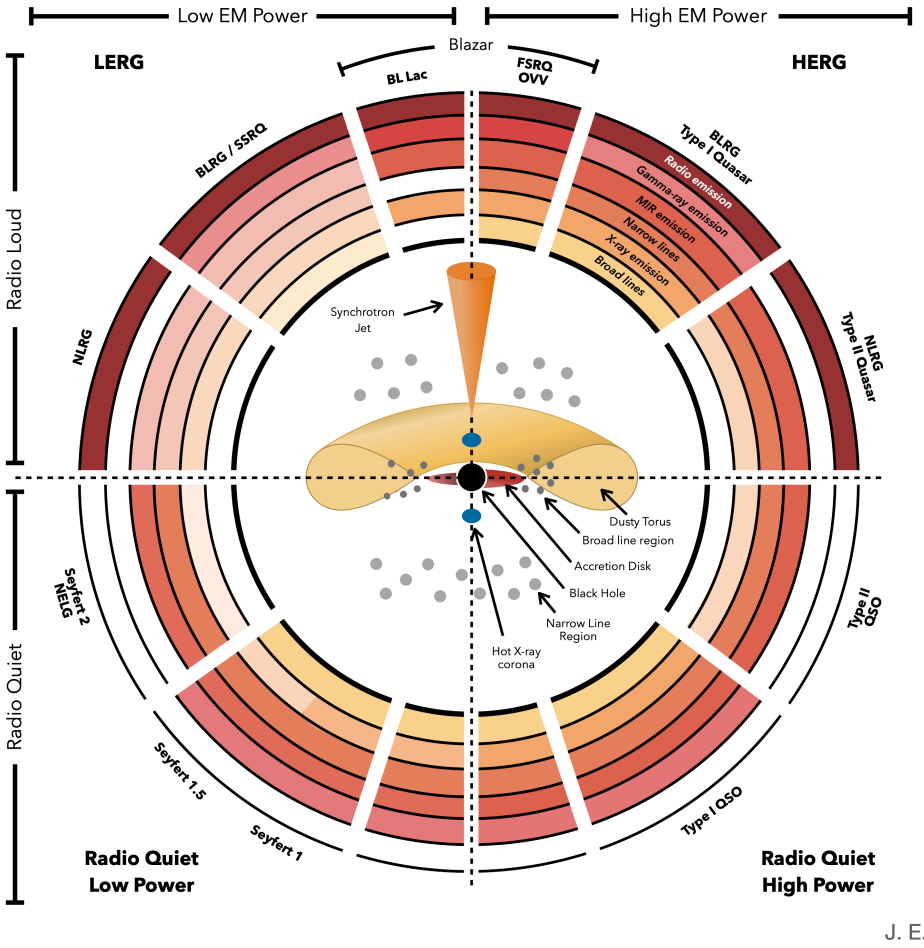


Figure 4.1: Schematic active galactic nuclei classification (Thorne et al. 2022). The type of object seen depends on the viewing angle, whether or not the AGN produces a significant jet (radio-loud and radio-quiet), and the rate of accretion onto the central supermassive black hole. The center of the schematic shows the typical components of an AGN, although the geometry of many of these components is still unknown. Some of the most commonly used names are displayed for different classes of AGN including broad line radio galaxy (BLRG), narrow line radio galaxy (NLRG), narrow emission line galaxy (NELG), flat spectrum radio quasar (FSRQ), steep spectrum radio quasar (SSRQ), optically violent variables (OVV), and quasi-stellar objects (QSO). Here, quasars and QSOs are separated into radio-loud and radio-quiet objects, however, these names are often used interchangeably. The transparency of the colour in each ring corresponds to the increasing strength or prevalence of a particular emission type.

Since Quasars are the brightest object in the sky at every wavelength, they can be observed at very great distances from us *i.e.* at very high redshift, and thus allow us to probe the distant Universe.

Fig. 4.2 shows a Ly-α quasar spectrum with a high signal observed by DESI. This spectrum exhibits a predominant emission line at $\lambda = 1215.67$Å due to the Ly-α emission. The peculiar shape of the spectrum with these broad bands makes automatic classification easy. However, the broadness of the emission lines also makes the redshift measurement less accurate. See also,

Fig. 4.9 for other quasar spectra.

The Ly-α forests, blueward the emission line, become visible to DESI spectrographs from $z \approx 2$ and are a valuable source of information to probe the Universe at this redshift (du Mas des Bourboux et al. 2020). In particular, the use of Ly-α forest helps the constraint on the neutrinos mass (Palanque-Delabrouille et al. 2015; 2020)

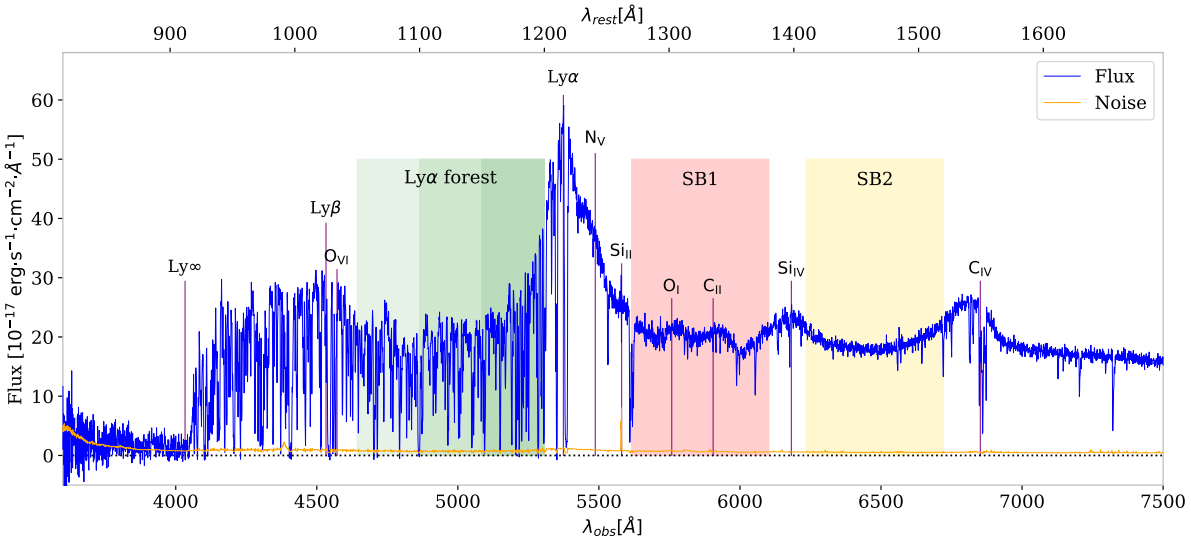


Figure 4.2: High-signal spectrum of a quasar located at a redshift $z = 3.42$ measured by DESI with an exposure time of $2,300$ seconds (Ravoux et al. 2023). This quasar was observed on 12th April 2021, in the SV program, on DESI tile 221 (TARGETID = 39627746095137037, RA = 217.263 °, DEC = −1.755 °). The quasar flux is represented in blue and its noise is in orange. The Ly-α forest is shown in green. The region in red and yellow are used for the Ly-α forest study.

4.1.2 A fabulous tracer for high-z Universe

Under Gaussian approximations, the statistical errors on the power spectrum can be expressed as the sum of the limitations of the finite volume of the survey, known as the *sample variance*, plus the fluctuation due to the sampling of the underlying density field, known as the *shot noise*. Since the observable Universe is finite, the sample variance cannot be reduced to zero. This is known as the *cosmic variance*.

Finally, the statistical errors are (Feldman et al. 1994)

$$\frac{\sigma_P}{P} = 2\pi \sqrt{\frac{2}{V k^2 \Delta k \Delta \mu}} \left(\frac{1 + nP}{nP} \right), \tag{4.1}$$

where V is the volume of the survey, n is the comoving number density of the considered tracer, and $1/n$ is the shot noise from the Poisson sampling of the density field when n is supposed constant in position. As explained in Section 1.3.1.2, the power spectrum (1.116) depends on k, μ. Thus, Δk (*resp.* $\Delta\mu$) is the bin width in k (*resp.* μ) and P is the average value of the power spectrum in $\Delta k, \Delta\mu$. The number of independent Fourier modes is given by $2 \times 2\pi k^2 \Delta k \Delta\mu V/(2\pi)^3$. The additional factor two arises because the density field is real and so the Fourier modes $\mathbf{k}$ and $-\mathbf{k}$ are not independent.

When the shot noise contribution becomes dominant over the sample variance *i.e.* $nP \ll 1$,

the tracer is said *shot noise limited*, as for the quasars, and the errors reduce as

$$\frac{\sigma_P}{P} \propto \frac{1}{\sqrt{V}} \frac{1+nP}{nP} \longrightarrow \frac{1}{\sqrt{V}} \frac{1}{nP}. \tag{4.2}$$

In this case, increasing the density of the tracer, typically increasing the number of targets, is a direct gain for the power spectrum measurement and then for the cosmological inferences. In the opposite case ($nP \gg 1$), the errors depend only on the volume of the survey and on the value of the power spectrum.

To quantify, the statistical power of each tracer in DESI, one can calculate the associated effective volume, as in Anderson et al. (2014), defined by

$$V_{\text{eff}} = \sum_i \left(\frac{\bar{n}\left(z_i\right) P_0}{1+\bar{n}\left(z_i\right) P_0} \right)^2 \Delta V\left(z_i\right), \tag{4.3}$$

where P_0 is computed at the effective redshift of the shell at $k = 0.1\ h\text{Mpc}^{-1}$, the redshift distributions are displayed in Fig. 3.18 and we consider a nominal sky coverage of $14,000\text{deg}^2$.

The result for the three main dark tracers of DESI is given in Table 4.1. Note that the computation does not take into account the bias evolution as a function of the redshift (see, for instance, Laurent et al. 2017).

Table 4.1: Effective volume for DESI tracers computed as in Anderson et al. (2014). The densities are from the Y1 catalog which will be used for the clustering measurement. For simplicity, we compute P_0 at the mean redshift of the sample.

Tracer	Density $[\text{deg}^{-2}]$	z range	z mean	bias	V_{eff} $[(\text{Gpc}/h)^3]$
LRG	537	$0.4-1.1$	0.7	2.3	19
ELG	2400	$0.6-1.6$	1.1	1.3	39
QSO	188	$0.8-3.1$	1.7	2.6	107

QSOs have the biggest effective volume and are therefore expected to have the best sensitivity to measure the power spectrum at large scales. In order to use QSOs as a tracer to measure the primordial non-gaussianity with the best possible sensitivity, we will pay particular attention to their target selection.

4.2 Quasar target selection for DESI

Because of their point-like morphology and with photometric characteristics that mimic faint blue stars in optical wavelengths, especially for the Ly-α QSOs, the QSO selection is challenging. Successful selection of a highly-complete and pure QSO sample are usually based on their UV excess (Richards et al. 2002, Ross et al. 2012). In DESI, we propose an alternative approach that detects their near-infrared excess as already demonstrated in eBOSS (Myers et al. 2015). Indeed, we use three optical bands (g, r, z) combined with WISE infrared photometry in the $W1$ and $W2$ bands to select our primary sample of QSOs. QSOs are ~ 2 mag brighter in the near-infrared at all redshifts compared to stars of similar optical magnitude and color, providing a powerful method for discriminating against contaminating stars.

In order to test the different target selection approaches and to optimize the exposure time for each target class before beginning five years of DESI operations (hereafter main survey), DESI has performed a Survey Validation (SV), organized in two phases with separate goals.

The first phase of SV, completed in four months, allowed us to optimize the selection algorithms, estimate the redshift distributions, and evaluate the projected cosmology constraints. It provided spectra over 45 fields containing a mix of luminous red galaxy targets, emission line galaxy targets and quasar targets. Among them, 42 fields have a total effective exposure time of $T_{\rm eff} \sim 4000$s and 3 fields correspond to ultra-deep observations ($T_{\rm eff} \sim 10,000$s). The latter observations have been visually inspected (VI) and those three fields provide a control sample to study the target selection.

The second stage, the 1% survey consisted of a full clustering program covering about 1% of the DESI survey with fiber assignments similar to the main survey and exposure times $\sim 30\%$ longer than the nominal exposure time ($T_{\rm eff} \sim 1000$s) projected for the main survey. It lasted approximately one month. During the second phase of SV, we used the final quasar selection (hereafter "main selection") that was optimized during the first phase of SV.

In the following, All magnitudes, will be quoted on the AB system, including magnitudes from the Wide-field Infrared Survey Explorer which are often given on the Vega system. In addition, except when mentioned otherwise, all computations with `HEALPix` pixels are done with $N_{\rm side} = 256$ (a pixel area of $\sim$0.05 deg^2) and all maps are plotted in a Mollweide projection with a `HEALPix` resolution of $N_{\rm side} = 64$.

4.2.1 Main Quasar Target Selection

In this section, we describe the target selection used in the 1% survey and in the main survey. This selection corresponds to bit 2 (`QSO`) of the maskbits `SV3_DESI_TARGET` and `DESI_TARGET` described in Myers et al. (2023).

4.2.1.1 Overview of the sample

The DESI survey uses QSOs as point tracers of the matter clustering mostly at redshifts lower than 2.1, in addition to using QSOs at higher redshift as backlights for clustering in the Ly-α forest. This approach enlarges the role of QSOs relative to the BOSS project (Ross et al. 2012), which only selected QSOs at $z > 2.15$ for use via the Ly-α forest, and enhances their role relative to eBOSS (Myers et al. 2015), where QSOs are used in a similar fashion as in DESI although with lower densities.

In DESI Collaboration et al. (2016b), based on the quasar luminosity function (QLF) of Palanque-Delabrouille et al. (2016), we inferred that a complete QSO sample, brighter than magnitude $r = 22.7$, would contain about 190 QSOs per deg^2 at $z < 2.1$ and about 70 at $z > 2.1$. Assuming a minimum efficiency of about 65%, the goal of DESI was to obtain the redshifts for 120 and 50 QSOs per deg^2 in the redshift ranges $z < 2.1$ and $z > 2.1$, respectively. With the Survey Validation during which we were able to test several extensions of our selection, we demonstrated that we can significantly exceed these statistics without significantly inflating our target budget (see Sec. 4.2.3). Therefore, in the main selection presented in this Section, we use a magnitude limit of $r = 23.0$ for an average density of ~ 310 targets per deg^2.

4.2.1.2 Strategy for the Selection

QSOs commonly exhibit hard spectra in the X-ray wavelength regime, bright Ly-α emission in the rest-frame UV, and a power-law spectrum behaving as $F_\nu \propto \nu^\alpha$ with $\alpha < 0$ in the mid-infrared bands (Stern et al. 2005, Donley et al. 2012). In the mid-optical colors, QSOs at most redshifts are not easily distinguished from the much more numerous stars. Successful

selection of a highly-complete and pure QSO sample must make use of either UV or infrared photometry. With the extended (WISE) mission that more than quadrupled the exposure time of the original (WISE) all-sky survey, and in the absence of any 'u'-band imaging over the whole DESI footprint, we decided to rely upon optical and infrared photometry for QSO selection.

Therefore, the DESI QSO target selection is a combination of optical-only and optical+IR colors. In order to illustrate this strategy, we use two colors, $grz - W$ vs. $g - z$ where grz is a weighted average of the grz band fluxes with flux(grz) = [flux(g) + 0.8×flux(r) + 0.5×flux(z)] / 2.3 and W a weighted average of $W1$ and $W2$ fluxes with flux(W)=0.75×flux$(W1)$+0.25×flux$(W2)$. In the Legacy Imaging Surveys (Dey et al. 2019), the conversion from linear fluxes to magnitudes is $m = 22.5 - 2.5\log_{10}(\text{flux})$. Fig. 4.3 shows the bulk of the QSO targets which are identified in an optical+IR selection where the excess infrared emission from QSOs results in a clear segregation from stars with similar optical fluxes. Stellar SEDs indeed sample the rapidly declining tail of the black-body spectrum at those wavelengths, where QSOs have a much flatter SED than stars. This method was previously demonstrated in eBOSS and Fig. 5 of Myers et al. (2015) exhibits the same separation between stars and QSOs thanks to WISE imaging.

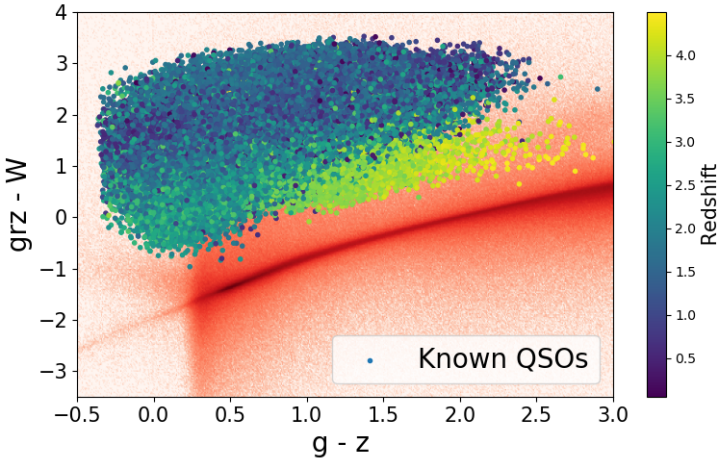


Figure 4.3: Colors in the optical or near-infrared of objects photometrically classified as stars (red) or spectroscopically classified as QSOs (from blue to yellow dots, depending on their redshift). The color $grz - W$ allows us to reject stars based on the "infrared excess" of QSOs.

4.2.1.3 Selection with Random Forest Algorithm

Neural-network-based algorithms implemented in BOSS (Yèche et al. 2010) were found to increase QSO selection efficiency by ≈ 20% compared to color cuts. Similarly, to improve the success rate for DESI, we use a machine-learning algorithm based on Random Forests (RF).

First, before utilizing the RF, we restrict the selection to objects with stellar morphology ('PSF' in DR9), to avoid an almost 10-fold contamination by galaxies that otherwise enter our selection region, and we impose $16.5 < r_{\text{AB}} < 23.0$. In addition, to reject stars, we apply a cut on the (WISE) magnitudes ($W1 < 22.3$ and $W2 < 22.3$). This cut is particularly efficient at getting rid of stars in the Sagittarius Stream, a region which exhibits an overdensity of QSO targets (see Fig. 4.4 and discussion in Section 4.3.2). We also require that the targets are not in the vicinity of bright stars, globular clusters, or large galaxies. Such "masked" sources have `MASKBITS`[1] of 1, 12 or 13 set in Legacy Surveys catalogs.

[1] https://www.legacysurvey.org/dr9/bitmasks/

Then, we train the RF using two samples: one of 'QSOs' similar to the objects we want to select and the other of 'stars' we want to discriminate against. The QSO sample consists of 332,650 known QSOs in the DESI footprint. The vast majority of those QSOs with $17.5 < r < 23.2$ are selected by their intrinsic time-variability in the SDSS Stripe 82 (an equatorial stripe in the South Galactic Cap defined by SDSS), using the method described in Palanque-Delabrouille et al. (2011) with SDSS light curves. This selection provides a training sample of QSOs that is not biased by information on QSO color, an essential ingredient for the RF training. The 'star' sample is obtained by considering 332,650 unresolved sources in Stripe 82 that are not known QSOs and do not exhibit any QSO-like variations in their SDSS light curve. We randomly select the stars from this much larger sample such that the r-band number counts of the stars matches the QSOs. We train the RF selection with 11 input parameters: the 10 possible colors using the five optical and NIR bands $grzW1W2$, and the r-band magnitude. In contrast to the RF method applied during the DESI commissioning (Yèche et al. 2020), the final selection uses a single RF covering the full QSO redshift range, which we retrained with the latest processing of imaging catalogs, DR9.

In order to achieve the required QSO target budget, ~ 310 targets per deg^2[2], and to ensure a uniform target density over the full DESI footprint, we apply slightly different thresholds on the RF probability in the three regions (North), South (DES) and South (non-DES), see the exact definition on Fig. 3.14. We also vary the RF probability threshold with r, following $p_{\rm th}(r) = \alpha - \beta \times \tanh(r - 20.5)$. For the three regions (North), South (DES) and South (non-DES), we choose (α, β) to equal $(0.88, 0.04)$, $(0.7, 0.05)$, and $(0.84, 0.04)$, respectively. The heat map of the resulting selection is illustrated in Fig. 4.4.

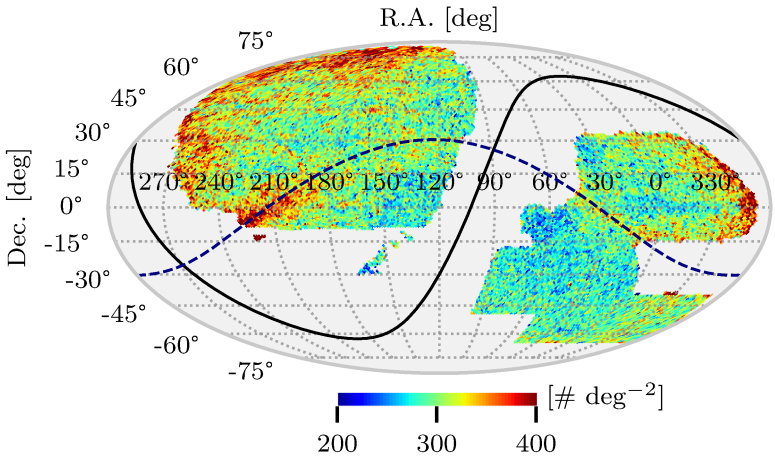


Figure 4.4: Density map of the DR9 QSO target selection. The solid black and dashed blue lines show respectively the Galactic plane and the plane of the Sagittarius Stream.

Fig. 4.5 shows the magnitude distribution of the QSO targets in the r, z, $W1$ and $W2$ bands and demonstrates that these distributions are similar for each of the imaging footprints highlighted in Fig. 3.14. The three magnitude limits imposed on the selection are clearly visible on the corresponding histograms. Note that the $r < 23.0$ limit also affects the z-band distribution, producing a sharp drop-off at the faint end for objects in the top-right panel. The two Gaussian distributions for $W1$ and $W2$ (bottom panels) demonstrate that the selection of a QSO target is not limited by the depth of the optical imaging (DECaLS or BASS/MzLS) but is sensitive to the determination of the fluxes in the WISE imaging.

[2]Since DES is less contaminated by stars than the other regions, the target density is on average smaller than 310 targets per deg^2.

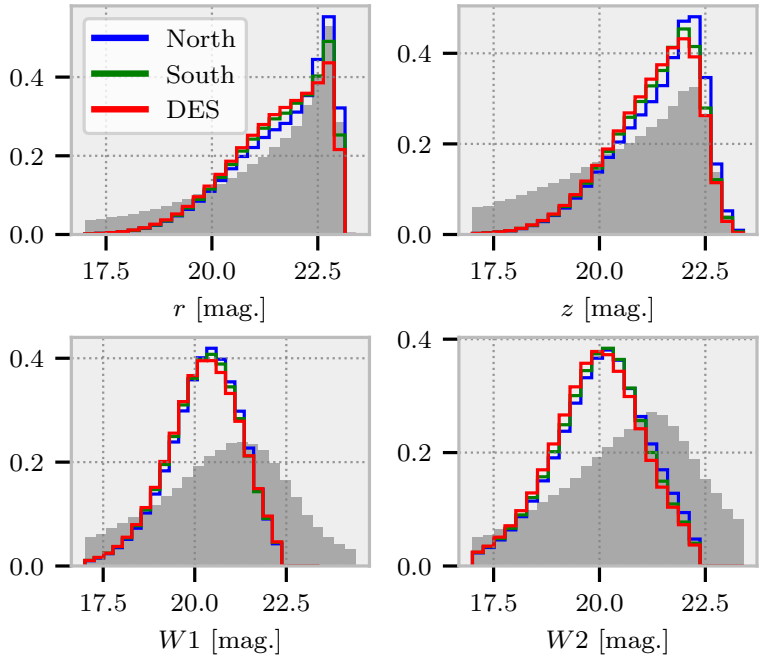


Figure 4.5: Magnitude distribution of the QSO targets in the r, z, $W1$ and $W2$ bands. These distributions are shown for the three independent imaging footprints in Fig. 3.14. Blue is for the North, green for the South and red for DES. Each grey histogram depicts the magnitude distribution for the parent sample of sources (PSF sources with $r < 23$) from which QSO targets are selected.

To help contextualize which bands guide the DESI QSO target selection, it is worth noting that the colors that carry the largest weight in the selection are first $z - W2$ and $z - W1$ and then $g - r$, $W1 - W2$ and $g - z$. This can be quantified using the importance features, see Section 4.3.4.2 for a complete description, as shown in Fig. 4.6. Although the feature r has the lowest Gini importance, the target selection is worse without including it in the training. As explained in Section 4.3.4.2, one can prefer to use the permutation importance to avoid this kind of artifact.

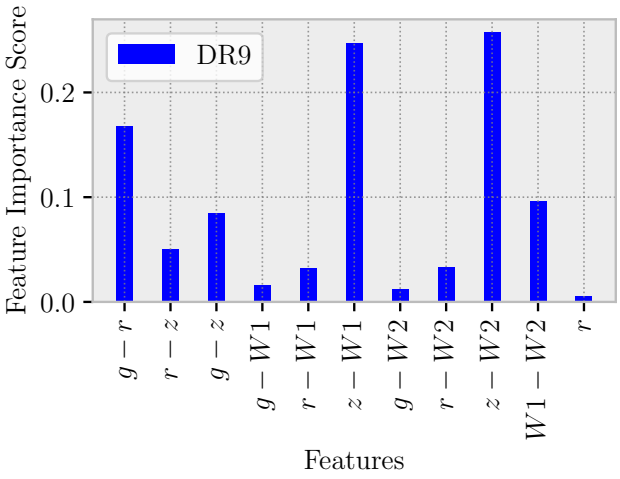


Figure 4.6: Feature importance for the QSO target selection, calculated using the Random Forest method based on Gini importance. As expected, the target selection is driven by $z - W2$ and $z - W1$ *i.e.* , by a difference between an optical and an infrared band.

The code for the QSO target selection of both the 1% survey and the main survey is public on GitHub and it is available at 1% QSO selection and main QSO selection.

4.2.2 Extended selection of QSO targets during Survey Validation

The goals of the first SV phase were to optimize the selection algorithms, estimate the redshift distributions, and evaluate the projected cosmology constraints. In this section, we describe the extensions of the main selection and the alternative QSO selection methods that we tested during SV.

In practice, as explained in Sec. 4.2.2.1, we varied the definition of the stellar morphology and the magnitude limit. We released the cuts on the RF probability and on the definition of the color boxes. The goal was to select fainter QSOs or those with higher redshift missed by the main selection. In parallel, we tested new algorithms using for instance, the intrinsic variability of the QSOs. All these variations of selections are grouped into 5 classes which are described in Sec. 4.2.2.2.

4.2.2.1 Alternative Selections

► **Source Morphology for the Quasar Selection** The first SV study was related to the definition of stellar morphology ('PSF' in DR9). Fig. 4.7 shows the potential gain that we expect using point-like sources in the COSMOS/HST region. For instance, we can extend the definition of 'PSF' sources to also include objects photometrically classified as 'extended' but having small relative χ^2 difference between PSF and extended morphological models ($\Delta(\chi^2)/\chi^2 < 0.015$). Using the DR9 Legacy Surveys Imaging catalogs, the relative χ^2 is defined as (`dchisq['REX']` − `dchisq['PSF']`)/`dchisq['PSF']`.

We will discuss in Sec. 4.2.3 the impact of these extensions and the optimization performed to achieve the final main selection of Sec. 4.2.1. For instance, in this specific case, we will study the redshift distribution of the QSOs recovered with the looser morphology restriction. It will allow us to assess the trade-off between a higher QSO completeness and an increase in the quasar target budget.

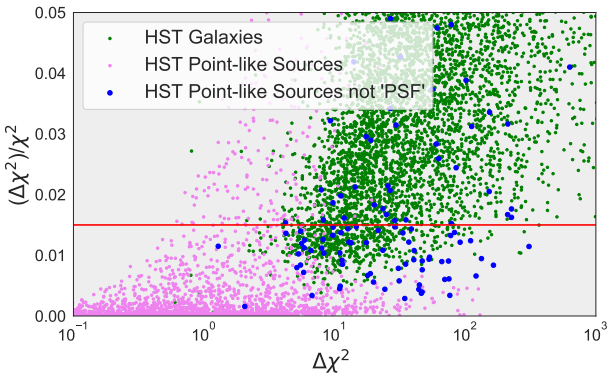


Figure 4.7: Relative χ^2 difference between extended and 'PSF' models as a function of the χ^2 difference, for COSMOS/HST objects. The violet dots correspond to objects confirmed as point-like sources in HST imaging. The green dots correspond to objects identified as extended galaxies in the HST imaging. The blue dots are HST point-like sources that are classified as extended objects in the DECaLS DR9 photometric catalogs.

► **Extension of the Color Selections** In parallel, we investigated two approaches for the QSO selection, one based on color cuts and the other on a machine-learning algorithm. At the time of the SV, both methods had reached a high level of maturity, and one of the SV goals was to select between the two approaches (Yèche et al. 2020). In addition to the pure performance in terms of number of true quasars selected per deg^2 for a given target budget, the relative sensitivity to systematic effects also has to be assessed. Finally, if the sample of spectroscopically-confirmed quasars selected by one of the two approaches is included in the sample selected by the other approach, we will retain the selection yielding the largest set of validated quasars.

The QSO luminosity function indicates that the size of the QSO samples at both $z < 2.1$ and $z > 2.1$ can be increased by extending the magnitude limit above $r = 22.7$. The benefits are particularly apparent for the higher redshift Ly-α forest QSOs. Therefore, we relaxed the magnitude limit to $r = 23.2$ for the extended SV RF selection, as shown by the distribution of r-band magnitude in the 'Ext. Random Forest Selection' on Fig. 4.8. We also developed an additional selection for Ly-α forest QSOs as faint as $22.7 < r < 23.5$ (see the distribution of r-band magnitudes of the 'High-z and Faint selection' in Fig. 4.8). In addition, a goal of SV was to determine how efficiently we can identify and classify high-redshift quasars with these extended selections for the nominal effective exposure time, $T_{\rm eff} \sim 1000$s (see definition in Schlafly et al. (2023)).

► **Selection of High-z QSOs** QSOs at $z > 5$ provide direct probes of the evolution of the intergalactic medium and supermassive black holes at early cosmic times. Current high-redshift QSO surveys either mainly focus on the bright end or are limited to a small deep field. We conducted a selection for $z \gtrsim 5$ faint QSOs using photometry from DECaLS grz and unWISE $W1$, $W2$. The selection method is based on the color selections that have been used in previous successful $z \sim 5-6$ QSO surveys (Wang et al. 2016, Yang et al. 2017). The main techniques are g/r-band dropout and the $r - z/z - W1$ color-color diagram. The unWISE $W1 - W2$ color is used to further reject M dwarfs. We have a survey depth of z band magnitude 21.4. We divide the selection into two sets based on two redshift ranges, $z_{\rm red} \geq 4.8$ and $4.3 < z_{\rm red} < 4.8$, and apply different color cuts according to QSO color-$z_{\rm red}$ tracks in $r - z/z - W1$ and $W1 - W2$ color space.

► **WISE Variability Selection** Finally, in Palanque-Delabrouille et al. (2011), it was demonstrated that the SDSS light curves on the stripe 82 provide a very efficient method to select the QSOs by their intrinsic variability. The DR9 (WISE) catalog offers, for each object, light curves with 15 epochs over a time period of about 10 years in the $W1$ and $W2$ bands. We adapted the method developed in Palanque-Delabrouille et al. (2011) to the (WISE) light curves. We selected objects with 'PSF' morphology and $18.0 < r < 23.0$, passing a low RF probability cut, $p > 0.1$, and exhibiting a high variability in their light curves. This variability technique is a robust, efficient and well-understood method, less sensitive to the spatial non-uniformity of the optical imaging. The goal was to study whether such a method can select quasars not already spotted by the usual methods based on optical and NIR colors.

4.2.2.2 Definition of the QSO Target Maskbits

We defined five classes of quasar selection for SV, grouping the extensions described above. Each selection can be identified by a combination of bits of `SV1_DESI_TARGET` defined in Myers et al. (2023). The code for the QSO target selection of SV is public on GitHub and a link to the code is provided for each class.

1. **Extended color Cut Selection**, `QSO_COLOR_4PASS` or `QSO_COLOR_8PASS` (~ 300 deg^{-2}): Compared to the Color Cut selection of Yèche et al. (2020), we relaxed all the definitions of the color boundaries; loosened the veto on the color box defined for stars, and applied a looser selection when requiring point-source morphology. Link to the code.

2. **Extended Random Forest Selection**, `QSO_RF_4PASS` or `QSO_RF_8PASS` (~ 570 deg^{-2}): Compared to the RF selection of (Yèche et al. 2020), the r-band magnitude limit is extended to $r = 23.2$, the RF probability is reduced, and a looser selection is applied to require point-source objects. Link to the code.

3. **High-z and Faint QSO Selection**, `QSO_HZ_F` (~ 115 deg^{-2}): The selection is extended to fainter objects $22.7 < r < 23.5$. We have also applied a looser cut on the RF probability than for the nominal selection but with an additional color cut to enhance the fraction of high-z QSOs. Link to the code.

4. **z $\sim$ 5 QSO Selection**, `QSO_Z5` (~ 20 deg^{-2}): We use g-band and r-band dropout techniques to select very high-z QSO candidates ($4.5 < z_{\rm red} < 5.5$). Link to the code.

5. **WISE Variability Selection**, `WISE_VAR_QSO` in secondary targets `SV1_SCND_TARGET` (~ 140 deg^{-2}): We use the intrinsic variability of the QSOs, based on the *WISE* light curves spanning over 10 years.

The r-band magnitude distribution for each class is shown in Fig. 4.8. Many objects are common to the different classes and the total density is not the simple sum of all the individual densities. Finally, the overall density is of the order of 700 targets per deg^2, to be compared to 260 targets per deg^2 for the original selection.

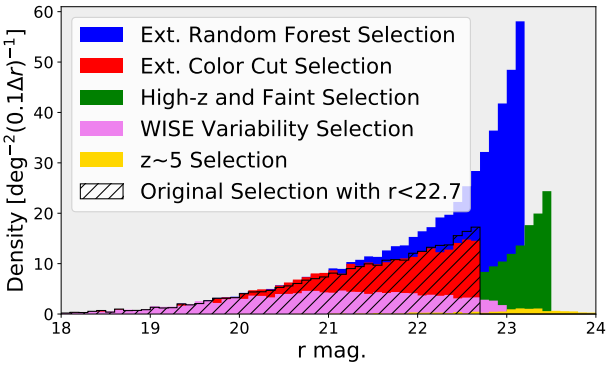


Figure 4.8: Target densities as a function of the r magnitude for the five classes of extended SV selections and the original selection with $r < 22.7$. All the selections are based on DECaLS DR9 imaging catalogs.

4.2.3 Optimization of quasar selection with Survey Validation

In this section, we describe our process to build a catalog of QSOs using the DESI spectroscopic information for each of the QSO targets we observed. We validate the catalog with a control sample of QSOs obtained after visual inspection of the spectra (Alexander et al. 2023). We then use this catalog to optimize the QSO selection (definition of point sources, magnitude limits, etc.) and we test the impact of the alternative selections (Color Cut selection, (WISE) variability, etc.) proposed in Sec. 4.2.2. For each alternative selection, we present its results and we estimate the potential gain in terms of QSO density by reference to the total target budget. Finally, we explain our choice of QSO selection for DESI which is described in Section 4.2.1.

Table 4.2: Description of the three datasets (first SV phase, 1% survey, main survey) used in this paper for QSO analysis. For the 1% survey, we only study the fields with an effective area greater than 0.4 deg^2).

	Number of Fields	Effective Area (deg^2)	Number of good spectra	Number of QSOs
First SV phase	45	90.5	78182	26094
1% survey	79	159.6	53307	33813
Main survey	305	1290.9	432383	264753

Table 4.3: Fractions of the spectrum types for the three ultra-deep fields that were visually inspected. Spectra that are of insufficient quality to assign a type are labeled 'inconclusive'. The first and second rows are respectively for the SV and Main selections.

	Fraction of QSOs	Fraction of stars	Fraction of galaxies	Fraction of inconclusive
SV sel.	33.5%	11.9%	39.8%	14.8%
Main sel.	71.0%	6.3%	16.1%	6.7%

4.2.3.1 Dataset and Control Sample Visually Inspected

The first phase of SV was used to optimize the QSO target selection. In this section we study 45 fields observed during this phase. They contain a mix of luminous red galaxy, emission line galaxy targets and quasar targets (see Table 4.2). Among them, 42 fields have a total effective exposure time of $T_{\mathrm{eff}} \sim 4000$s and 3 fields correspond to ultra-deep observations ($T_{\mathrm{eff}} =$ 7200s, 10820s, 8200s) (see a few examples of spectra in Fig. 4.9).

The latter observations have been visually inspected (VI) and those three fields provide a pure sample of QSOs that we use as a control sample when building the QSO catalog (see section 4.2.3.2). The breakdown of the visual inspection results is summarized in Table 4.3. As the main purpose of SV selection was to collect all the possible QSOs, the selection was extremely loose and we cannot draw any conclusion about the contaminants.

By contrast, the second row of Table 4.3 gives us a description of the contaminant of the QSO main selection. Roughly, one-quarter of the contaminants are stars and the other three-quarters are galaxies. The fraction of contaminant increases for fainter targets, especially for galaxies as illustrated in Fig. 4.10a. Comparison of Fig. 4.3 and Fig. 4.10b shows that the location in the color-color space, of the two contaminants, stars and galaxies, are in the middle of the QSO color space, demonstrating the difficulties to improve the QSO selection.

4.2.3.2 Quasar catalog

The process to produce the QSO catalog is illustrated by the flow chart of Fig. 4.11. The method is based on three algorithms: the DESI pipeline classifier Redrock (RR), a broad Mg II line finder (MgII) and a machine learning-based classifier QuasarNET (QN).

The RR algorithm (Bailey 2023) is a template-fitting classifier. It uses a set of templates for each class (star, galaxy or QSO) constructed from spectra observed in SDSS. After PCA decomposition, theses templates provide a linear basis. Linear combinations of the basis components are fitted to each spectrum for each redshift within a suitable range. From these fits, a best class and a best redshift is determined, corresponding to the template class-redshift combination that

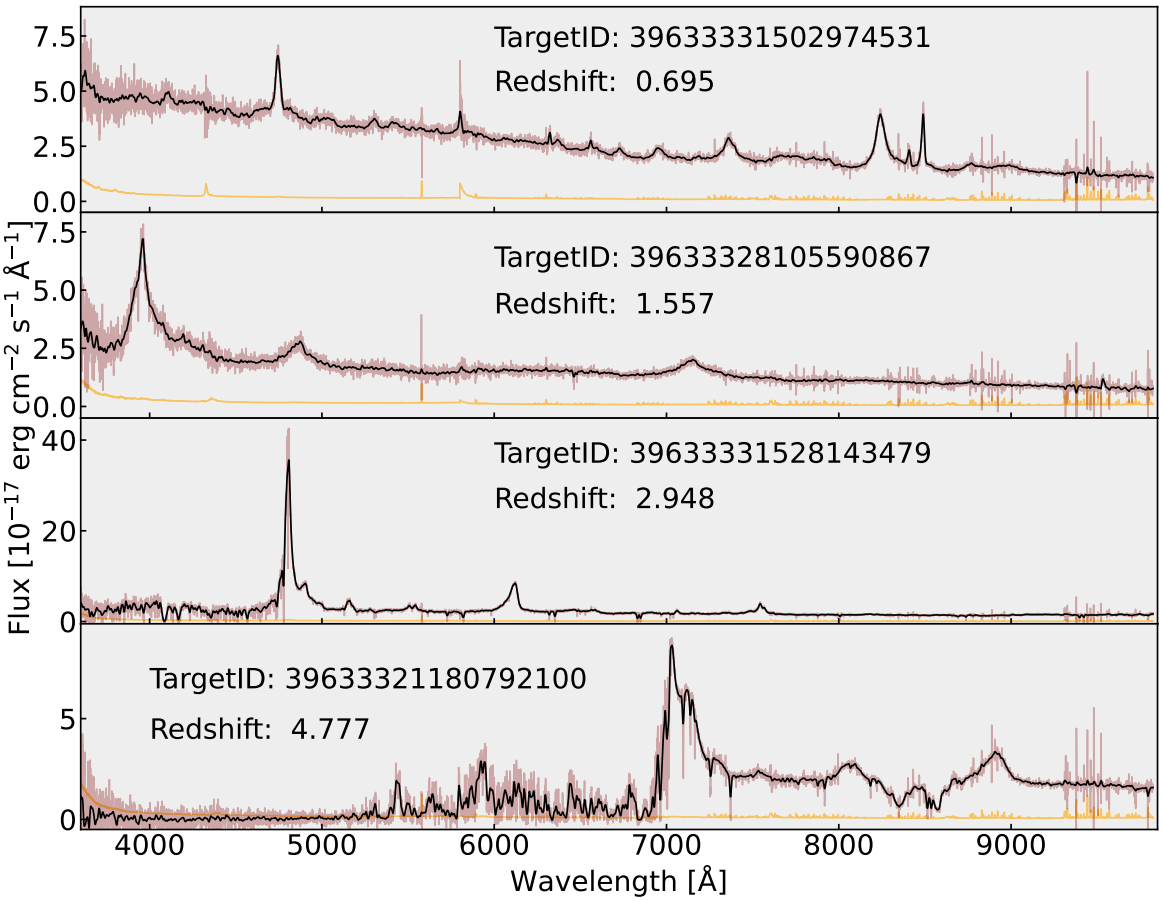
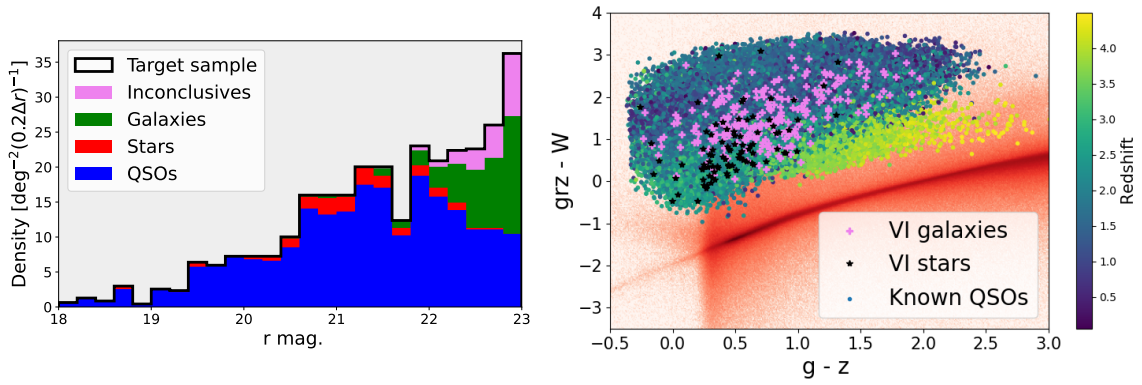


Figure 4.9: Four spectra of the ultra-deep field used for visual inspection. The field is centered at R.A. = 106.740° and Dec. = 56.100°. The effective exposure time is $T_{\text{eff}} = 10,820$s. The spectra cover the range of redshifts observed in DESI. The redshifts are 0.695, 1.557, 2.948 and 4.777. The last spectrum is a very rare case of a target selected by both the RF and the high-z selections. The maroon curves are the DESI spectra. The black curves are obtained after smoothing the spectra with a Gaussian filter. The orange curves represent the noise spectrum.

resulted in the lowest $\Delta\chi^2$. Therefore, as an output, RR provides both the class of the object (star, galaxy or QSO) and its best-fit redshift.

The MgII algorithm identifies spectra with a Mg II broad line. It is an afterburner, run after RR and using RR outputs as inputs. The goal is to change the initial classification of the object from Galaxy to QSO if the spectrum exhibits a Mg II broad line. The method consists in fitting a Gaussian in a 250 Å window centered at the position of Mg II line given by RR. We consider the Mg II line as a broad line if the improvement of χ^2 is better than 16, the width of the Gaussian greater than 10 Å and the significance of the amplitude of the Gaussian greater than 3. The algorithm possibly changes the source classification but never modifies the redshift given by RR.

The QN algorithm (Busca and Balland 2018, Farr et al. 2020) is a deep convolutional neural network (CNN) classifier, taking a smoothed spectrum as an input before carrying out four layers of convolutions. The output from these convolutions is then passed to a fifth, fully-connected layer, before feeding into a number of "line finder" units. Each of these units consists of a fully-connected layer, trained to identify a particular emission line. In our case, we use the following six lines: Ly α, C IV, C II, Mg II, H α and H β and an object is classified as a QSO if at least one



(a) (b)

Figure 4.10: (a) Composition of the main selection. (b) colors in the optical or near-infrared of objects photometrically classified as stars (red) or spectroscopically classified as QSOs (from blue to yellow dots, depending on their redshift). The black stars and the violet crosses correspond respectively to star and galaxy contaminants.

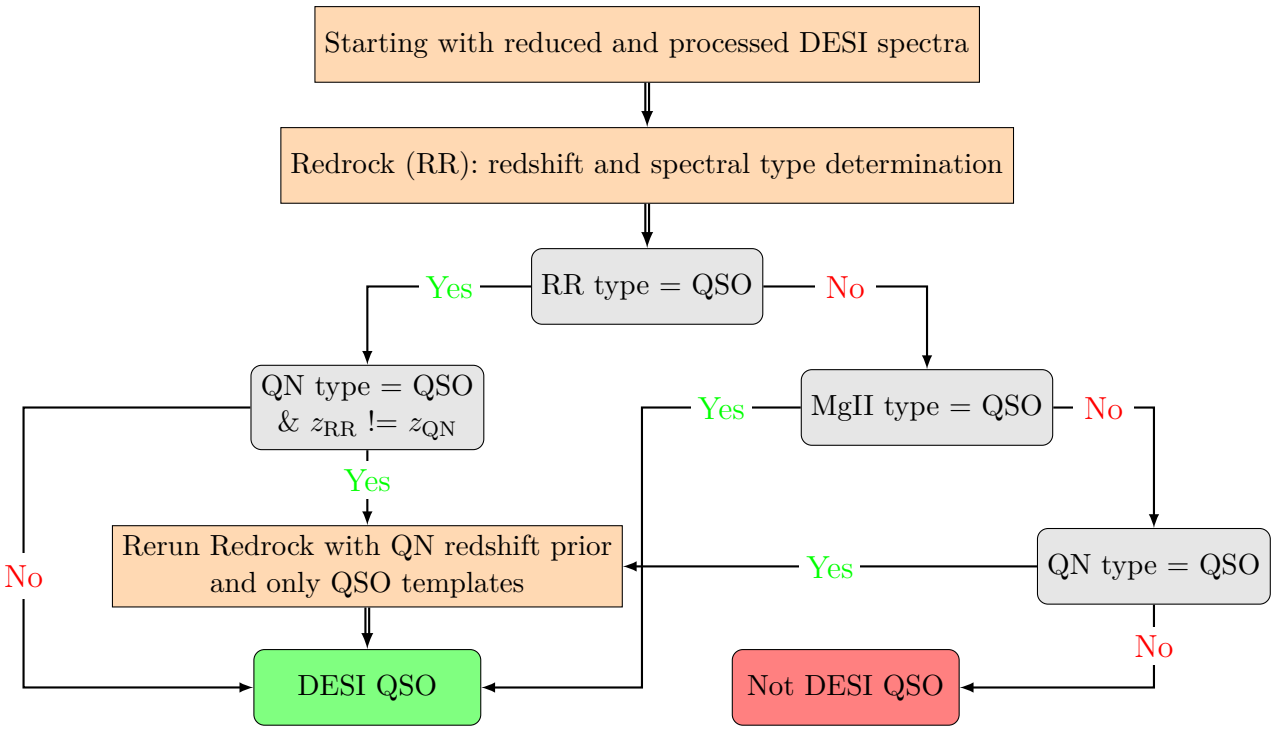


Figure 4.11: Flow chart to produce the quasar catalog.

of the six confidence probabilities is greater than 0.5.

Our strategy to build the final QSO catalog was established thanks to a control sample of QSOs obtained by visual inspection of their DESI spectra (Alexander et al. 2023). This truth sample contains $\sim$ 1330 QSOs passing the selection summarized in Sec. 4.2.2, see results in Table 4.3.

In order to monitor the quality of our catalog, we define two quantities, the *efficiency* and the *purity*. Those two parameters are performance metrics similar to parameters used in classification problems (Powers 2011), respectively, the *recall* and the *precision* . We define the *efficiency*, ε, as the fraction of the QSOs of the control sample that is selected in the catalog

and the *purity*, p, as the fraction of the catalog objects confirmed as QSOs. Numerically, the ε and p are defined as $\varepsilon = N_c^{QSO}/N_{cs}^{QSO}$ and $p = N_c^{QSO}/N_c$, where N_c^{QSO}, N_{cs}^{QSO} and N_c are respectively, the number of QSOs in the catalog, the number of QSOs in the control sample and the number of objects in the catalog.

Fig. 4.12 shows the performance achieved when the RR, MgII and QN algorithms are successively applied. From Fig. 4.12, we learn that by using the QSO class from RR alone, we obtain a catalog with a very high purity and an efficiency of the order of 80%. Adding the QSOs identified by MgII algorithm, low-z QSO are recovered. Finally, the QN algorithm allows us to recover faint QSOs missed by the RR or MgII algorithms. For the main selection described in Sec. 4.2.1, the total efficiency and purity are respectively $99.2 \pm 0.3\%$ and $98.3 \pm 0.4\%$. These results reflect the performance of the QSO catalog that will be used in the DESI science analyses and not the composition of the parent QSO target sample described in Table 4.3 and in Fig. 4.10a.

For the contaminants of the catalog, we have limited statistics, only 17 spectra. Therefore it is difficult to draw definitive conclusions. Of the 17 spectra, none corresponds to that of a star and 8 spectra do not have sufficient quality to assign a type. Of the 9 galaxy spectra, all the spectra but one have the correct redshift in the QSO catalog. These objects correspond to a transition phase during which the quasar is formed. Considering those galaxies with a good redshift as good tracers of the matter, the purity increases to $99.1 \pm 0.3\%$.

To summarize the flow chart of Fig. 4.11, we first classify the object as `QSO` if it is classified as `QSO` by RR. We check if the redshift is confirmed by QN, otherwise we refit the redshift with RR using a top hat prior of ± 0.05 around the redshift given by QN. Then, if the RR classification is `GALAXY` and the MgII classification is `QSO`, we classify the object as `QSO` and keep the redshift given by RR. Finally, if the object is classified as `QSO` by QN but neither by RR nor MgII, we classify it as `QSO` in the QSO catalog but we refit the redshift with RR using a ± 0.05 top hat prior around the redshift given by QN. In this way, all the redshift are obtained by a single algorithm, RR, providing a consistent measurement of the redshift.

In summary, we validated the automated QSO catalog with visually inspected objects. We achieve both excellent purity and excellent efficiency. In the rest of this paper, the QSO catalog is therefore built according to the strategy described above. The numbers of QSOs for all the datasets are given in Table 4.2.

4.2.3.3 Source Morphology for the Quasar Selection

The distribution of $\Delta(\chi^2)/\chi^2$ for point source objects in COSMOS/HST is illustrated in Fig. 4.7. It indicates that we can potentially improve the QSO selection by accepting objects with $\Delta(\chi^2)/\chi^2 < 0.015$. By relaxing the stellar morphology definition in such a way, the target density of the main selection (310 targets per deg^2) is increased by 70 targets per deg^2.

During SV, this option was tested. Fig. 4.13 shows the fraction of additional QSOs selected when relaxing the morphological criterion as a function of the redshift and the r magnitude. The improvement is mainly visible for faint QSOs with $z < 1$, which do not contribute to neither QSO clustering nor Ly-α forest studies. In addition, they only add 14 QSOs per deg^2 to a total of 200 QSOs per deg^2 for the main selection.

In conclusion, because the relaxed morphological selection only increases the number of QSOs at low redshifts and because the cost in terms of target budget is significant (+20%), we do not retain this extended definition of stellar morphology and we use the 'PSF' morphology definition of DR9 catalogs to select point-like sources.

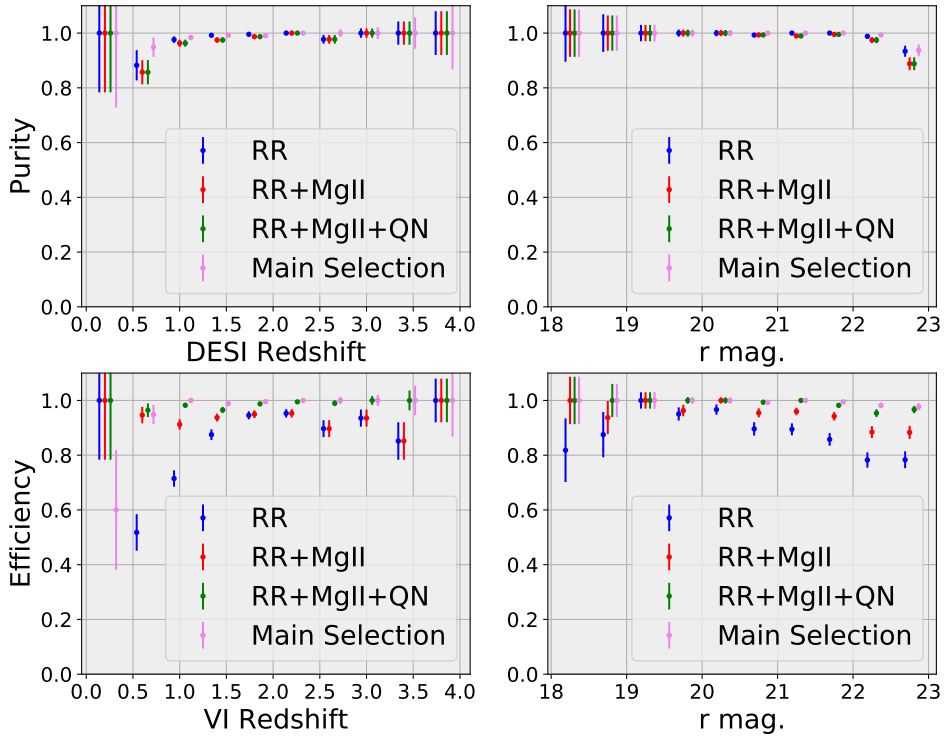


Figure 4.12: Efficiency and purity as a function of redshift and r magnitude, using the VI catalog as control sample. The efficiency is the fraction of the control sample that is selected in the catalog. The purity is the fraction of the catalog objects that are confirmed QSOs. Starting with QSO targets selected as described in Sec. 4.2.2, the three algorithms, RR, MgII and QN, are successively applied. The violet curve corresponds to the main selection described in Sec. 4.2.1 using the three algorithms (RR+MgII+QN).

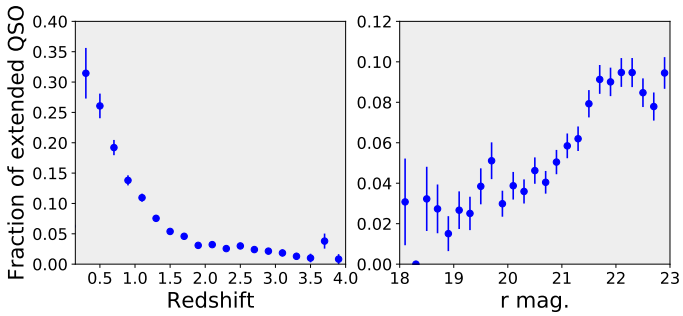


Figure 4.13: Fraction of additional QSOs selected by relaxing the morphological criterion as a function of the redshift and the r magnitude.

4.2.3.4 Results for the Alternative Selections

The SV phase also allowed us to study several alternative selections described in Sec. 4.2.2. Fig. 4.14 summarizes all the results and compares these alternative methods to the main selection based on a RF approach (see Sec. 4.2.1).

For a fixed target density, ~ 310 target per deg^2, the RF selection (main selection) retains 15% more QSOs than the Color Cut selection on average over all redshifts, and 21% more for the Ly-α forest QSOs. Taking the union of the RF and the color cut selections would increase the target budget by 20%. In addition, only 3% of the QSOs selected by the Color Cut method

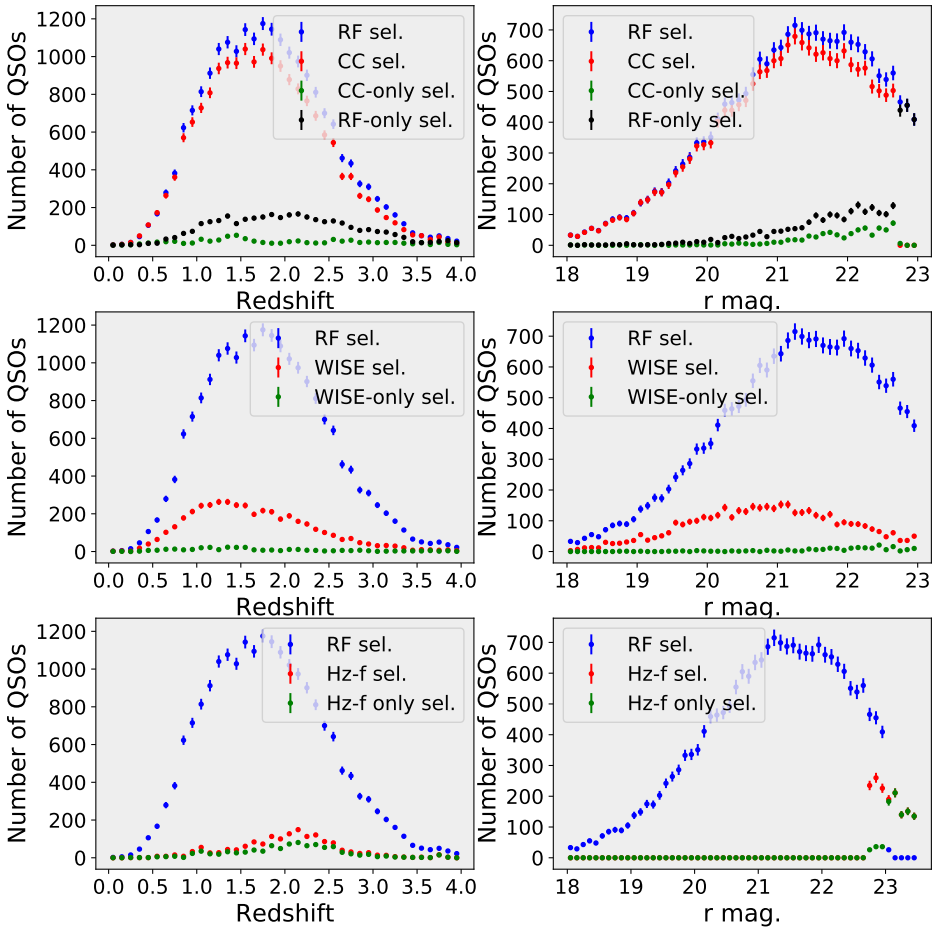


Figure 4.14: Study of alternative selections. Number of QSOs as a function of the redshift and the r magnitude. Each row of two plots tests successively a Color Cut selection (CC), a selection based on variability detected in WISE light curves (WISE) and a high-z faint quasar selection (Hz-f).

do not pass the RF selection, and the first row of Fig. 4.14 shows that they mostly have a low redshift. As the vast majority of the QSOs with $z > 1.0$ selected by the Color Cut method are included in the RF sample, we do not use the Color Cut selection.

A selection based on the detection of the QSO intrinsic variability with the WISE light curves represents a very interesting alternative because it shows a better spatial uniformity. However, the conclusions are similar to those of the Color Cut selection. The union with the main selection would increase by 15% the target budget with a QSO gain of 3%, mainly at low redshift (see second row of Fig. 4.14). Therefore, this selection was not retained.

We have also extended the RF selection to very faint objects $22.7 < r < 23.5$ with an additional color allowing us to select high-z quasars. This selection was extremely expensive in terms of target budgets (+30% for $r > 23.0$) and the gain in terms of QSOs was extremely small, as we can see on the third row of Fig. 4.14, especially for $r > 23.0$. In the main selection, we extended the magnitude limit cut from the original $r = 22.7$ upper bound to $r = 23.0$. In contrast, it was not worth selecting targets above $r = 23.0$.

Finally, this $z \gtrsim 5$ QSO selection has identified ~ 60 QSOs at $3.9 \leq z \leq 5.7$ during SV observations. Since at $z \sim 5$ the Lyα emission line is in the i band, color selection that does not include i band photometry will help to construct a sample without dependence on Lyα line luminosity. This selection does identify weak-line and strong broad-absorption-line QSOs

missed by the previous $z \sim 5$ selection based on $r - i/i - z$ colors (McGreer et al. 2013, Wang et al. 2016). However, this selection has high contamination rate due to the lack of i band data. The success rate is about 2-3% and most of the contaminants are M dwarfs. About half of the $z \sim 3.9 - 5$ QSOs can also be selected by the QSO RF selection. Therefore, this selection is not retained. An updated selection adding i band photometry from Pan-STARR1 (Chambers et al. 2016) has been developed as a secondary program in the 1% and year 1 main surveys, focusing on QSOs in a higher redshift range, $z \sim 5 - 6.5$.

In conclusion, all these studies validate the decisions made for the main selection described in Sec. 4.2.1: we select $16.5 < r < 23.0$ objects with a stellar morphology ('PSF' in DR9) and with a RF probability greater than the probability threshold, $p_{\rm th}(r)$. To ensure uniformity of the target density over the whole footprint, $p_{\rm th}(r)$ is optimized independently in each of the three imaging regions.

4.2.4 Validation of the main quasar selection in DESI

In this section, we study the performance of the main selection that was deployed both for the 1% and the main surveys. The resulting catalog of QSOs is obtained with the approach presented in Sec. 4.2.3.2.

4.2.4.1 Methodology

The instrumental conditions varied a lot during both the SV and the beginning of the main survey. For instance, at the beginning of SV, the fiber reach was limited because of technical developments on the positioners of the focal plane. As a result, only a small fraction of the QSO targets could be observed. This limitation was gradually removed, making data analysis and the comparison between fields more complex. Similarly, some observations were performed with a subset only of the ten spectrographs. To account for the large variability of the instrumental conditions during observations, we use the number of quasars per deg^2 obtained for each field.

First, for a given field (tile), we compute the *effective surface* defined as the ratio of the number of QSO targets with a spectrum over the number of QSO targets in the field, multiplied by the surface of the focal plane (8.2 deg^2). Note that the numerator does not include targets which are not assigned to a fibre or for spectra which do not pass the spectroscopic quality flag `COADD_FIBERSTATUS`. The effective surface varies from 1.6 deg^2 for the first tiles of SV, to 4.6 deg^2 for the tiles of main survey. For the 1% and the main surveys, the total effective surface are 160 deg^2 and 1290 deg^2 respectively (see Table 4.2).

We then divide the number of QSOs (defined as in Sec. 4.2.3.2) for a given field by its effective surface. Therefore, the number of quasars per deg^2 is a quantity insensitive to the instrumental conditions.

In addition, as both in the 1% survey and the main survey, the QSOs can be re-observed several times, we only use the first observation, meaning that we require respectively for the 1% and main surveys, `PRIORITY == 103400` and `PRIORITY == 3400`.

4.2.4.2 Performance of the Main Selection

First, we estimate the efficiency and the purity of the automated QSO catalog, as defined in Section 4.2.3.2 for the main selection with nominal exposure time conditions. To achieve this, we coadded the different exposures of the three tiles visually inspected into coadds of $\sim$ 1000s and we apply a posteriori the main selection. Using as the truth, the classification obtained by

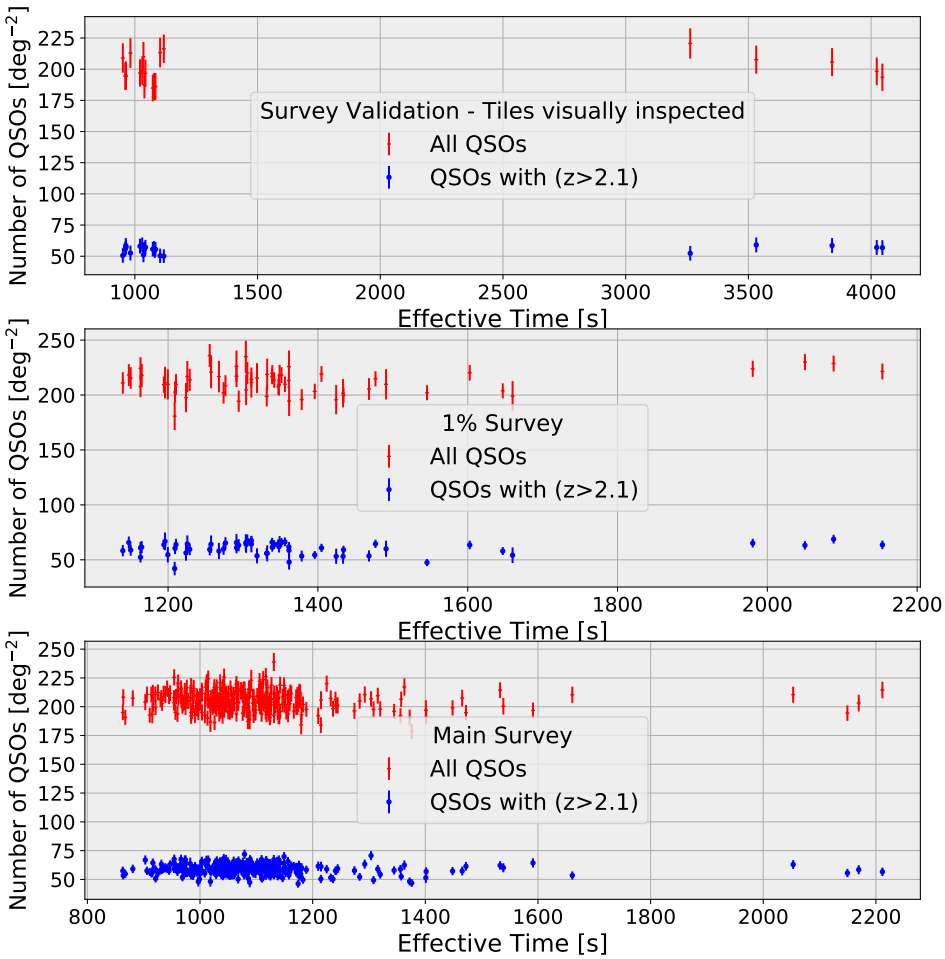


Figure 4.15: Number of quasars per deg^2 as a function of the effective time for SV1 ($\sim$ 1000s or $\sim$ 4000s), 1% survey and the main survey. Each point corresponds to a tile.

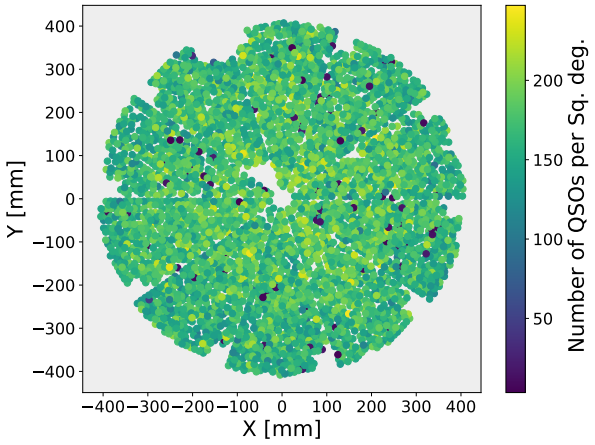


Figure 4.16: Number of quasars per deg^2 as a function of the target location on the focal plane. Each petal of the focal plane exhibits a hole at its periphery, corresponding to Guide, Focus, and Alignment (GFA) sensors. One of the 10 petals presents additional holes, due to connection issues with the positioners that were repaired during 2021 summer.

visually inspecting spectra containing all the coadds, we measure a $99.4 \pm 0.1\%$ purity and a $93.5 \pm 0.1\%$ efficiency, for effective exposure time, corresponding to $T_{\rm eff} \sim 1000$s. Considering the galaxy contaminant with a good redshift as good tracers of the matter, the purity increases to $99.7 \pm 0.1\%$. This very high purity of the automated catalog with nominal conditions allows us to use this catalog in the rest of the paper for the validation of the main selection.

Then, we study the performance of the main selection as a function of the *effective time*, $T_{\rm eff}$. In Fig. 4.15, three different datasets are studied: 1) the three tiles visually inspected, for which we coadded the different exposures in coadds of ~ 1000s or ~ 4000s effective time, 2) the 1% survey with an average ~ 1300s effective time, 3) the main survey with an average ~ 1000s effective time.

The result of the top plot of Fig. 4.15 had a crucial role in our choice of the final selection. It clearly shows that the number of quasars has very little dependence on the effective observation time. Whether for $T_{\rm eff} \sim 1000$s or $T_{\rm eff} \sim 4000$s, the number of QSOs is ~ 200 QSOs and ~ 60 QSOs per deg^2 for all QSOs and Ly-α forest QSOs, respectively. This stability of the results made it possible to extrapolate the results obtained for the SV ($T_{\rm eff} \sim 4000$s) to the main survey ($T_{\rm eff} \sim 1000$s).

The other two plots of Fig. 4.15 again show that the number of QSOs is very stable as a function of $T_{\rm eff}$, even when $T_{\rm eff}$ is below the nominal time, defined for the main survey ($T_{\rm eff} = 1000$s). By construction, during the main survey, the effective time will suffer from a certain dispersion, $\Delta T_{\rm eff} \sim \pm 150$s, but the stability of the number of quasars proves that QSO clustering analyses will not have to correct for a possible first-order effect related to exposure time. Similarly, the excellent uniformity of the number of QSOs as a function of target location over the focal plane, as illustrated by Fig. 4.16, should facilitate clustering analyses.

In conclusion, the performance of the QSO main selection is extremely stable in $T_{\rm eff}$ and uniform as a function of the target location on the focal plane.

4.2.4.3 Comparison with SDSS catalog

As the QSO targets have the highest priority in the DESI fiber assignment, the first two months of main survey already correspond to an effective surface of 1291 deg^2 for the QSO targets.

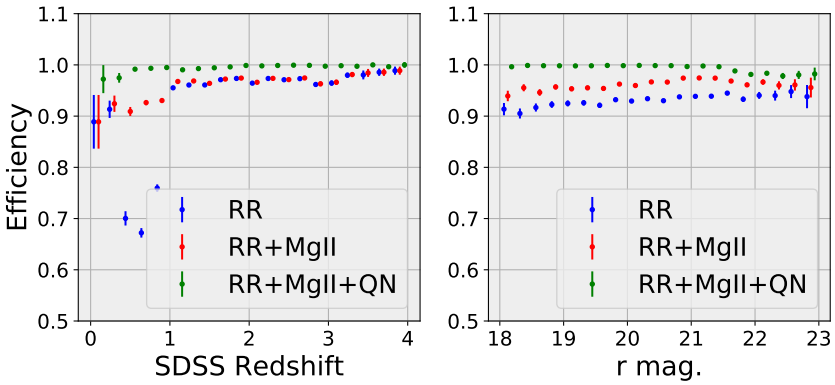


Figure 4.17: Efficiency as a function of redshift and r mag., using the DR16Q SDSS catalog as a control sample. The efficiency is the fraction of the DR16Q SDSS catalog that is selected in the DESI QSO catalog. The three algorithms, RR, MgII and QN are successively applied.

A large fraction of the DESI footprint is covered by the DR16Q SDSS QSO catalog (Lyke et al. 2020). In the DESI main survey, 49,148 QSO targets are also in DR16Q. We use these QSOs as a control sample with which we measure the efficiency (but not the purity because the DR16Q control sample is not complete) defined in Section 4.2.3.2. The results shown in Fig. 4.17 are quite similar to those obtained with the visually inspected control sample of QSOs (see Fig. 4.12). The RR algorithm has an efficiency at the order of 90%. The MgII algorithm allows us to recover low-z QSOs and finally QN algorithm allows us to achieve a 99% efficiency.

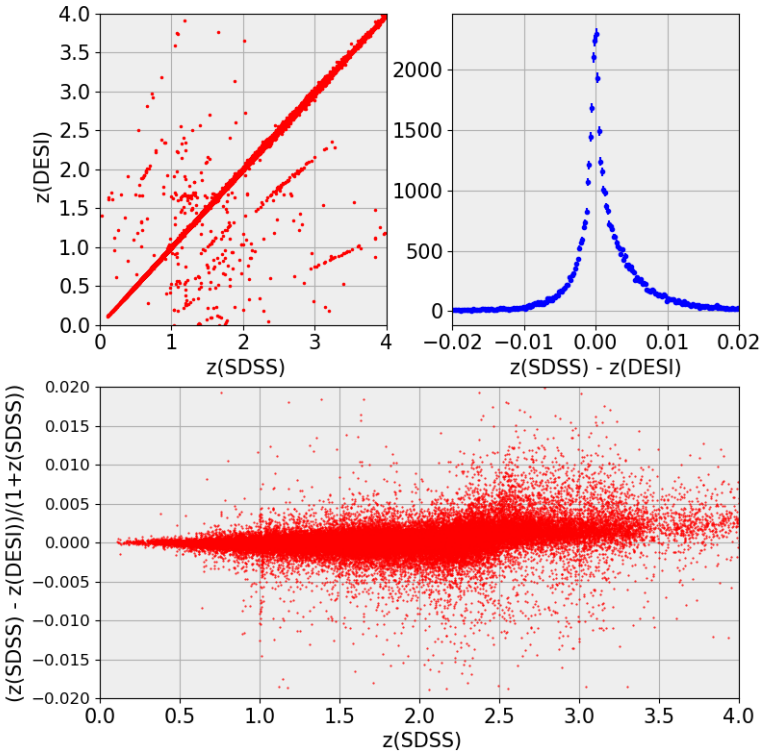


Figure 4.18: Comparison of DESI redshifts with SDSS redshifts. The objects are matched between the SDSS DR16Q catalog and the QSO catalog for the first two months of the DESI main survey.

In Fig. 4.18, we compare the redshift measurements of the DESI and DR16Q catalogs. The top left plot shows that the vast majority of QSOs have consistent redshifts. The off-diagonal QSOs (0.8% of the sample) most often correspond to an incorrect association of QSO emission

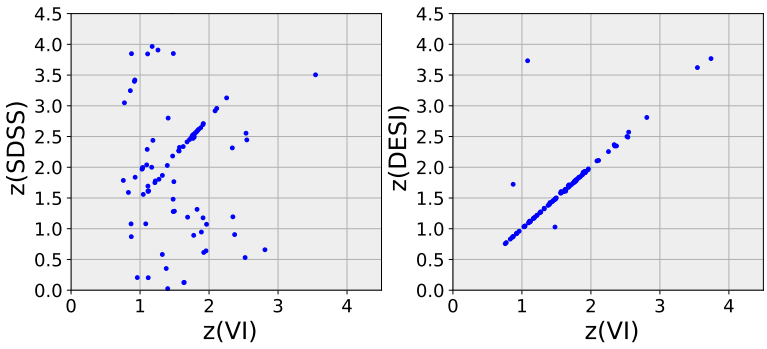


Figure 4.19: Comparison of the DESI redshifts and the SDSS redshifts with the VI redshifts when the SDSS and DESI redshifts are inconsistent (i.e. for QSOs that lie off the z(SDSS) = z(DESI) diagonal on the top left plot of Fig. 4.18).

lines, which results in lines that lie off the z(SDSS) = z(DESI) diagonal.

A visual inspection of one third of the off-diagonal QSOs which exhibit inconsistent redshifts between the two catalogs is summarized in Fig. 4.19. Only three objects out of 99 show a discrepancy between the visual redshift and the DESI redshift. In contrast, all the SDSS redshifts of the off-diagonal QSOs are inconsistent with the visual inspection redshifts.

The core of the redshift difference distribution, $\delta = z(\text{SDSS}) - z(\text{DESI})$, is shown in the top right plot of Fig. 4.18. It is clearly asymmetric, and the mean is significantly different from zero: $\mu(\delta) = (1.1 \pm 0.01) \cdot 10^{-3}$. The bottom plot of Fig. 4.18 seems to demonstrate that the asymmetry in δ appears only above redshift 2.5, when the Mg II line cannot be used to measure the redshift. A direct comparison of DR16Q redshifts with the systemic redshifts measured with spectra of the reverberation mapping project (Shen et al. 2016) or a more recent publication(Wu and Shen 2022) tend to confirm this discrepancy.

4.2.4.4 DESI redshift resolution

In the 1% survey, all the QSOs with $z > 1.6$ have been observed at four times the nominal exposure time in order to test the infrastructure that will allow DESI to observe the Ly-α QSOs four times longer that the rest of the QSOs. The 1% survey thus provides several repeats of the same QSO, allowing us to study the DESI redshift resolution. There are 103,350 pairs with $z > 1.6$ that can be used for comparison.

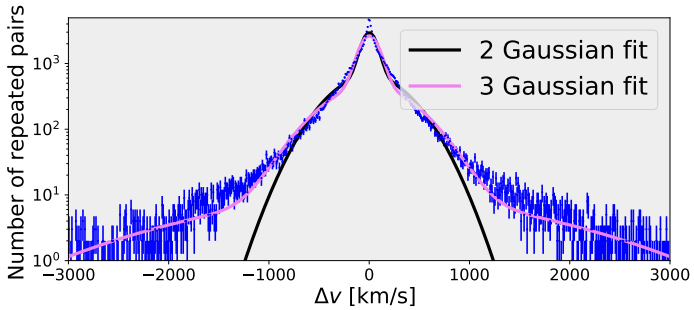


Figure 4.20: Comparison of the DESI redshifts for several repeats of the same QSO obtained with the 1% survey. The black and violet curves correspond respectively to two-Gaussian and three-Gaussian models.

For each pair (i, j) of redshifts, we compute the redshift difference, $\Delta v = (z_i - z_j)/(1 +$

Table 4.4: Number of QSOs per deg^2 for the 1% and main surveys, obtained with a 310 deg^{-2} main selection target density. The second and third columns are for all the QSOs and for the $z > 2.1$ QSOs, respectively.

	Number of QSOs (deg^{-2})	Number of Ly-α QSOs (deg^{-2})
DESI requirements	170	50
1% survey	211.9 ± 1.2	61.0 ± 0.6
Main survey	205.1 ± 0.4	59.1 ± 0.2

$(z_i + z_j)/2) \times c$ (see Fig. 4.20). The standard deviation of the Δv distribution is 372 km s^{-1}, indicating a redshift resolution of the order of 263 km s^{-1}. However, Fig. 4.20 shows a non-Gaussian distribution with very wide tails. A two-Gaussian model encounters difficulties in reproducing the tails (black curve). The Δv distribution is better modeled by three Gaussians with $\sigma_1 = 95$ km s^{-1}, $\sigma_2 = 400$ km s^{-1} and $\sigma_3 = 1500$ km s^{-1}, corresponding to 53%, 44% and 3%, respectively, of the total distribution.

4.2.4.5 Results

The results in terms of number of QSOs per deg^2 are summarized in Table 4.4. With a 310 deg^{-2} target density, the main quasar selection selects more than 200 deg^{-2} quasars, including 60 deg^{-2} quasars with $z > 2.1$. The QSO densities are exceeding the project requirements by 20 %. We expect a similar gain of 20% in the measurement of the cosmological parameters from clustering of either QSOs or Ly-α QSOs compared to the forecasts given in DESI Collaboration et al. (2016b).

We measure a slight difference between 1% survey and main survey, on the order of a few percents. This is partly due to the difference in the effective exposure time but mainly due to the regions of the sky observed. These first two months of the main survey are located near the Galactic Plane, a region where the imaging is of lower quality, which explains the small observed discrepancy.

The comparison of the distribution of the QSO number as a function of the redshift is remarkably identical for the North and South imaging (see Fig. 4.21). The only area of small discrepancy is at low z, a region where the target selection depends notably on the definition of the stellar morphology ('PSF'). The morphology which is driven by the z band in the North imaging is different than in the South imaging where the three optical bands contribute almost equally.

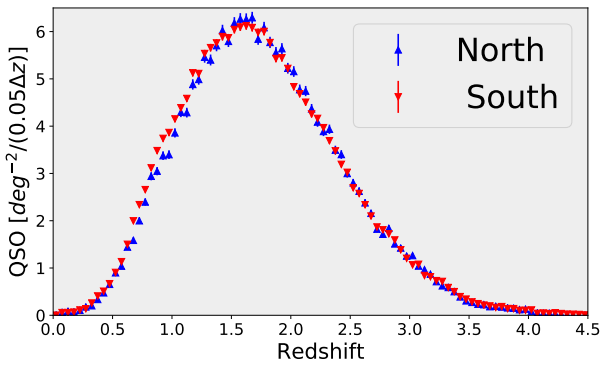


Figure 4.21: dN/dz for North and South regions.

Finally, we compare our results to the density of QSOs as a function of redshift that we can derive from the Quasar Luminosity Function (QLF) of Palanque-Delabrouille et al. (2016). In Fig. 4.22, the QLF is corrected for target selection completeness, $\varepsilon(z, r)$, which depends on redshift and r magnitude. This selection completeness is determined from the QSOs in the RF test sample that were not used in the RF training and that pass the selection of Sec. 4.2.1. For $r < 22.7$ (blue curve), we obtain an excellent agreement between the prediction from the QLF and the observed number of QSOs. The very small discrepancy observed for $r < 23.0$ (red curve) comes, on the one hand, from uncertainties in the QLF, in particular for faint QSOs, and, on the other hand, from the limited number of QSOs available in the RF test sample beyond 22.7 in r.

In conclusion, with a 310 deg^{-2} target density, the main selection based on a RF approach selects over 200 deg^{-2} quasars, including 60 deg^{-2} quasars with $z > 2.1$, exceeding the project requirements by 20%. These QSO densities are in excellent agreement with QLF predictions.

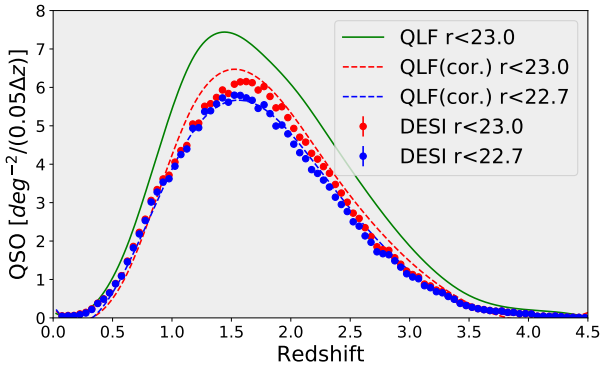


Figure 4.22: Comparison between QLF predictions and the measured quasar density for $r < 22.7$ and $r < 23.0$. The green curve is the QLF integrated up to $r = 23.0$. The red and blue curves are computed after correcting by the target selection completeness, $\varepsilon(z, r)$. The blue and red dots correspond to the QSO density obtained with the main survey for $r < 22.7$ and $r < 23.0$ QSO targets, respectively.

4.3 Mitigation of imaging systematics in the QSO target selection

4.3.1 Context

As described in Section 2.2.3, a promising approach to probe inflation is through the tiny imprint left on the matter power spectrum by inflation-induced primordial non-Gaussianity. This measurement Ross et al. (2013), Castorina et al. (2019), Mueller et al. (2022) is known to be limited by systematic effects on large scales, most of which are due to imaging systematics imprinted on the density of spectroscopic targets during the target selection.

Hence, the aim of this section is to mitigate imaging systematics for the DESI QSO target selection. We compute the angular clustering properties of the QSO targets in order to validate the selection method and to provide a control sample of QSO targets stripped of residual biases from imaging systematics or selection criteria. This is particularly important as the QSO target selection is known to be strongly contaminated by stars in addition to being impacted by imaging systematics. This work also serves as a crucial input to the selection of the QSO targets for DESI, which will soon be finalized for the five-year duration of the survey, in order to avoid strong imprints into the spectroscopic QSO sample due to imaging systematics that occur during the

target selection. Finally, it is also a proof of concept. Since the stellar contamination of the target selection will be removed after the spectroscopic survey, the same method could be used to mitigate systematics in the DESI spectroscopic QSO sample.

Different strategies have been developed to deal with imaging systematic effects and improve the reliability of clustering studies. In this work, we follow the approach that was used for SDSS studies (Myers et al. 2006, Ross et al. 2011, Ho et al. 2012, Ross et al. 2017; 2020, Raichoor et al. 2021) and for Dark Energy Survey (DES Collaboration et al. 2021) studies (Leistedt et al. 2016, Elvin-Poole et al. 2018). This method models the variation of target density as a linear function of imaging features (see, *e.g.* Myers et al. 2015, Prakash et al. 2016) in order to remove fluctuations caused by imaging systematics. A correction weight is then computed and applied to the data. Since this method smooths density fluctuations, one needs to check to what level the mitigation procedure affects the cosmological signal. Another less common approach is based on mode projection (Rybicki and Press 1992, Tegmark et al. 1998, Leistedt et al. 2013, Elsner et al. 2016, Kalus et al. 2019): Modes (in Fourier space) or pixels (in configuration space) are assigned increased variance where the systematics map exhibits large values, such that the covariance matrix has larger values in the presence of systematics. This is a robust method, which, however, only mitigates systematics using also a linear combination of the imaging maps. It cannot model the non-linear effects that are now observed, as illustrated in Ho et al. (2012). The correction-weights and mode-projection strategies can be combined in a common framework as explained in Weaverdyck and Huterer (2021).

To circumvent the assumption that only known features can completely explain all imaging systematics, one can also use a forward-modeling approach (Suchyta et al. 2016, Burleigh 2018, Kong et al. 2020). Such an approach accounts for source detection and target selection processes in an end-to-end fashion, by injecting fake galaxies into raw images, running source detection on the images, and applying target selection algorithms to the resulting sources.

Some studies of how imaging systematics affect DESI target selection have already been undertaken. For instance, Kitanidis et al. (2020) gave a first overview of imaging systematics for different DESI target classes selected from Data Release 7 (DR7) of the DESI Legacy Imaging Surveys (see e.g. Dey et al. 2019), and found that the DESI QSO target sample suffered from strong contaminating effects. In another study of DR7 of the Legacy Imaging Surveys, Rezaie et al. (2020) analysed how imaging systematic affect the eBOSS-like ELG selection, and used an artificial neural network to mitigate non-linear effects.

In this section, we will therefore analyse the final QSO target selection that is used by DESI in its nominal survey, and we will explore machine-learning approaches based on Random Forests (RFs) and Neural Networks (NNs), which we will compare to each other as well as to a traditional linear treatment, to mitigate systematic effects on the QSO selection. The mitigation process will be tested by measuring the angular properties of the QSO target selection, before and after applying the weights derived from these techniques.

4.3.2 Photometric properties of the QSO selection

Fig. 4.23 shows the density map of the QSO targets from the main selection described in Sec. 4.2.1. In particular, it exhibits several regions with higher density of QSO targets than average:

- Overdensity near the Galactic plane: the stellar density is higher near the Galactic plane (black line in Fig. 4.23), which increases the stellar contamination in the QSO target selection. This effect is not obvious in the region bounded by R.A., Dec. $\in [120°, 140°] \times [-10°, 15°]$, because the lower $W1/W2$ PSF depth counters the excess of targets caused

by the higher stellar density.

- Overdensity in the Sagittarius Stream (see also Section 4.3.3.1): the stellar population of the Sagittarius Stream, indicated by the blue dashed line in Fig. 4.23, is different from the Galactic stellar population. Most of the stars in the stream are bluer than Galactic stars and tend to have similar colors to the bulk of the QSO population. We empirically noted that Sgr. Stream stars are very faint in the two NIR bands, $W1$ and $W2$ compared to Galactic-plane stars, which justifies our NIR cut ($W1 < 22.3$ and $W2 < 22.3$) of Sec. 4.2.1. This overdensity is mainly visible in the NGC, but it can also be observed in SGC at $0° <$ R.A. $< 30°$.

- Overdensity in the North: the DESI QSO target density increases with declination. This overdensity could be due to the poorer PSF depth in the z band in this region. This is likely caused by imaging depth decreasing at higher declination due to increasing airmass that was not compensated for by additional exposure time in the MzLS observing strategy. Since the z band plays a crucial role in the QSO selection, the discriminating power between stars and DESI QSO targets is reduced at higher declinations.

- The DES footprint, which benefits from a one mag. deeper photometry in all optical bands is, as expected, the least contaminated region. Note that DES region is similarly contaminated in the region of the crossing of the Sagittarius Stream.

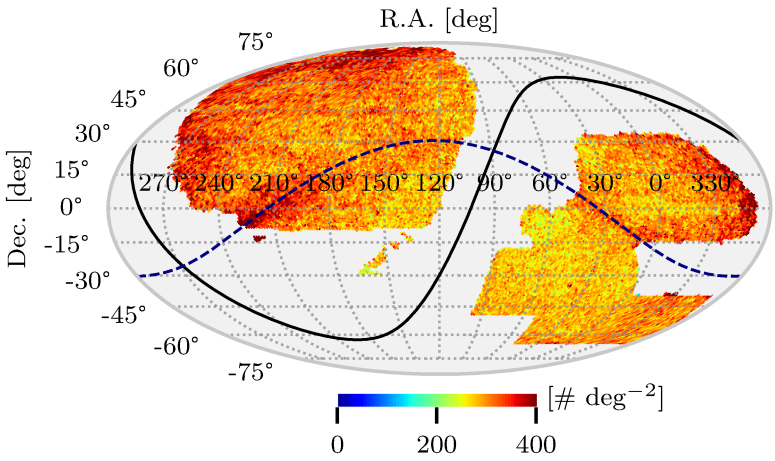


Figure 4.23: Density map of the DR9 QSO target selection. The solid black line indicates the Galactic plane and the blue dashed line indicates the plane of the Sagittarius Stream.

Fig. 4.24 shows the relative QSO target density as a function of each observational parameter, allowing us to identify the main sources of systematic effects in the QSO target selection. Observational parameters (called features in the next) are described in Section 4.3.3.1. We observe very different behaviors in the three regions:

- In the South (DES) region, because of the deeper photometry in optical bands, all the fluctuations of the relative density are at the order of only a few percent, typically, one order of magnitude below the level of fluctuations in the other two regions (North and South (non-DES)).

- In the North region, the morphology is driven by the z band since the MzLS telescope is the one with the best seeing. In addition, the segregation between stars and QSOs is based on the comparison of the optical z band with the two NIR ($W1$,$W2$) bands. Therefore,

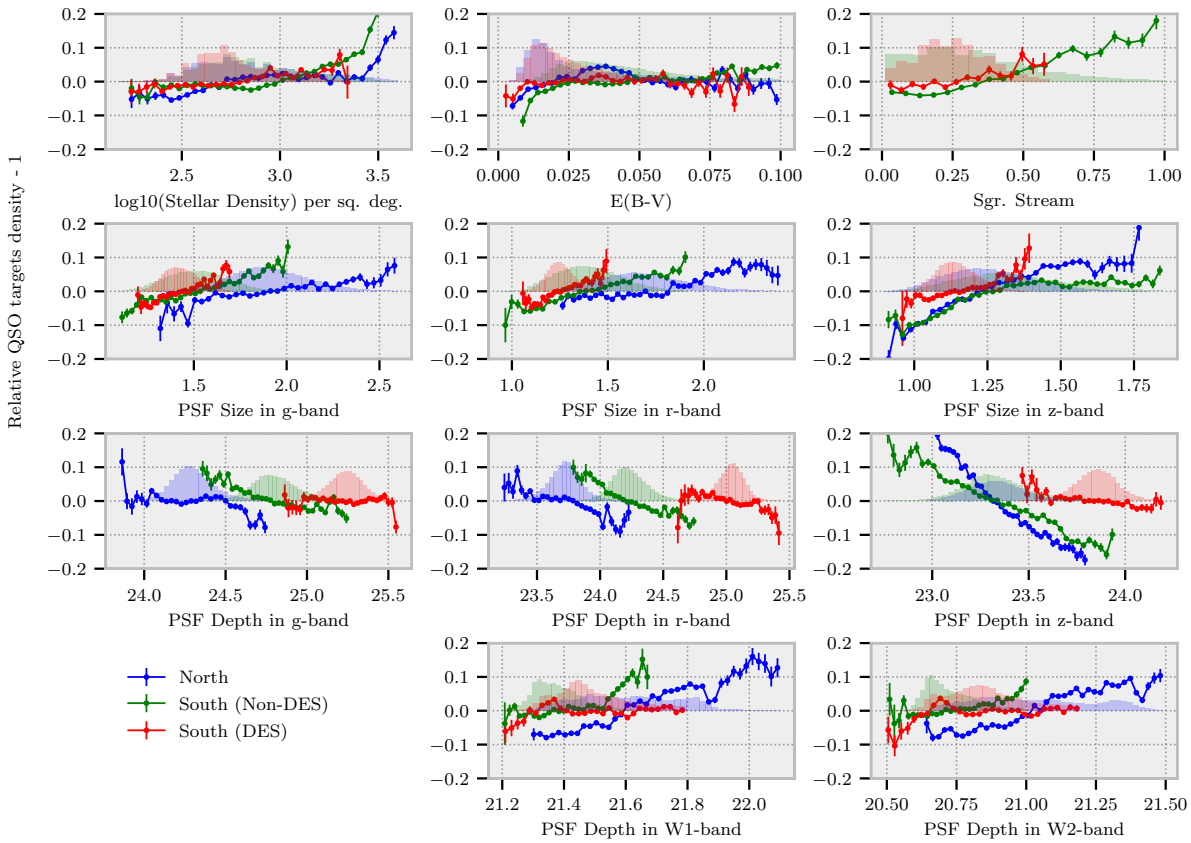


Figure 4.24: Relative QSO target density in the North, South (Non-DES) and South (DES) regions as a function of each observational parameter (see Sec. 4.3.3.1 for the definition of the parameters). The relative QSO target density is a mean value after rejecting outliers. The histograms represent the distributions of each observational parameter in the three regions. The color code is blue, green and red, respectively, for the North, South (Non-DES) and South (DES) regions.

we observe a strong dependence of the QSO density on the z band seeing as well as on the z, $W1$ and $W2$ depths.

- In the South (non-DES), we observe the same trends except that the QSO density is less sensitive to the z band seeing, because the morphology is determined from an almost balanced combination of the three optical bands. Finally, the $W1$ and $W2$ imaging is shallower in the South region compared to North one, because WISE produced many more images around the North Ecliptic Pole located in the North region. However, the behavior is essentially the same: the blue and green curves (see Fig. 4.24) are just shifted by ~ 0.5 magnitude.

4.3.3 Methodology

All the density and feature maps discussed in this section will be pixelized using `HEALPix`[3] (Gorski et al. 2005) with $N_{\rm side} = 256$. All `HEALPix` operations are done using the `healpy`[4] package (Zonca et al. 2019). The choice of $N_{\rm side}$ (*i.e.* the size of the pixelization) is justified in 4.3.5.2.

[3] http://healpix.sf.net
[4] https://healpy.readthedocs.io/en/latest/

This method is implemented in `regressis`[5], and it is currently used to compute the default weights to correct the imaging systematics in the catalog creation pipeline of DESI. This method was also used in Krolewski et al. (2023) to constrain $f_{\mathrm{NL}^{\mathrm{loc}}}$ by cross-correlating the QSO target density with the CMB lensing.

4.3.3.1 Features

The aim of the systematic mitigation is to correct for spurious density fluctuations in the target selection without suppressing the cosmological signal contained in the targets' clustering. We restrict our study to "features" directly linked to the observational properties such as imaging quality and physical properties that could altered the observations as the stellar density. We are careful to avoid incorporating parameters that relate to specific positions in the sky. For instance, we do not use the Modified Julian Date (MJD) as a feature (as in, *e.g.* , Rezaie et al. 2020), since the date of an observation directly translates into the position observed on that date.

▶ **Observational Features from DR9** We use ten observational features to describe the systematic effects in the DESI QSO target selection. Whether our features are sufficient to describe spurious density fluctuations will be verified after the mitigation procedure by checking the isotropy of the corrected target selection map (see Section 4.3.4.3). The impact of mitigating systematics on the observational features themselves will be presented in Section 4.3.4.1.

The feature maps are generated using the script `bin/make_imaging_weight_map` from the `desitarget` package[6]. The maps, which we describe below, are extracted from the random catalogues provided as part of DR9 of the Legacy Imaging Surveys[7], with the exception of the stellar density map:

- Stellar density (Stardens) [deg^{-2}]: Density of point sources from Gaia DR2 (Gaia Collaboration et al. 2018) in the magnitude range: $12 <$ PHOT_G_MEAN_MAG < 17.
- E(B-V) [mag]: Galactic extinction from Schlegel et al. (1998) as modified by Schlafly and Finkbeiner (2011).
- PSF Depth [1/nanomaggies^2] in r, g, z, $W1$, $W2$: PSF depth is the 5-sigma point-source magnitude depth[8]. The dependence of target selection on PSF depth is governed by two competing effects. On the one hand, the number of resolved objects increases with PSF depth. On the other hand, the determination of the flux is better for brighter objects, which means that contaminants are more easily rejected, resulting in a reduced number of targets. In this study, the z depth does not limit the target selection (cf. Section 4.2): it only affects the measurement of the z flux. Therefore, the target density decreases with increased z PSF Depth. In contrast, fluxes in $W1$ and $W2$ are obtained with forced-photometry. This allows fluxes to be measured for fainter objects that are detected at marginal significance in the WISE imaging (see Dey et al. 2019). Because such fluxes are noisy, the resulting colors scatter in regions of insufficient depth. Corresponding objects are rejected by the target selection not because they are not quasars but because their

[5] https://github.com/echaussidon/regressis
[6] https://github.com/desihub/desitarget
[7] https://www.legacysurvey.org/dr9/files
[8] For a 5σ point source detection limit in band x, $5/\sqrt{x}$ gives the PSF Depth as flux in nanomaggies and $-2.5\left(\log_{10}(5/\sqrt{x}) - 9\right)$ gives the corresponding magnitude (see https://www.legacysurvey.org/dr9/catalogs/).

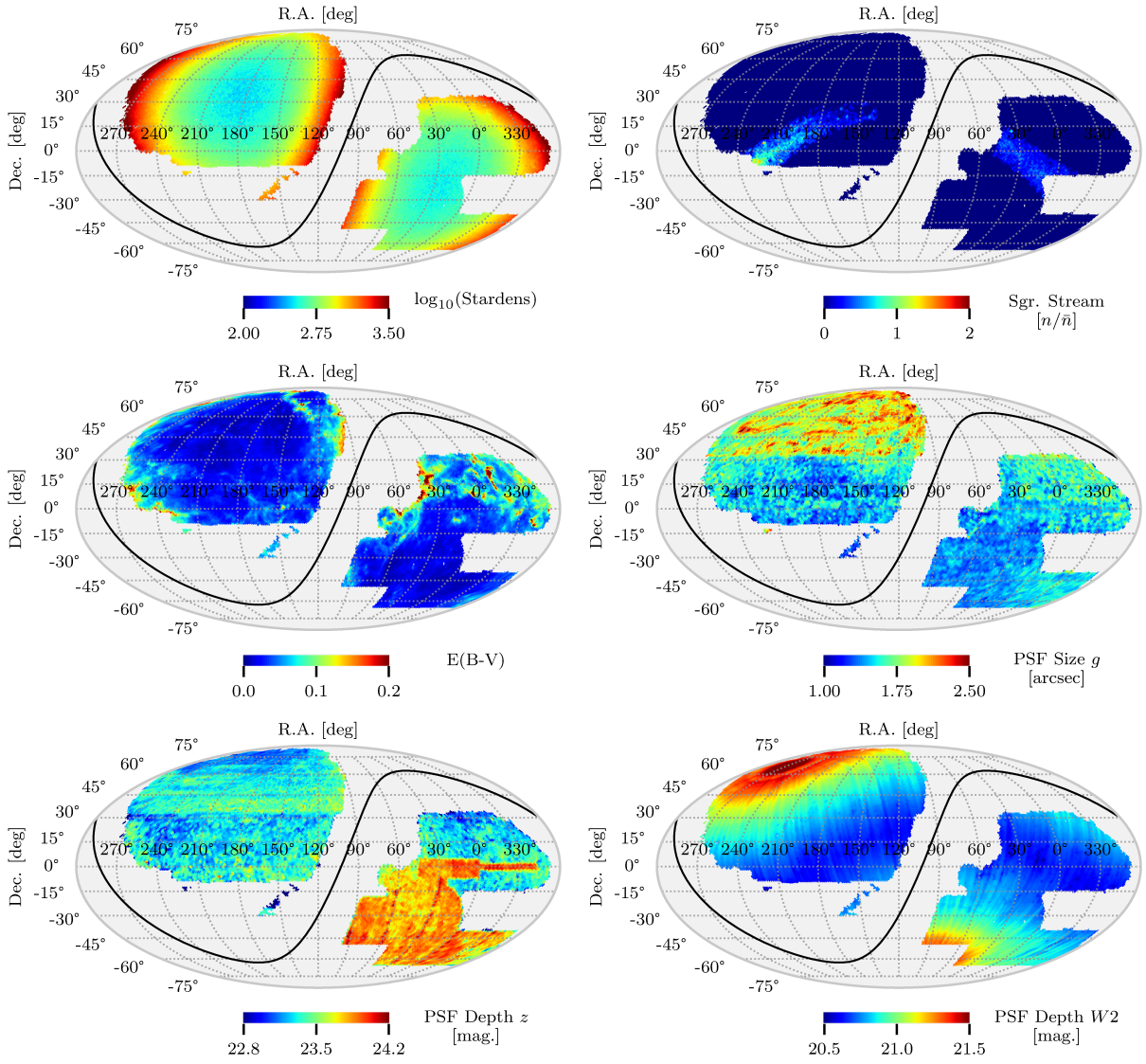


Figure 4.25: Maps of the most important features used in our systematics mitigation. The difference between the three imaging footprints highlighted in Fig. 3.14 is clearly visible in the *PSF Depth z* and *PSF Size g* feature maps.

colors are not in the region populated by QSOs. Conversely, when the PSF Depth in $W1$ or $W2$ is higher, the colors are more precise, and a greater number of targets are selected as genuine QSOs.

- PSF Size [arcsec] in r, g, z: Inverse-noise-weighted average of the full width at half maximum of the point spread function, also called the delivered image quality. A small value of PSF size corresponds to a good image resolution, which leads to more precise fluxes and improved target selection.

Fig. 4.25 shows 5 of the 10 observational feature maps from DR9. It is important to note that

these 10 features are correlated, and the level of correlation depends on the imaging footprint (North, South, DES). Stellar density and E(B-V) are positively correlated across the entire sky but E(B-V) contains additional information. This can be seen by comparing the top two left panels in Fig. 4.25 in the region near (R.A., Dec.) = (0°, 15°). The $W1$ and $W2$ PSF Depths are highly correlated in all of the three imaging footprints of interest. However, we include both as our analysis will show that the $W2$ PSF Depth is particularly heavily weighted by the Random Forest in the North, indicating that this feature has extra information which is not contained in the $W1$ PSF Depth. For the optical observational features, two cases are of particular note. In the North, the g and r bands were collected by the same camera and so their observational features are positively correlated, but these features are not correlated with the z-band, which was independently obtained by MzLS. In the South and in DES, the three bands were collected with the same camera and therefore are all positively correlated. Finally, the $W1/W2$ PSF Depths are more correlated with the other features in the DES region, as compared to the North and the South, since all the feature maps are more uniform in the DES footprint.

► **Sagittarius Stream Model** The target density map (Fig. 4.4) shows a significant excess in the Sagittarius Stream region. The Sagittarius galaxy is one of the closest galaxies orbiting around the Milky Way. The gravitational forces create two tidal arms called streams wrapping the Milky Way with the same orbit (Newberg et al. 2002, Majewski et al. 2003).

The Sagittarius contamination (see the blue dashed line in Fig. 4.4) occurs mainly in the South footprint but also touches the DES footprint. However, none of the feature maps discussed so far contain a pattern matching this contamination. Thus, an additional feature is required in our analysis. To model this feature, we use the Sagittarius Stream catalogue derived in Antoja et al. (2020). This catalogue is built from the Gaia DR2 catalogue, identifying stars in the Stream via their proper motions. Matching the Antoja et al. catalogue to the SDSS DR16 QSO catalogue (Lyke et al. 2020) on position, we find that some of the stars are actually known QSOs. To generate the Stream feature shown in Fig. 4.25 (top right panel), we remove the known QSOs and apply an $r > 18$ cut to only select faint stars. A fainter cut would better match the QSO selection, but the Antoja et al. catalogue does not contain a sufficient number of objects faint in r-mag to apply a fainter cut.

► **Three imaging footprints** The three footprints of the Legacy Imaging Surveys (North, South, DES; as defined in Section 3.2.3.1), exhibit distinct imaging properties. As shown in Fig. 4.26, while the PSF size in the z band is similar in the North and South, the r-band depth distributions are very different in the North, South and DES footprints. For instance, a PSF depth of 24.7 in r leads to a small overall target overdensity in DES (cf. Fig. 4.30) while it corresponds to an underdensity in the South (cf. Fig. 4.29). The imaging properties in the South footprint are similar in the North (NGC) and South (SGC) Galactic Cap. There is therefore no reason to split the South footprint in two.

In addition,the selection threshold in the QSO target selection is set independently in each of these three footprints. We therefore model systematic effects in the three footprints independently.

4.3.3.2 Systematic Model

We will identify each position on the sky with the corresponding pixel number i from a `HEALPix` pixelization. For each imaging footprint, we only consider pixels that contain at least one object from the random catalogues. The density of selected targets inside a pixel i is noted N_i. It is

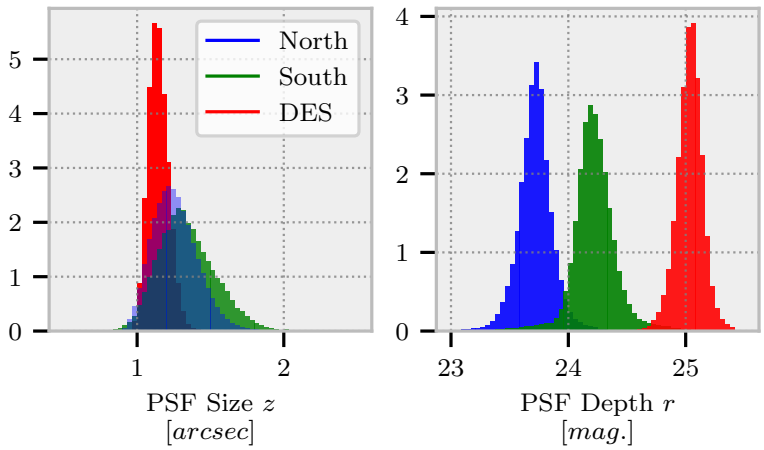


Figure 4.26: Density distribution of the PSF Size z and PSF Depth r for the three imaging footprints shown in Fig. 3.14. These two histograms show that a joint analysis on the three footprints is difficult since some features share the same support and others do not.

derived from the observational target density as

$$N_i^{\text{obs}} = f_i \times N_i, \tag{4.4}$$

where f_i is the observed fractional area of pixel i calculated as the number of random targets inside the pixel divided by the nominal density of randoms, and N_i is the true number of quasars in the pixel in the absence of any systematic effects. N_i is related to the mean target density N_0 by

$$N_i = N_0 \left(1 + \delta_i\right), \tag{4.5}$$

where δ_i is the overdensity that contains the cosmological information.

N_0 is unknown and has to be estimated from the target selection as the mean of the pixel density over the footprint weighted by f: $\hat{N}_0 = \langle N_i \rangle_i = \langle f_i^{-1} \times N_i^{\text{obs}} \rangle_i$. Since the target selection is contaminated by stars, we choose different regions known to be less contaminated to perform this estimation in the three footprints. We use the same region used to tune the probability selection during the quasar target selection, which was a box of R.A., Dec. $\in [120°, 240°] \times [32.2°, 40°]$ for the North footprint, a box of R.A., Dec. $\in [120°, 240°] \times [24°, 32.3°]$ for the South footprint and the entire footprint for DES. Note that even these regions are contaminated by stars, such that the actual QSO target density is smaller.

Systematic effects will be taken into account using an additional term F_i such that

$$N_i = N_0 \left(1 + \delta_i\right) \times F_i. \tag{4.6}$$

The aim of the imaging systematics mitigation is to describe F_i as a function of a set of observational features. These features must not depend on the sky position, to avoid suppressing the cosmological signal. We assume that the imaging systematics can be completely explained by our set of observational features. This assumption can be validated by the uniformity of the target selection density map after mitigation (cf. Section 4.3.4.3).

The features in the pixel i are denoted s_i which is an n-dimensional vector where n is the number of features ($n = 11$ in our case: 10 observational features from Section 4.3.3.1 and the Sagittarius Stream feature from Section 4.3.3.1). F should depend only on these features and not on the pixel number. We thus rewrite F_i as $F(s_i)$, and N now also depends on the observational features:

$$N_i(s_i) = N_0 \left(1 + \delta_i\right) \times F(s_i). \tag{4.7}$$

We denote $A(S)$ to be the subset of pixel numbers which belong to the same region S in the space of the feature maps. Averaging over many pixels will suppress the density contrast: $\langle \delta_i \rangle_{i \in A(S)} = 0$. Thus, if S is sufficiently large, the contamination signal is given by

$$F(S) = \frac{\langle N_i(S) \rangle_{i \in A(S)}}{N_0}. \tag{4.8}$$

This averaging is controlled by hyper parameters of the different regression methods. For instance in the particular case of the Random Forest (see Section 4.3.3.3), the averaging is controlled by the minimum number of objects in a leaf.

The systematic correction will be modelled by a weight to apply in each pixel, defined as

$$w_i^{\rm sys} = \frac{1}{F(s_i)}. \tag{4.9}$$

The correction is only efficient on scales at least as large as the typical size of the pixel at $N_{\rm side} = 256$, *i.e.* about $\theta = 0.22$ deg. Hence, the correction is constant within each pixel.

The regression is performed using only reliable pixels, which we choose to be pixels with $f_i > 0.9$, and is then applied to all pixels. This criterion removes pixels that contain too few targets, which could bias the regression. The pixels that are excluded represent only about 3.8% of the DR9 footprint, lying mainly at the edges of the footprint and in the region south of Dec. $< -10°$ in the NGC. For studies with a smaller pixel size, e.g. $N_{\rm side} = 512$, the nominal density of the randoms would have to be increased to limit the Poisson noise when determining f_i.

4.3.3.3 Regression Methods

We test different methods, utilizing the same feature set, to perform the regression presented above. The initial correction, obtained with a linear regression, turns out to be insufficient (as illustrated in Fig. 4.28 – 4.30), necessitating non-linear regression approaches. We therefore test two classical machine-learning methods based on `scikit-learn` (Pedregosa et al. 2011), namely, the Random Forest and Neural Network methods.

▶ **Linear** The contamination function F is described as a linear function of the observational features: $F(s_i) = a_0 + \sum_{j=1}^{11} a_j s_{ij}$. The coefficients a_j will be estimated with a least square minimization using the `iminuit` package (Dembinski et al. 2020). The χ^2 used for the minimization is defined as

$$\chi^2 = \sum_i \frac{1}{N_i} \left(F(s_i) - \frac{N_i}{N_0} \right)^2 + c_{\rm reg} \times \left(\langle F(s_i) \rangle_i - 1 \right)^2, \tag{4.10}$$

where $\sqrt{N_i}$ is an estimate of the error for the object count inside a pixel and $c_{\rm reg}$ is a penalty term to regularize the regression. Since the distribution of N_i/N_0 is not symmetric around 1, the higher number of pixels with $N_i/N_0 < 1$ forces the contamination function to not be centered around 1. We therefore use a penalty term to flatten the density map around the chosen mean density. The value of the penalty term depends on the number of pixels used to build χ^2. In our configuration, we use $c_{\rm reg} = 2e6$ and we check that as long as $c_{\rm reg}$ is sufficiently large, its value does not change the result of the regression.

▶ **K-fold training** Machine learning methods tend to overfit the data and have to be used carefully. Since we cannot create a training sample independent of the data set, we have to use a K-fold training to avoid over-training.

A K-fold training is a method that splits the data into K folds. For each fold, the method is trained with the remaining K-1 folds then the regression is performed on the isolated fold. This method guarantees that no data used for the regression is used for the training. A small value of K ensures no overfitting but the mitigation will be inefficient since the training set would not contain enough information. A high value of K ensures an efficient mitigation but is more prone to overfitting. In our case, we choose to work with 6 folds in the North and the DES footprint and with 12 folds in the South. Each fold has an area of about 830 $\deg^2$.

The locations of the folds have to be carefully chosen, since the contamination has distinct causes. For example, if all the borders of the Galactic plane in the South footprint were to be put in the same fold, the machine learning algorithm would not be able to explain related contamination with the K-1 remaining folds, since all the relevant information would have been removed from the training set. Therefore, we construct folds from small patches of the sky to spread the information across all folds. Such folds can be constructed easily in the `HEALPix` "nested" scheme using the `scikit-learn` function `GroupKFold`. The size of the patch matters, as patches that are too small lead to overfitting since all the information is present in all the folds. We use patches (groups in `scikit-learn` language) of 1000 pixels, which corresponds to an area of $\sim$52 $\deg^2$ for each patch.

The patch distribution is shown in Fig. 4.27, which demonstrates that no particular region is described by a single color. We checked that the estimated weight remains stable if we slightly vary the number of folds or the size of the patch, and used mocks (cf. Appendix 5.1) to ensure that no overfitting occurs.

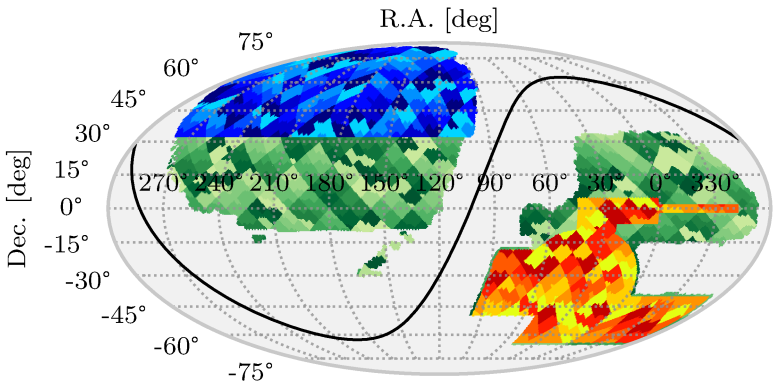


Figure 4.27: Distribution of folds across the three imaging footprints. There are 6 folds in the North (blue region), 12 in the South (green region) and 6 in DES (red region). Folds were split into small patches so the specific effects that contaminate target selection are always spread across several folds. The area of each patch is $\sim$52 $\deg^2$.

► **Random Forest (RF)** The first machine learning method used in our regression analysis is the well-known Random Forest algorithm. It is easy to parameterise and gives a helpful classification of the features as a function of their importance during the regression.

For the regression, we use the same set of hyper-parameters for each footprint and we do not normalise the data set. We choose a forest of 200 trees and we fix the minimum number of samples contained inside a leaf at 20. This means the average to estimate $F(S)$ (cf. Eq. (4.8)) is computed with at least 20 pixels. The mean number of pixels in each leaf is 80. We checked that the minimum sample size in a leaf has no strong impact on the regression during the K-fold training.

▶ **Neural Network (NN)** We also use the Multi-layer Perceptron (MLP) algorithm, a fully connected type of Neural Network. Finding the best hyper-parameters for a neural network is difficult. We base our choice on the result of a grid search performed on simulated QSO samples (*i.e.* mocks). The mocks used for this study are described in Appendix 5.1.

For the regression, we use the same set of hyper-parameters for each footprint, but when dedicated mocks for DESI become available, different hyper-parameterisations will be possible. We use an MLP with 2 hidden layers, comprising 10 neurons for the first layer and 8 for the second (*i.e.* a $(10, 8)$ formalism). The data are normalised on each fold dividing each feature by the standard deviation estimated in the considered fold. We hyper-parameterise the MLP with a sigmoid activation function, a batch size of 1000 and a tolerance of 1e-5. We use the Adam solver (Kingma and Ba 2015) to perform the minimization during the training.

4.3.4 Systematics Mitigation

We apply the method presented in Section 4.3.3 to correct for observational systematics in the DESI QSO target selection. We first describe the origin of these systematics and explain the role of the most important features. We then present the target density after it has been corrected by our systematic mitigation method.

4.3.4.1 Systematic Plots

To illustrate our method, we plot the relative QSO target density as a function of each observational feature. We will refer to these plots as "systematic plots". In our systematic plots, we center the relative density around 0. The goal of the correction is to obtain a relative density that is independent of the value of the observational feature. We produce systematic plots for the North (see Fig. 4.28), the South (see Fig. 4.29) and for DES (see Fig. 4.30).

Below, we give a brief description of the systematic plots in each footprint:

- North (Fig. 4.28): z and $W1/W2$ are crucial for the DESI QSO target selection. The plot of the relative density as a function of the PSF Depth z or the PSF Size z shows that when the z observational feature values are for bad observational conditions (small PSF Depth and high PSF Size), the discriminating power of the target selection algorithm is poorer and the relative density higher. In addition to the importance of the z band for the target selection, the z band benefits from better image quality (smaller PSF Size and higher PSF Depth) than g or r. The fluctuation of the relative density as a function of the g/r observational features are therefore weaker, but they follow the same general pattern. As explained in Section 4.3.3.1, the impact of the $W1/W2$ PSF Depths differs from the $g/r/z$ depths since the depth of the WISE colors are crucial for selecting QSOs. Hence, the number of targets increases with $W1/W2$ PSF Depth. In addition, the z PSF Depth is smaller in regions where the $W1/W2$ PSF Depth is larger, as shown in the bottom panels of Fig. 4.25. The combined effect increases the target density in this region.

- South (Fig. 4.29): the plots as a function of the z and $W1/W2$ PSF depths at high values exhibit similar behaviour as in the North. However, the effect is less significant since it is mainly visible in the R.A., Dec. $\in [210°, 270°] \times [15°, 30°]$ region. In comparison to the North, the plots as a function of the $g/r/z$ features all show similar trends to each other since these bands were collected with the same camera. The z-band dependence is the strongest, as expected since it is one of the most important bands for the QSO selection. The plots as a function of the $W1/W2$ PSF Depth exhibit seemingly unexpected

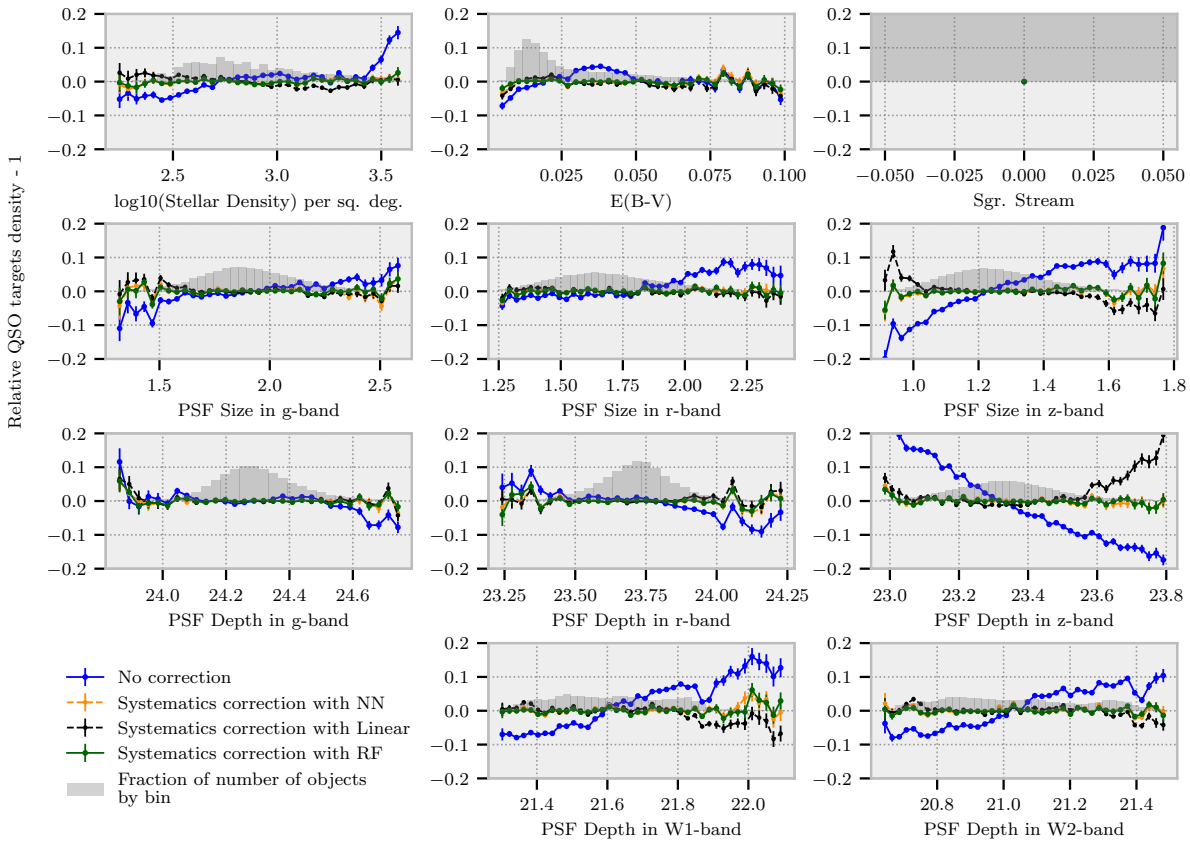


Figure 4.28: Relative QSO target density in the North footprint as a function of each observational feature. The relative QSO target density is a mean value after rejecting outliers *i.e.* pixels with a coverage lower than 90%. The blue lines depict the raw DESI QSO target selection. The green (*resp.* yellow / black) lines depict the QSO target selection after correcting for systematic effects using the RF (*resp.* NN / Linear) regression. The histogram represents the fraction of objects in each bin for each observational feature and the error bars are the estimated standard deviation of the normalized target density in each bin. The three methods all successfully flatten the relative QSO target density as a function of each observational feature. However, the linear method is less efficient than the other two methods.

behaviour at low values of the PSF Depth, where the relative density is almost flat. As explained in Section 4.3.2, the excess of stars near the Galactic plane or the Sagittarius Stream counteracts any expected decrease of targets due to lower values of the $W1/W2$ PSF Depth (cf., in particular, the R.A., Dec. $\in [120°, 150°] \times [-10°, 15°]$ region). The plots in the top panel of Fig. 4.29 (stellar density, dust and Sagittarius Stream) indicate a higher relative density due to the presence of stars. These features explain the stellar contamination near the Galactic plane and inside the Sagittarius Stream. Some stars have similar colors to QSOs and are therefore selected as QSO targets. More stars thus enter the QSO selection in regions of higher stellar density.

- DES (Fig. 4.30): The fluctuations of the relative density are much lower than in the two other footprints. The observational features for $g/r/z$, and especially for the z band, are better in DES than in both the South and the North (cf. the median value of these features in Table 3.3 and the distribution of the PSF Depth Size z and the PSF Depth r in Fig. 4.26). This results in excellent discrimination of QSOs from stars in the target selection process. So, DES is the least contaminated region and exhibits smaller fluctuations in relative density as a function of the observational features.

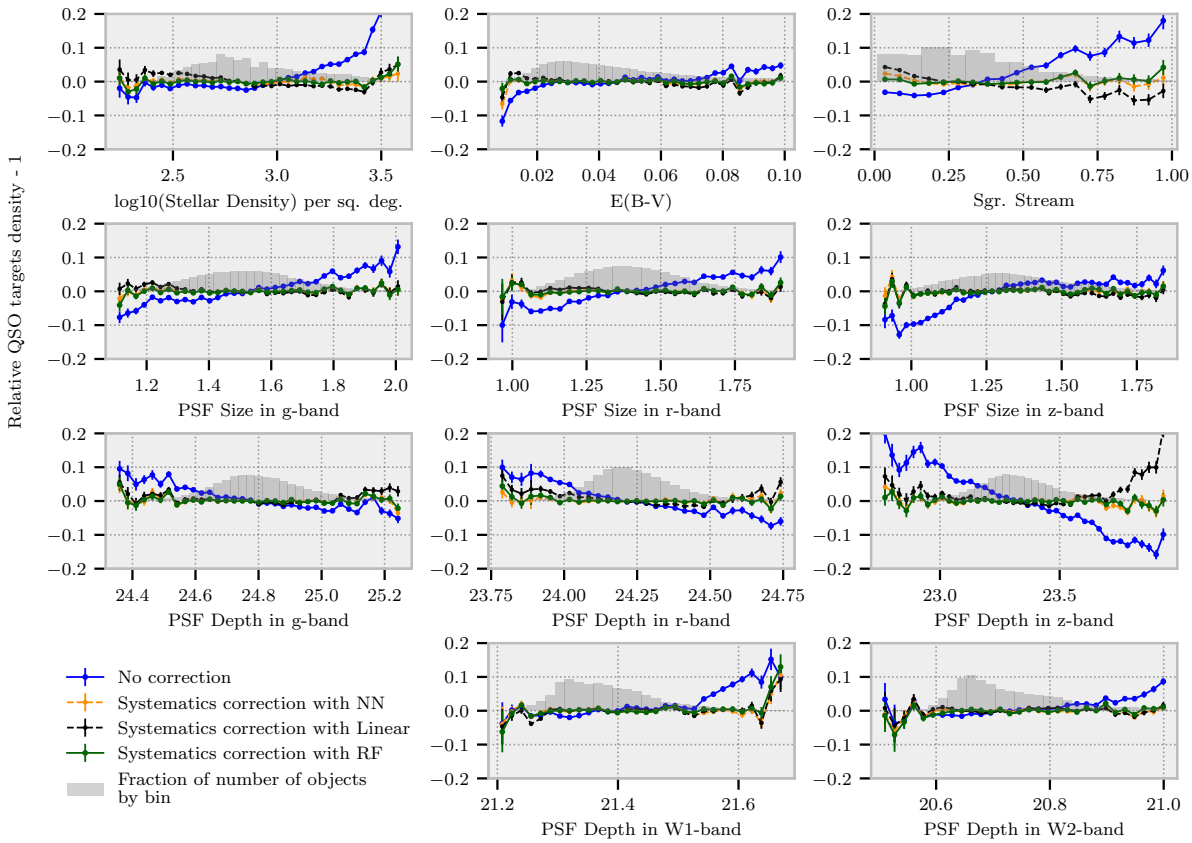


Figure 4.29: Same as Fig. 4.28 but for the South imaging footprint.

In Fig.s 4.28 – 4.30, we also plot the three different regression methods we apply for each footprint: RF in green, NN in yellow and Linear in black. These plots show that the mitigation works well and correctly flattens each feature. The difference in efficiency between the Linear and the non-linear methods is particularly obvious for the North and South footprints, where the contamination is stronger. For example, the linear correction fails when the relative density as a function of the PSF depth z is large, as shown in Fig. 4.28. This illustrates the non-linearity of the contamination and justifies the use of the machine learning methods. Both the RF and the NN perform well in correcting the non-linear systematics, although small differences can be found between these two methods. A more detailed comparison between the RF and NN will be done in Section 4.3.5.2.

4.3.4.2 Importance Features

The Random Forest algorithm includes a specific tool called *importance features*. Importance is a measure of which features most affect the regression. The importance features metric for the RF implemented in `scikit-learn` is based on the mean decrease in impurity (MDI), which is also called the Gini importance. The metric corresponds to averaging (over all the trees of the forest) the impurity reduction of each node weighted by the ratio of the training set passing through each node. The *permutation importance* was also tested and yields similar result in our data set than the Gini importance.

A feature with a higher importance value is more discriminating than a feature with a lower importance value, so the importance values are useful to identify the observational features that lead to contamination, to first order. It is worth noting that a feature with a low importance

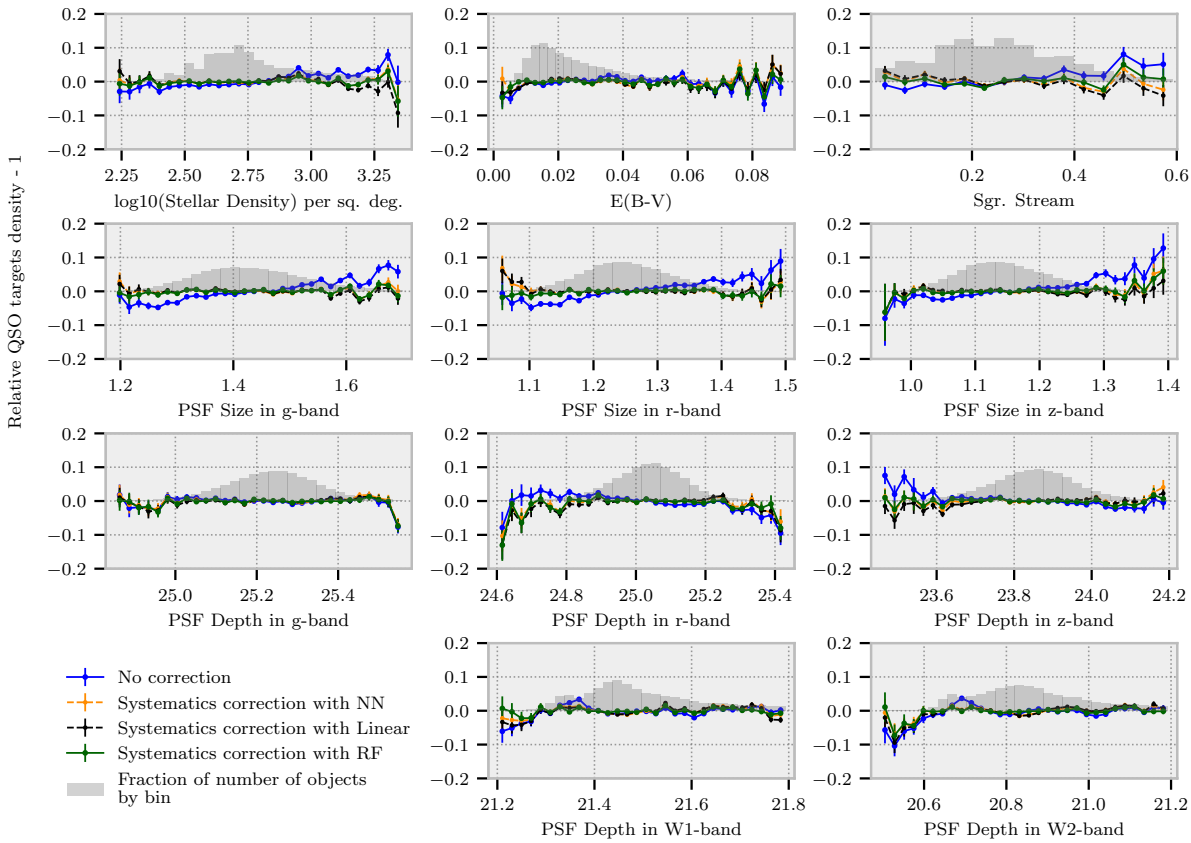


Figure 4.30: Same as Fig. 4.28 but for the DES imaging footprint.

value is not necessarily useless and can still improve the training compared to a case where we remove it. For instance, the Sagittarius Stream feature is necessary to correct the over-density in the Sagittarius region but it is not a high-value importance feature since it is useful only for a small number of pixels that vary in a manner that is quite different from other pixels in the footprint. This well-known bias can be circumvented using the *permutation importance*, where the Sagittarius Stream feature is ranked as one of the most important in the South.

In Fig. 4.31, we plot the importance for each feature. We only plot the six most important features for each footprint, since the other features have about the same value as the sixth most important feature. We recover the expected most important features described in Section 4.3.4.1:

- North: The importance feature analysis makes it clear that the PSF Depth z and the PSF Depth $W2$ play a key role in the contamination of the target selection in the North.

- South: The South region is strongly contaminated by stars from the Galactic plane as highlighted by the importance of the stardens and EBV features. The Sagittarius Stream feature has a low importance even if it is crucial to correct the Sagittarius Stream region (as explained above). The PSF Depth z plays an important role in this overdensity, as already noted. Surprisingly, the $W2$ PSF Depth does not appear as an importance feature, even though it explains the under-density near the anti-Galactic pole.

- DES: No clear importance feature emerges, as expected given that the DES region is the least contaminated, most uniform, region.

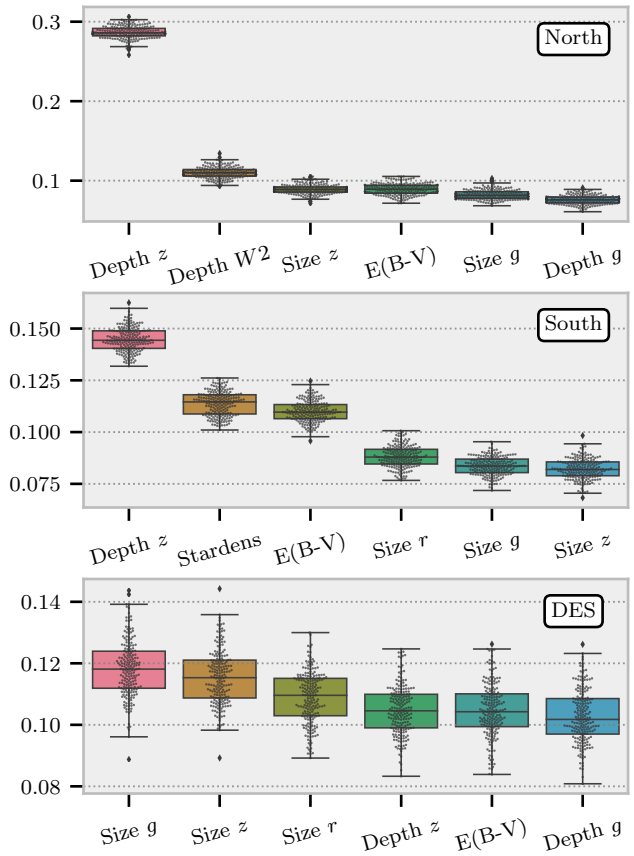


Figure 4.31: Feature importance calculated using the Random Forest method. Only the 6 most important features are plotted for each footprint. Each dot represents the feature importance value found with one decision tree. The box plot is assembled across all the trees of the forest. The values of the importance's cannot be compared between the three regions.

4.3.4.3 Quasar target selection after correction

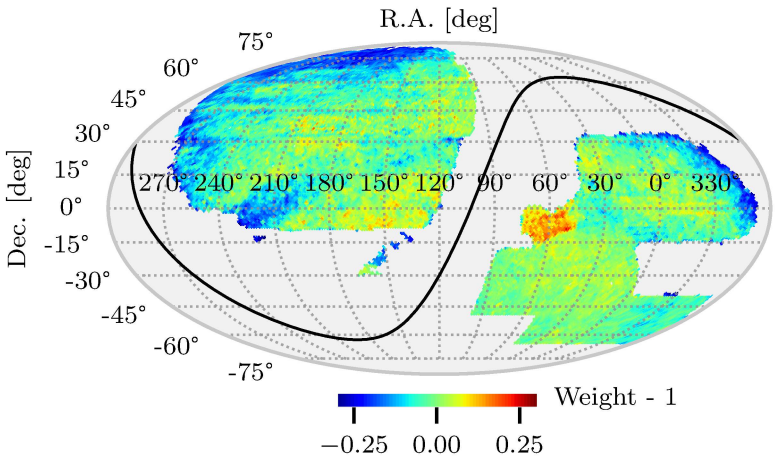


Figure 4.32: Distribution of the systematic correction weights (weight - 1) across the Legacy Imaging Surveys footprint.

We construct a pixel weight map to correct the over or under densities of the target selection. Fig. 4.32 shows the weight map obtained with the RF regression. We then multiply the density

of the DESI QSO target selection by the weight on a pixel-by-pixel basis. Fig. 4.33 shows the corrected QSO selection. The map is almost completely uniform at this resolution. The largest overdensities vanish after mitigation, which confirms that the set of observational features that we considered suffices to explain most of the observed large-scale target density variations.

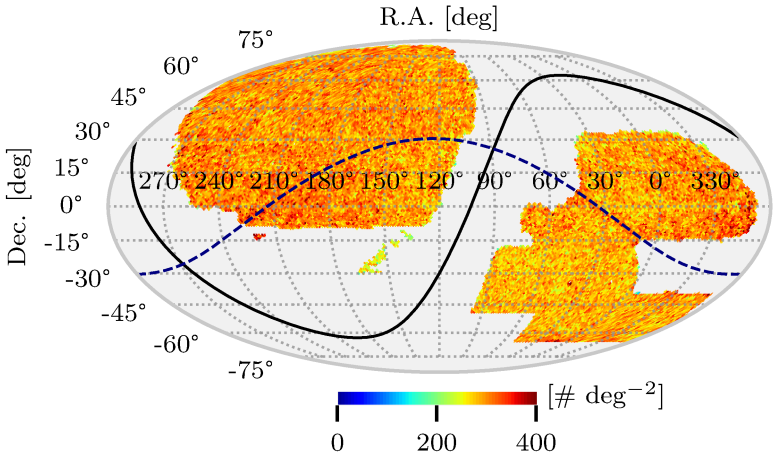


Figure 4.33: Density map of the QSO target selection after mitigating systematics using the RF method. The density map is quasi-uniform compared to the initial density map shown in Fig. 4.23. The solid black line depicts the Galactic plane and the blue dashed line depicts the plane of the Sagittarius Stream.

The density distribution of the QSO target selection is shown in Fig. 4.34. The full histograms are before applying the systematic mitigation and the lines are for after. The systematic mitigation acts on the width of the histograms removing the over or under-density pixels. This effect is smaller in DES than in the North and in the South which confirms the visual inspection of the Fig. 4.23 and 4.33. The difference in mean density between DES and the two other photometric footprints was mentioned in Section 4.2.1.

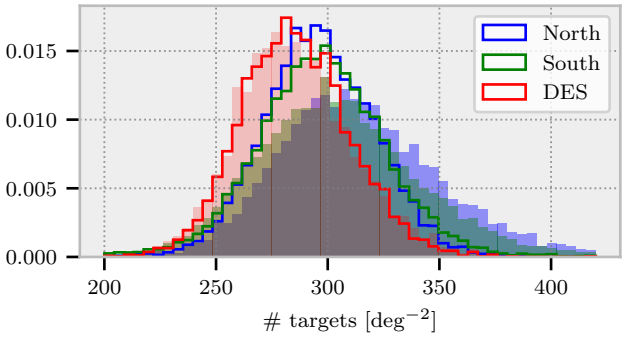


Figure 4.34: Density distribution of the QSO target selection for non corrected (full histogram) and corrected (line) cases in the three photometric footprints. After the correction, the width of the histogram is smaller in each photometric footprint. As mentioned in Section 4.2.1, the density in DES is lower than in the North or in the South.

4.3.5 Angular Correlation and Clustering

Our systematic mitigation method is able to correct fluctuations in density as a function of our observational features. In this section, we measure the impact of the systematic mitigation on the angular correlation function.

4.3.5.1 2-point Correlation function

► **Definition** The 3-dimensional 2-point correlation function $\xi(\mathbf{r})$, see Section 1.2.1.2, describes the excess probability to find a pair of objects inside two infinitesimal volumes dV_1 and dV_2 separated by $\mathbf{r}$:

$$dP(\mathbf{r}) = \bar{n}^2 \, (1 + \xi(\mathbf{r})) \, dV_1 dV_2, \tag{4.11}$$

where $\bar{n}$ is the mean density. It is related to the contrast density by $\xi(\mathbf{r}) = \langle \delta(\mathbf{x}) \delta(\mathbf{x} + \mathbf{r}) \rangle_{\mathbf{x}}$. The cosmological principle ensures that ξ depends only on r.

The same definition can be extended to the 2-dimensional case *i.e.* the angular correlation where volumes are replaced by solid angles and distances by angular distances:

$$dP(\theta) = n_0^2 (1 + \omega(\theta)) d\Omega_1 d\Omega_2, \tag{4.12}$$

where n_0 is the mean angular density.

► **2-point correlation function estimator** We estimate the angular correlation function $w(\theta)$ using the Landy-Szalay estimator derived in Landy and Szalay (1993):

$$\hat{w}(\theta) = \frac{aDD - bDR + RR}{RR}, \tag{4.13}$$

where DD, DR, RR refer to the weighted pair counts $data-data$, $data-random$ and $random-random$ with an angular separation θ. The normalization terms are

$$a = \frac{\sum\limits_{i \neq j} w_i^R w_j^R}{\sum\limits_{i \neq j} w_i^D w_j^D} \quad \text{and} \quad b = \frac{\sum\limits_{i \neq j} w_i^R w_j^R}{\sum\limits_i w_i^D \sum\limits_j w_j^R}, \tag{4.14}$$

where w_i^D (*resp.* w_i^R) is the weight for the data (*resp.* random). The Landy-Szalay (LS) estimator is known to be unbiased and to have minimal variance in the limit of an infinitely large random catalogue with a volume greater than the scales considered, and for weak correlations ($w(\theta) << 1$).

We use the package `CUTE`[9] (Alonso 2012) to perform the estimation. `CUTE` is a fast implementation written in C and using OpenMP and MPI.

4.3.5.2 Angular correlation of the DR9 quasar target selection

The systematic mitigation method is checked to avoid over-fitting. This is done in Appendix 5.1 by applying the mitigation method to contaminated mocks. Then, we compute the angular correlation of the DR9 quasar target selection corrected by the systematic weights that we calculated in Section 4.3.4.

► **Comparison with SDSS DR16** We calculate the angular correlation function of the raw and corrected (RF method) QSO target densities for the three different imaging footprints. For comparison, we also calculate the angular correlation function of SDSS DR16 quasars. Note that the correlation with SDSS quasars cannot be computed at angles smaller than 62 arsec due to fiber collisions (e.g. Dawson et al. 2016). For SDSS DR16 quasars, we use the systematic weights provided in Rezaie et al. (2021) based on a neural network treatment (see Ross et al. 2020, for

[9] https://github.com/damonge/CUTE

the standard treatment). The results are shown in Fig. 4.35. Note that the error bars shown in Fig. 4.35 are the standard deviation of the LS estimator (Landy and Szalay 1993) except for the correlation function of SDSS quasars, for which the errors are estimated using the standard deviation across 100 EZ-mocks from Zhao et al. (2021).

After mitigating for systematics, the angular correlation in the North and in the DES region (again, see Fig. 3.14 for the definition of these regions) are comparable with the correlation computed with the SDSS DR16 sample. The SDSS DR16 sample has been carefully corrected for systematic effects (as derived in previous work), and so is expected to be largely free from any contamination. In the South, even after mitigating for systematics, we do not recover the same level of correlation as for the SDSS DR16 sample. Reasons for this difference are discussed further in Appendix 5.2.

It is worth noting that the correction is larger at larger angles, *i.e.* the impact of systematic effects is higher on large scales. Therefore, mitigation of photometric systematics is critical for studies that require information from large scales, such as studies of primordial non-Gaussianity.

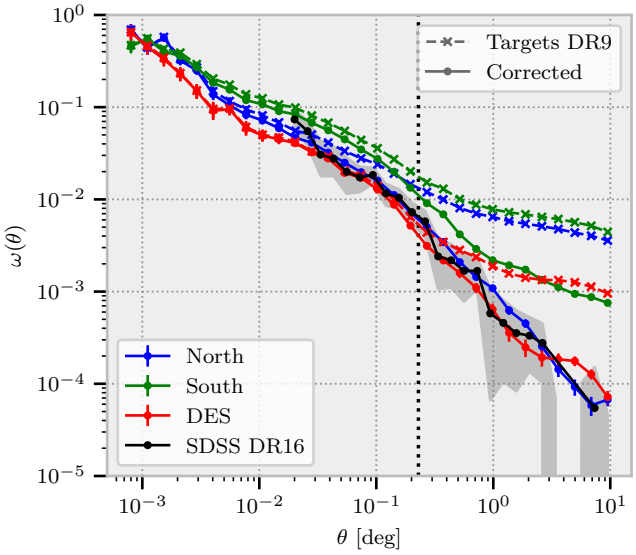


Figure 4.35: The angular 2-point correlation function of the DR9 QSO targets. The dashed lines represent the nominal DESI QSO target selection. The solid lines are for the corrected QSO target selection with RF method. The black line is the angular correlation function from SDSS DR16 (which cannot be calculated at angles smaller than 62 arcsec, due to fiber collisions). The grey region despicts the error for SDSS DR16 estimated as the standard deviation across 100 EZ-mocks. The black dotted line corresponds to the typical resolution of our regression analyses *i.e.* the pixel size at $N_{\rm side} = 256$. The solid green line is discussed in detail in Appendix 5.2.

► **Regression method comparison** Fig. 4.36 shows the difference between the three methods introduced in Section 4.3.4 on the North, on the South and on DES. As expected, the linear method is less effective than the other two in the highly contaminated North (top panel), highlighting the necessity of our more complex machine-learning-based regressions. The two machine-learning-based methods give similar results in the North. None of the three methods properly correct for the contamination in the South (middle panel); this is discussed further in Appendix 5.2.

The information in Fig. 4.30 suggests that all three methods introduced in Section 4.3.4 are quite effective at mitigating systematics in the DES region. However, the bottom panel of Fig. 4.36 demonstrates that the linear method and the NN method less effectively correct the

angular correlation function in the DES region as compared to the RF method. The difference between the linear and the RF methods comes from the non-linear part of the contamination in this region. The systematics plots show that in the highly contaminated regions, the linear method corrects less than the RF.

The difference between the NN and the RF is more subtle since the training information, *i.e.* the area and the chosen folds in that area, is similar. The explanation comes from the difference between the two algorithms. The NN method is less efficient to correct small, highly contaminated regions. The RF creates boxes in feature space and separates the most contaminated regions from the rest. The estimation of the correction weight is then possible everywhere.

To solve this problem, a regularization term can be introduced to force the NN to also consider small regions. The choice of the regularization value strongly depends on the size of the small regions and on their location in feature space. Without realistic mocks for the DESI QSO sample, this additional hyper-parameter cannot be easily optimized. The training time for the NN also varies considerably as a function of the value of the tolerance chosen to sample the NN hyper-parameters, whereas, with our parameterisation, the RF method is quicker to train. Since the RF correction already obtains good results in DES, we do not need to improve the current NN method. We leave fine optimizing of the NN method for future work when realistic DESI QSO mocks become available.

► **Resolution of the correction** The size of the pixels used when determining weights to mitigate systematics is critical, because it gives the scale at which the correction is most effective. We perform all the analysis above with $N_{\rm side} = 256$ corresponding to a characteristic angle of ∼0.22 deg *i.e.* 12.6 h^{-1}Mpc at $z \sim 1.7$ (Planck 2015; Planck Collaboration et al. 2016). This pixel size is chosen such that there are sufficient targets inside each pixel to learn about contamination without introducing significant Poisson noise in the mitigation.

For the DESI QSO target selection, a pixel with a size of $N_{\rm side} = 256$ typically contains a median of ∼16 QSO targets. At a size of $N_{\rm side} = 512$, the number of targets decreases to ∼4. With such a small number of targets, the per-pixel density is too noisy for machine learning methods to relate fluctuations in density to observational features. For the same reason, we needed to increase the pixel-size used in our analysis to $N_{\rm side} = 128$ when studying eBOSS mocks, since the density of the eBOSS mocks was lower than the DESI QSO target selection density (see Appendix 5.1).

Other DESI targets, such as ELGs and LRGs, will have a higher density and it will be possible to decrease the size of the pixels used to estimate correction weights by a factor of 2–4.

► **Systematic checks: restrictive QSO target selection** Since the DESI QSO targets are selected with an RF classification, the stellar contamination depends on the value of the selection threshold. This dependence propagates into the angular correlation, since we broadly expect a lower r_0 for a less contaminated sample.

Fig. 4.37 shows the angular correlations in the North for DESI QSO targets with different values of the selection threshold (dashed lines) and of the corresponding corrected samples (full lines). The systematic weights are generated with the RF method for each target selection. The nominal probability in the North footprint used in the QSO selection depends on r and is given by $p(r) = 0.88 - 0.04 \times \tanh(r - 20.5)$. This threshold is lower than for the other regions. When the threshold increases, the selection is less contaminated since the RF will select a higher fraction of bonafide QSOs.

The amplitude of the angular correlation in Fig. 4.37 decreases as the probability threshold

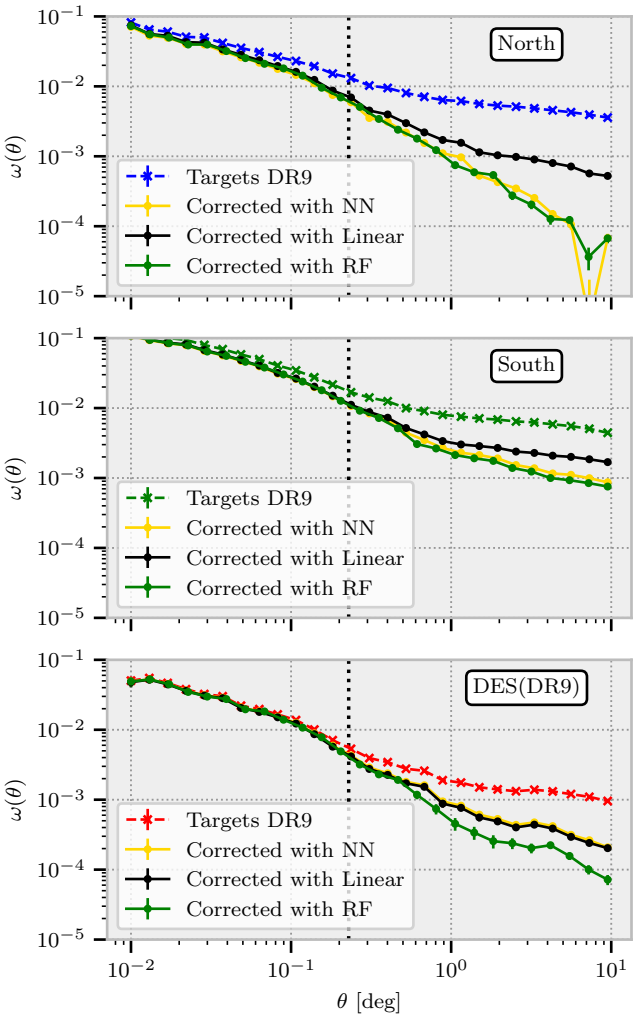


Figure 4.36: The angular correlation function for the three different methods introduced in Section 4.3.4 for the three regions highlighted in Fig. 3.14. The Linear method is less effective than the two machine-learning-based methods, especially in the most strongly contaminated region (the North). The correction in the South does not improve properly the correlation for any of the three methods. The RF and NN methods produce similar results in the North but slightly different results in DES. This is because the RF is more robust to sampling high levels of contamination from a small number of pixels.

is increased, suggesting that the excess of correlation is mostly due to stellar contamination. This also confirms that our method to mitigate the systematics does not over-fit the data since when the contamination is removed, the correlation converges to the same level as the SDSS QSO correlation (cf. Fig. 4.35).

Finally, we note that the excess correlation in the South is *not* reduced when the probability threshold of the RF selection is increased. This is because many stars in the Sagittarius Stream have selection features that resemble QSOs. These stars are assigned near unit probability by the DESI QSO targeting algorithm, such that they are not removed by a more restrictive selection.

4.3.5.3 Limber parameters

▶ **Limber approximation** One can use the Limber approximation to relate the angular 2-point correlation function to the 3D one with (1.47). The integral calculation requires a fiducial

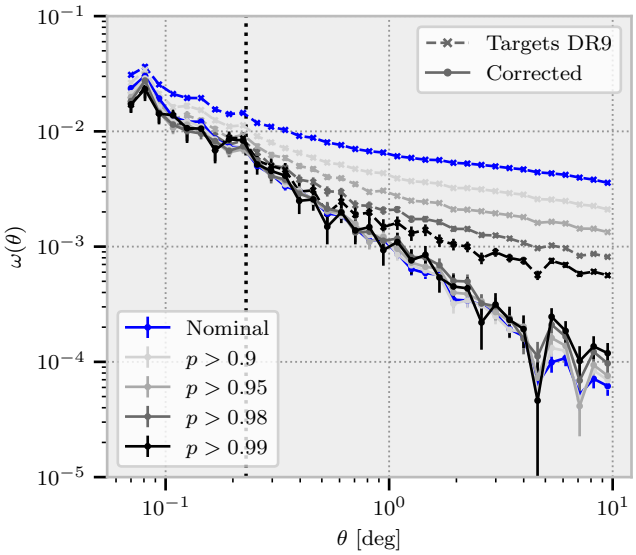


Figure 4.37: Evolution of the angular correlation function in the North region when the probability threshold of the quasar target selection is increased. The nominal probability in the North is $p(r) = 0.88 - 0.04 \times \tanh(r - 20.5)$. The dashed lines depict raw target selections with different probability thresholds. The solid lines show the corrected versions of the same selections. The black dotted line corresponds to the typical resolution of the correction *i.e.* the pixel size at $N_{\rm side} = 256$.

cosmology. We choose a Λ cold dark matter (ΛCDM) cosmology following the Planck 2015 parameters from Planck Collaboration et al. (2016): $\Omega_{m,0} = 0.308$, $\Omega_{\Lambda,0} = 0.691$, $\Omega_{b,0} = 0.048$, $h = 0.677$, $\sigma_8 = 0.815$, $n_s = 0.967$.

► **Fitting of Limber parameters** The power-law parameterisation of the correlation function $\xi(r)$ can be then constrained by the angular correlation function $w(\theta)$ at small angles. We proceed by estimating these power-law parameters and comparing them to previous measurements done with SDSS data by Myers et al. (2009) and with 2dF data by Croom et al. (2005).

The Limber parameters are estimated in the three imaging regions highlighted in Fig. 3.14 and also in the DES region with only Dec. > -30, which we will refer to as DES(Dec. > -30). This region will almost correspond to the intersection between DES and the expected nominal

Table 4.5: Limber parameters for the DESI QSO target selection in the imaging footprints depicted in Fig. 3.14. The errors are estimated using a sub-sampling method. We provide the measurements for both the non-corrected and corrected cases. The correction is performed with the RF method. $\bar{z}$ is the mean redshift of the sample.

	r_p min $[h^{-1}$Mpc$]$	r_p max $[h^{-1}$Mpc$]$	$\bar{z}$	r_0 $[h^{-1}$Mpc$]$	γ	r_0 $[h^{-1}$Mpc$]$	γ
Croom	1.0	25	1.35	-	-	5.84 ± 0.33	1.64 ± 0.04
Myers	1.6	40	2	-	-	4.56 ± 0.48	1.5 (fixed)
				DESI QSO Targets		Corrected	
North	0.045	45	1.7	10.15 ± 0.70	1.61 ± 0.03	7.49 ± 0.57	1.89 ± 0.02
South	0.045	45	1.7	12.88 ± 0.95	1.64 ± 0.03	10.33 ± 0.84	1.80 ± 0.02
DES (Dec. > -30)	0.045	45	1.7	6.76 ± 0.58	1.78 ± 0.04	6.15 ± 0.45	1.88 ± 0.04
DES	0.045	45	1.7	7.19 ± 0.34	1.79 ± 0.03	6.47 ± 0.31	1.89 ± 0.02

DESI footprint. The results are shown in Table 4.5 and fit for the corrected DESI QSO targets in DES region is plotted in Fig. 4.38. The correlation function is fitted from $1e^{-3}$ to 0.8 deg, corresponding to a transverse separation of 0.045 to 45 h^{-1}Mpc at redshift 1.7. The mean redshift of the DESI QSO targets (~ 1.7) is obtained thanks to the selection function introduced in Section 4.3.5.3 which corresponds to the estimated redshift distribution of the sample.

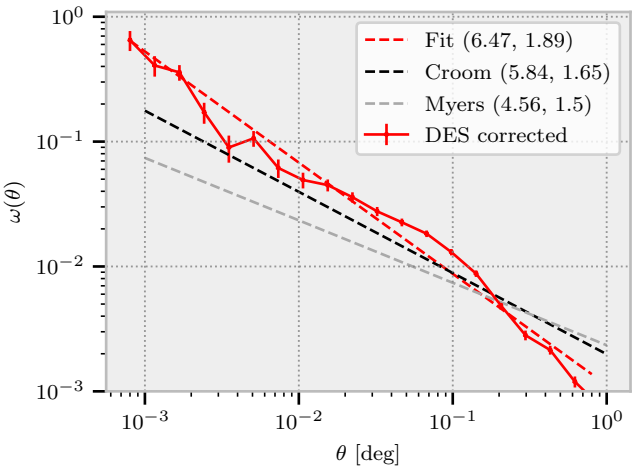


Figure 4.38: Fit of the Limber parameters in the DES region for the corrected DESI QSO targets (red line). The Limber parameters (r_0, γ) are given in each case. The black (*resp.* grey) dashed line shows the Limber function measured by Croom et al. (2005) (*resp.* Myers et al. 2009). The parameters for the other regions are given in Table 4.5.

The errors on the parameters r_0 and γ are estimated using a sub-sampling method: 10 (*resp.* 18, 10, 4) sub-regions of similar area (~ 450 deg^2) are used for the North (*resp.* South, DES, DES Dec. > -30)). Sub-sampling is used to probe the variability of the angular correlation function in different areas of the footprint, where stellar contamination and systematic effects should differ. If the systematics are properly mitigated, the error of each parameter should be reduced *i.e.* the corrected angular correlation functions over each sub-region should be more similar.

The value r_0, which parameterises the amplitude of the correlation function, captures the offset of different measured angular correlation functions. For instance, the fact that the angular correlation in the South is higher than in DES, manifests in a value of r_0 that is higher in the South than in DES. As systematics tend to lead to an increase in amplitude, r_0 tends to be higher when systematics are dominant. For example, the fact that the value of r_0 that we measure in the DES footprint is generally comparable to the values found for previous measurements (Croom et al. 2005, Myers et al. 2009), suggests that systematics have been mitigated well in the DES region.

The parameter γ describes the slope of the angular correlation. A higher value of γ means a steeper slope. It is worth noting that since the relevant correction scales of the systematic weights are larger than the pixel size used for our analyses (~ 0.2 deg), the angular correlations of the corrected targets have a steeper slope than for the raw target samples. The comparison of γ between the DESI QSO targets and previous measurements is not particularly relevant since any excess correlation caused by systematics on small scales, *i.e.* below the typical angular resolution of our corrections, cannot be removed with our mitigation procedure.

Though no spectroscopic confirmation of DESI QSO targets is available across a significant fraction of the footprint used in our study, our correction for angular systematics enables us to provide a Limber amplitude parameter r_0 which is consistent with those found by Croom

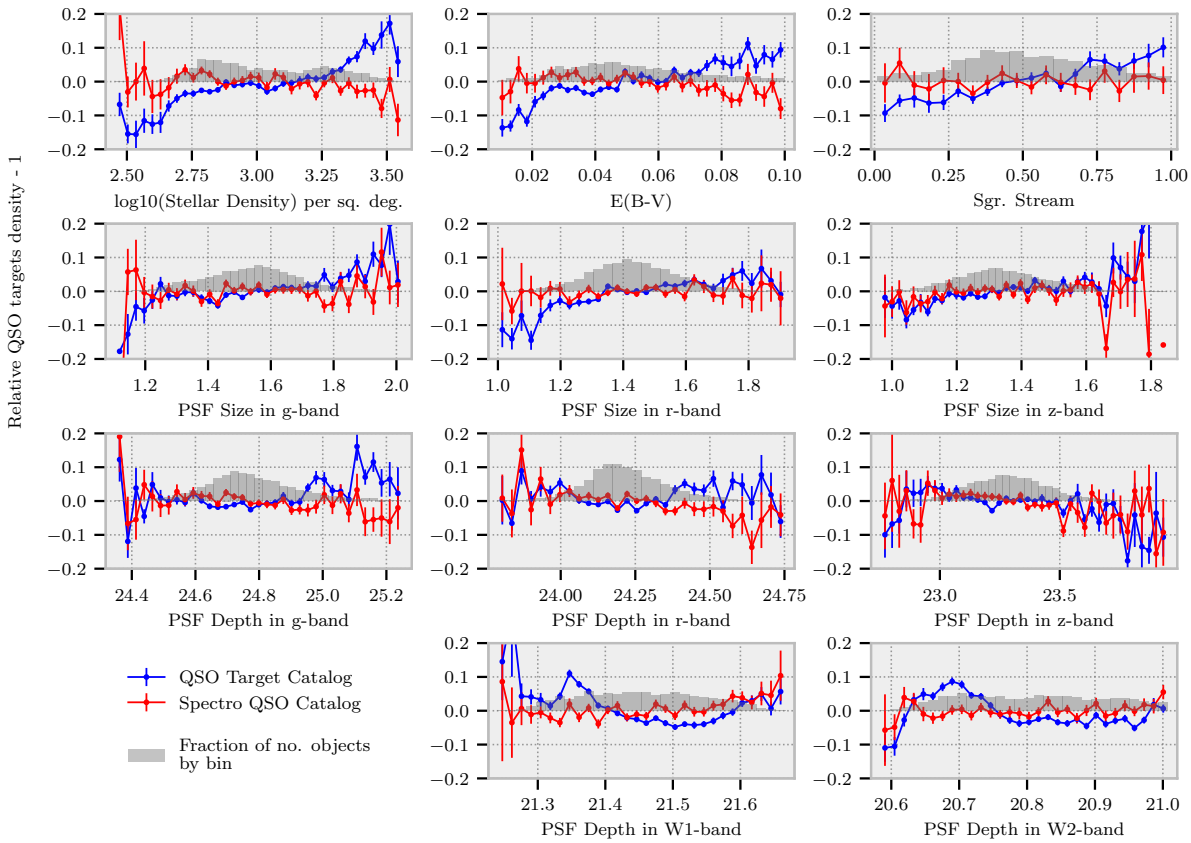


Figure 4.39: Relative QSO target or QSO densities for a part of South (Non-DES) region as a function of each observational parameter. The blue curve represents the density for all the QSO targets while the red curve is obtained for the spectroscopically confirmed QSOs only.

et al. (2005), Myers et al. (2009) using spectroscopic data. Nevertheless, the fitting of the Limber parameters will be greatly improved by information from the DESI spectroscopic survey. Follow-up spectroscopy will enable us to remove stars from the DESI QSO target sample, so any clustering analyses will be free of the stellar contamination that increases the amplitude of the angular correlation measurements.

4.3.6 Imaging systematics validation

In order to identify the main sources of systematic effects in the QSO target selection, Fig. 4.24 shows the relative QSO target density as a function of each observational parameter. Those plots are obtained for all the QSO targets, including the contaminants, galaxies and stars. It is particularly interesting to reproduce such a figure for the objects retained in the automated QSO catalog, presented in Section 4.2.3.2, since the purity of this catalog is $> 99\%$ *i.e.* all the contaminants were removed.

Fig. 4.39 allows us to compare the systematic effects for the original QSO target catalog (blue curve) and the catalog of QSOs spectroscopically confirmed (red curve) in a part of South (Non-DES) region, observed during the first two months of main survey. This subset of DESI footprint corresponds to a region, strongly contaminated by stars from the Sagittarius Stream and therefore particularly interesting to study. We observe that all the strong trends in the target densities related to the stellar density in the Galactic plane or the Sagittarius Stream, vanish in the spectroscopic catalog. Weaker systematic effects are also removed. The method

presented in this section should suppress the remaining effects.

4.4 DESI QSO Y1

With a 310 deg^{-2} target density, the main survey selection allows DESI to select more than 200 deg^{-2} quasars, including 60 deg^{-2} quasars at redshifts $z > 2.1$, exceeding the project requirements by 20% (DESI Collaboration et al. 2016b), and since QSOs have the highest priority during the observation, the Y1 QSO sample is already expected to have competitive results compare to the ones from the QSO of eBOSS DR16 (Neveux et al. 2020, Hou et al. 2021, Mueller et al. 2022).

4.4.1 Footprint and QSO density

The DESI Y1 data release contains all the data of the main survey from May, 14$^{\text{th}}$ 2021 up to the Contreras fire in early June 2022. Fig. 4.40 shows the number of passes in the Y1 footprint. Survey coverage is unfortunately uneven across the 7560 deg^2. However, the region with at least three passes covers about 2,100 deg^2, and the QSO sample can be almost considered as a complete sample in this region.

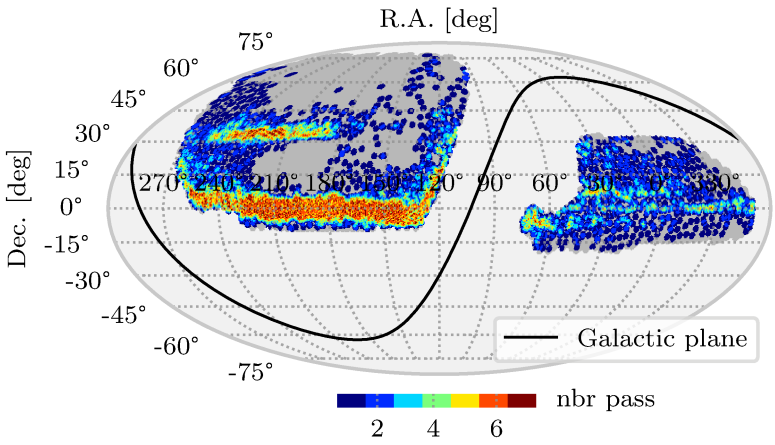


Figure 4.40: Number of passes in Y1 footprint. The QSO sample is 55% (*resp.* 80%) complete after one pass (*two passes*) and can be almost considered complete in the three passes region. The three passes region covers about 2,100 deg^2.

For the quasars, all the QSO targets which are detected as a quasar by the pipeline presented in Section 4.2.3.2 with $0.8 < z < 3.5$, will be kept in the final catalog used for the clustering measurement. The creation of the clustering catalogs is briefly explained for all the tracers in DESI Collaboration et al. (2023b). Fig. 4.41 shows the density distribution of the QSO sample for clustering corrected by the completeness. The map shows no significant over-density regions, illustrating the stellar contamination that occurs in the Sagittarius stream or near the galactic plane as shown in Fig. 4.23. Note the difference in density between the North and the South photometry.

Fig. 4.42 shows the redshift distribution of all the quasar targets that are identified by the pipeline, see Section 4.2.3.2, as true quasars. In particular, the standard clustering measurements, as defined in DESI Collaboration et al. (2016a), will use the quasars with $0.8 < z < 2.1$. To increase the volume probed and hence reduce the statistical errors on power spectrum measurement, we will also use the quasars with $2.1 < z < 3.5$.

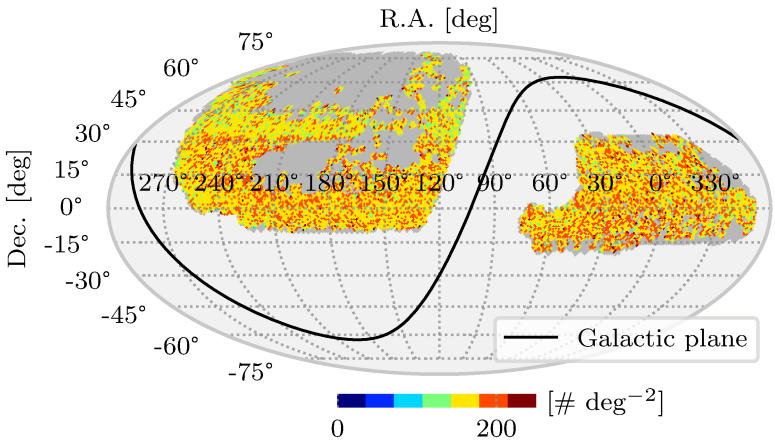


Figure 4.41: Y1 QSO ($0.8 < z < 3.5$) density distribution corrected by the completeness. This map does not exhibit strong over or under-density region. In particular, there is no overdensity along the Sagittarius stream or near the galactic plane.

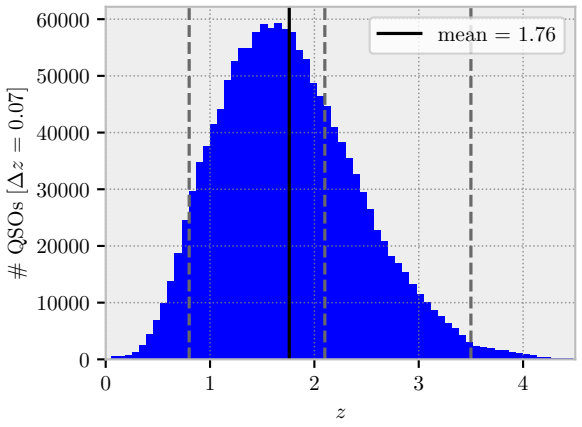


Figure 4.42: Y1 QSO redshift distribution. The clustering catalog will contain all quasars with $0.8 < z < 3.5$.

4.4.2 Imaging systematics

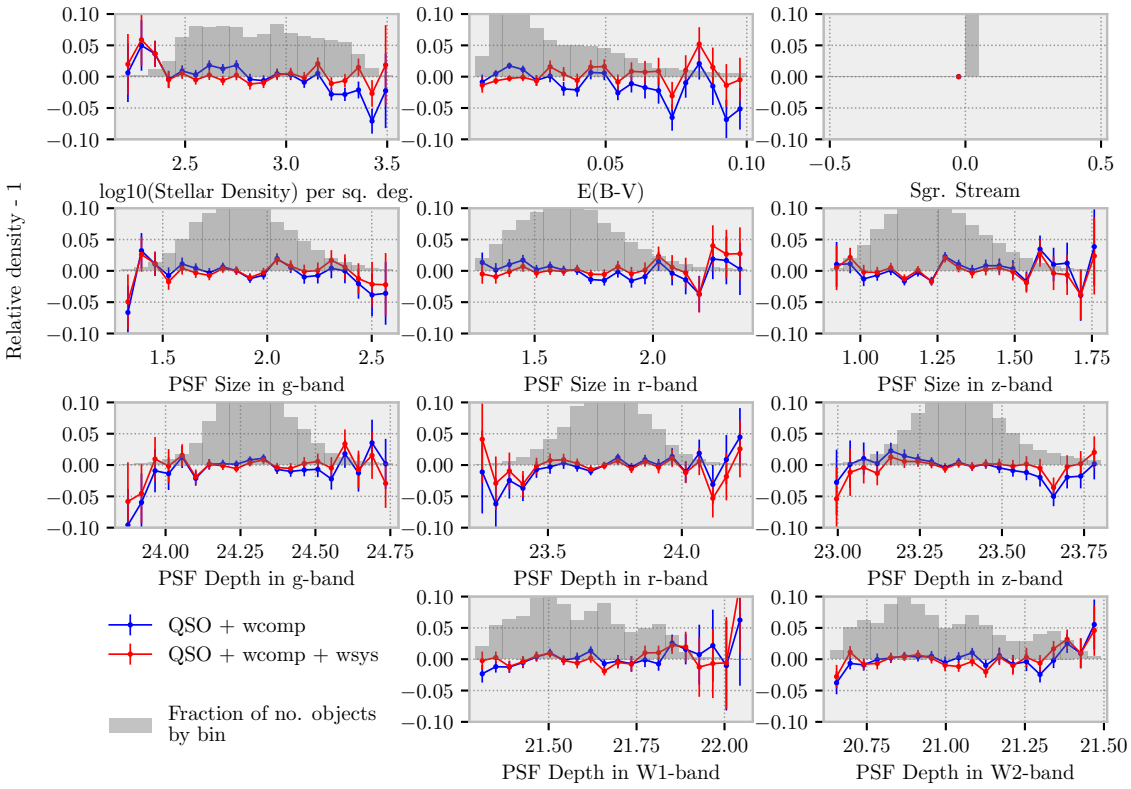
Since the regions South NGC and South SGC will not be used together during power spectrum measurements, we show systematic plots for the two separate regions, Fig 4.43b-4.44a. Although the imaging systematic effects are much less pronounced than in the case of targets, the remaining fluctuations will have to be carefully corrected to be able to measure the power spectrum on large scales without bias. This is one of the main topic of the next chapter.

Note that the imaging systematic plots in Fig. 4.43-4.44 are plotted between -0.1 and 0.1. Wile the previous figures, Fig. 4.28-4.30, 4.39, are plotted between -0.2 and 0.2.

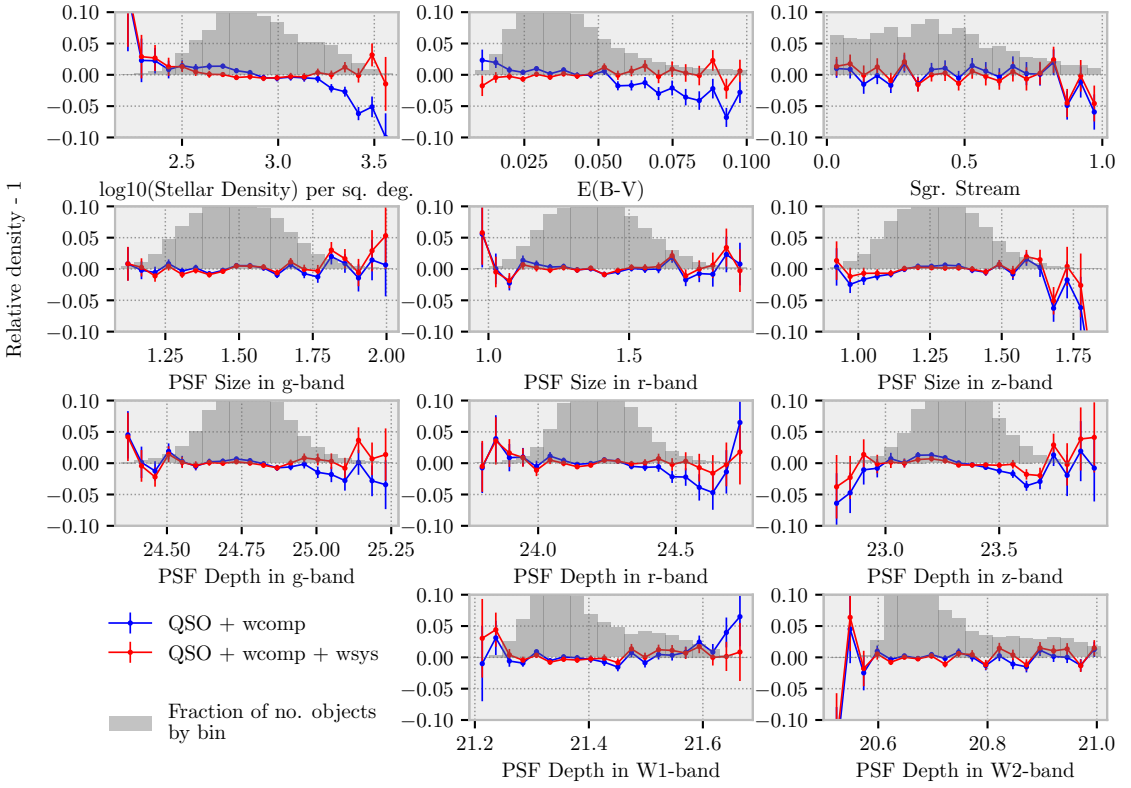
The imaging systematics displayed here are weaker than in the target case confirming the prediction done in Fig. 4.39. Indeed, thanks to the purity of the catalog, all the stellar contamination that occurs during the QSO target selection is removed after the spectroscopic classification. The large trends observed on the stellar density and Sagittarius stream features in Fig. 4.28 are then removed. The dependence in E(B-V) is also removed, since E(B-V) is strongly contaminated with the stellar density and the trend was not fully due to the E(B-V) feature.

This can be also observed in the sky distribution shown in Fig. 4.41, where no excess of

objects can be identified near the galactic plane or in the Sagittarius stream.

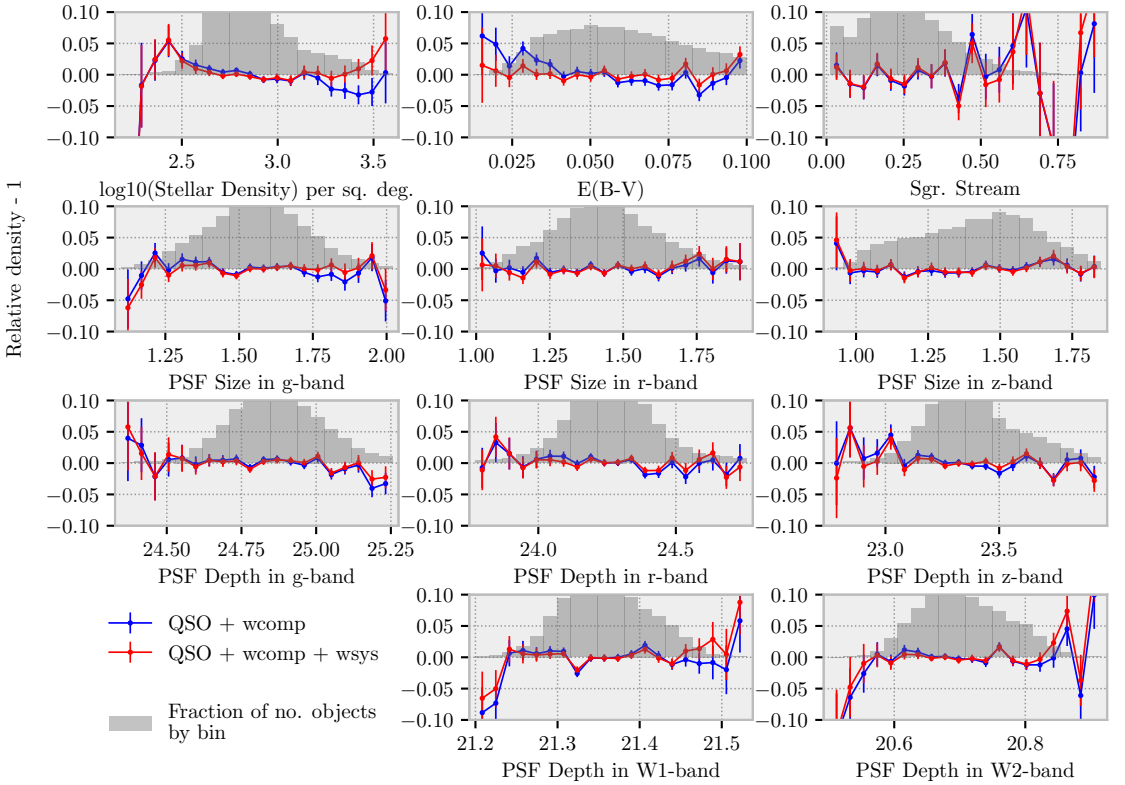


(a) North

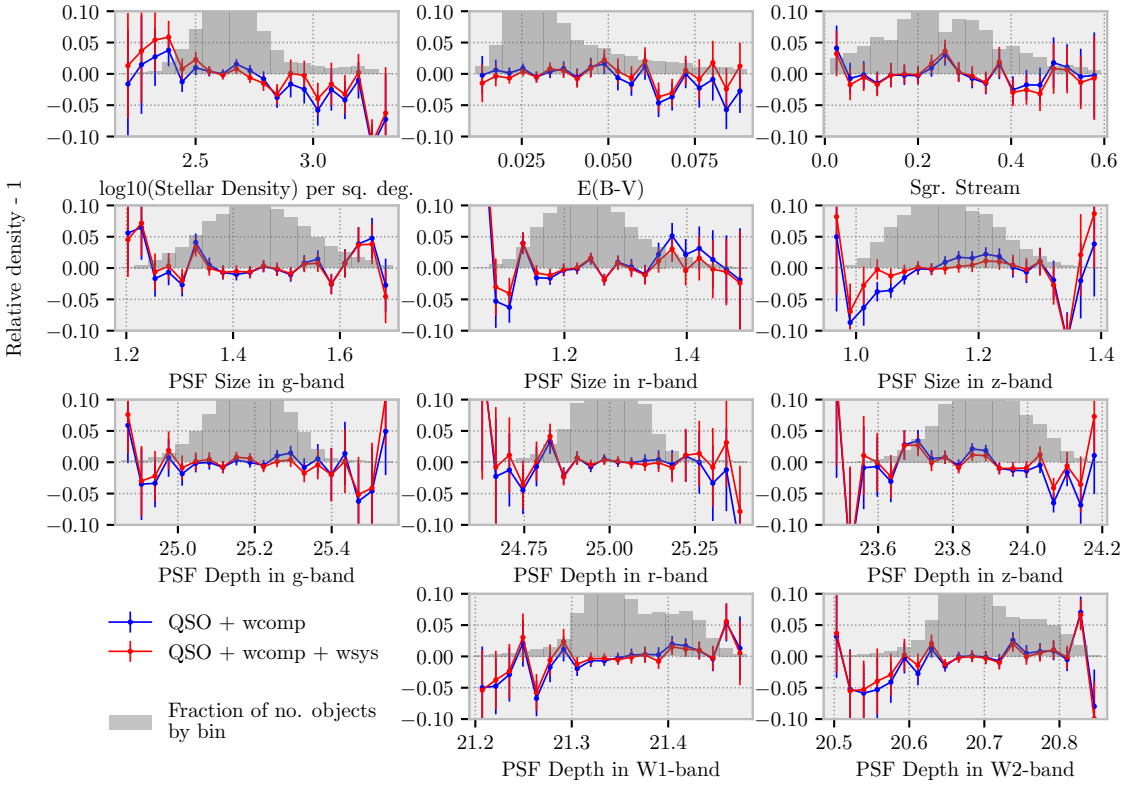


(b) South NGC

Figure 4.43: Same as Fig. 4.28, but here the blue lines are for the catalog from the spectroscopic data with completeness weights and the red ones are for the catalog from spectroscopic data with completeness weights and with imaging weights computed with `regressis`. (a) is for North and (b) for South NGC footprint.



(a) South SGC



(b) DES

Figure 4.44: Same as Fig. 4.43. (a) is for South SGC and (b) for DES footprint.

Chapter 5

Measurement of primordial non-gaussianity with DESI quasars

Contents

Deep in the human unconscious is a pervasive need for a logical universe that makes sense. But the real universe is always one step beyond logic.

Frank Herbert, Dune (1965)

Now that DESI is producing the largest 3D map of the Universe ever made, it is time to take the next step, by analyzing all this new data. DESI will have several data release (DR) during all its survey. In the following, we will work with DR1, the data release of the first year of observation, see Section 3.2.2.3.

This first data release (DR1) is particularly appealing for the quasars sample since they have the highest priority during the observation and then are collected first. So we are already expecting competitive constraints on the measurement of primordial non-gaussianity.

After explaining how to measure the power spectrum, and therefore primordial non-gaussianity, from a galaxy survey, we will validate the weights correcting for imaging systematics using simulations with non-zero $f_{\rm NL}^{\rm loc}$ values. Then, we will propose a method for blinding the signal produced by the scale-dependent relation, with the aim of optimizing the measurement without any confirmation bias. Finally, we will conclude with a preliminary analysis using the DR1 quasars.

Section 5.3 is from Chaussidon et al. (2023) and all this analysis should be published with the release of DESI Y1 data expected in late 2024.

5.1 Measuring primordial non-gaussianity

5.1.1 Measuring the power spectrum from a spectroscopic survey

5.1.1.1 Power spectrum estimator

▶ **FKP estimator** Under the assumption that fluctuations are Gaussian, Feldman et al. (1994) derived the optimal estimator for the power spectrum:

$$\hat{P}\left(\mathbf{k}_\mu\right) = \frac{1}{AV_{\mathbf{k}_\mu}} \int_{V_{\mathbf{k}_\mu}} \mathrm{d}\mathbf{k} \left[\int \mathrm{d}\mathbf{x_1} \int \mathrm{d}\mathbf{x_2} e^{i\mathbf{k}\cdot(\mathbf{x_2}-\mathbf{x_1})} w_{\rm FKP}\left(\mathbf{x_1},\mathbf{k}_\mu\right) w_{\rm FKP}\left(\mathbf{x_2},\mathbf{k}_\mu\right) \mathcal{F}\left(\mathbf{x_1}\right) \mathcal{F}\left(\mathbf{x_2}\right) - \int \mathrm{d}\mathbf{x} \bar{n}(\mathbf{x}) w_{\rm FKP}\left(\mathbf{x},\mathbf{k}_\mu\right)^2 \right], \tag{5.1}$$

where $\hat{P}\left(\mathbf{k}_\mu\right)$ is the estimation of the average power spectrum in a volume $V_{\mathbf{k}_\mu}$ corresponding to the binning used in $\mathbf{k}_\mu$ space for the measurement. The last term removes the shot noise. The normalization factor is given by

$$A = \int \mathrm{d}\mathbf{x} \bar{n}(\mathbf{x})^2 w_{\rm FKP}\left(\mathbf{x},\mathbf{k}_\mu\right)^2, \tag{5.2}$$

and $\mathcal{F}$ is the fluctuation field, known also as the FKP field,

$$\mathcal{F}(\mathbf{x}) = n(\mathbf{x}) - \bar{n}(\mathbf{x}) = \bar{n}(\mathbf{x})\left[\delta(\mathbf{x}) - 1\right]. \tag{5.3}$$

$\bar{n}(\mathbf{x})$ is the expected number density of galaxies in the absence of clustering, and it is computed via a synthetic catalog, knowns as *randoms*, that reproduce geometrically the exact observed data distribution[1]. The fluctuation field becomes

$$\mathcal{F}(\mathbf{x}) = n(\mathbf{x}) - \alpha n_s(\mathbf{x}), \tag{5.4}$$

with

$$\alpha = \frac{\int \mathrm{d}\mathbf{x} W(\mathbf{x})}{\int d^3\mathbf{x} n_r(\mathbf{x})}, \tag{5.5}$$

where we have introduced, the *survey selection function*: $W(\mathbf{x}) = \langle n(\mathbf{x})\rangle$.

Remark: As shown in Feldman et al. (1994), the average of the FKP fields is given by

$$\left\langle \mathcal{F}(\mathbf{x})\mathcal{F}\left(\mathbf{x}'\right)\right\rangle = \bar{n}(\mathbf{x})\bar{n}\left(\mathbf{x}'\right)\xi\left(\mathbf{x},\mathbf{x}'\right) + \bar{n}(\mathbf{x})\delta_D^{(3)}\left(\mathbf{x}-\mathbf{x}'\right), \tag{5.6}$$

that explains the normalization factor A and the subtraction of the shot noise contribution in (5.1), to obtain an unbiased estimator, such that

$$\left\langle \hat{P}(\mathbf{k}_\mu)\right\rangle = \frac{1}{AV_{\mathbf{k}_\mu}}\int_{V_{\mathbf{k}_\mu}} \mathrm{d}\mathbf{k}\int \mathrm{d}\mathbf{x_1}\int \mathrm{d}\mathbf{x_2} e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}\bar{n}(\mathbf{x_1})w_{\mathrm{FKP}}\left(\mathbf{x_1},\mathbf{k}_\mu\right)\bar{n}(\mathbf{x_2})w_{\mathrm{FKP}}\left(\mathbf{x_2},\mathbf{k}_\mu\right)\xi(\mathbf{x_1},\mathbf{x_2}). \tag{5.7}$$

The weights w_{FKP}, known as the FKP weights[2], are the optimal weighting scheme for the power spectrum measurement and minimize the expected errors of $\hat{P}\left(\mathbf{k}_\mu\right)$:

$$w_{\mathrm{FKP}}(\mathbf{x},\mathbf{k}) = \frac{1}{1+\bar{n}(\mathbf{x})P(\mathbf{k})}. \tag{5.8}$$

They are the optimal compromise between the density of the tracer (the greater the number of objects, the smaller the errors) and the value of the power spectrum (the larger the power spectrum, the easier it is to measure). They depend on $\mathbf{k}$ and on $P(\mathbf{k})$ which is unknown a priori. We can easily iterate on the value of P(k), although this is not necessarily very decisive, and it is common to take the value of the fiducial power spectrum as a reference. On the other hand, for computational reasons, the FKP weights are fixed at a specific value of $\mathbf{k}$ to be independent of it. They are set at to the scale of interest. In our case, we want to probe the scale-dependent bias with the power spectrum (2.105) around k_{eq} *i.e.* where the amplitude is maximal. Hence, we will use the maximal amplitude measured in the data: $P(\mathbf{k}) = P_0 = 3\times 10^4\ h^{-3}\mathrm{Mpc}^3$, see Fig. 5.35.

► **Multipoles estimator** Few years later, Yamamoto et al. (2006) proposed similar estimators for the multipoles of the power spectrum (1.119):

$$\hat{P}_\ell\left(k_\mu\right) = \frac{2\ell+1}{AV_{k_\mu}}\int_{V_{k_\mu}} \mathrm{d}\mathbf{k}\int \mathrm{d}\mathbf{x}_1\int \mathrm{d}\mathbf{x}_2 e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)} w_{\mathrm{FKP}}\left(\mathbf{x}_1\right) w_{\mathrm{FKP}}\left(\mathbf{x}_2\right)\mathcal{F}\left(\mathbf{x}_1\right)\mathcal{F}\left(\mathbf{x}_2\right)\mathcal{L}_\ell\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}_m\right), \\ - P_\ell^{\mathrm{noise}}\left(k_\mu\right) \tag{5.9}$$

[1] In the following, we will assume that the density of randoms in the synthetic catalog is sufficiently high to avoid any bias in the determination of $\bar{n}(\mathbf{x})$. In general, a density 10 times that of the data is sufficient.

[2] Feldman et al. (1994) were written by Feldman, Kaiser and Peacock.

using the *midpoint* line-of-sight $\mathbf{x}_m = (\mathbf{x}_1 + \mathbf{x}_2)/2$. This is only well defined under the so-called *local plane-parallel* approximation in which the two galaxies are far from the observer such that $|\mathbf{x}_2 - \mathbf{x}_1| / |\mathbf{x}_m| \ll 1$. The shot noise contribution is

$$\begin{aligned} P_\ell^{\text{noise}}(k_\mu) &= \frac{2\ell+1}{AV_{k_\mu}} \int_{V_{k_\mu}} \mathrm{d}\mathbf{k} \int \mathrm{d}\mathbf{x} \bar{n}(\mathbf{x}) w_{\text{FKP}}(\mathbf{x})^2 \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}) \\ &= \frac{\delta_{\ell 0}}{A} \int \mathrm{d}\mathbf{x} \bar{n}(\mathbf{x}) w_{\mathbf{FKP}}(\mathbf{x})^2 . \end{aligned} \tag{5.10}$$

To reduce the computation time of (5.9), one can split the double integral by choosing one galaxy in each pair as the line-of-sight, as noticed by Yamamoto et al. (2006),

$$\hat{P}_\ell(k_\mu) = \frac{2\ell+1}{AV_{k_\mu}} \int_{V_{k_\mu}} \mathrm{d}\mathbf{k} F_0(\mathbf{k}) F_\ell(-\mathbf{k}) - P_\ell^{\text{noise}}(k_\mu) , \tag{5.11}$$

where we have introduced

$$F_\ell(\mathbf{k}) = \int \mathrm{d}\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} w_{\text{FKP}}(\mathbf{x}) \mathcal{F}(\mathbf{x}) \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}). \tag{5.12}$$

As noticed by Hand et al. (2017), (5.12) can be written as a sum of Fourier transforms. Indeed, by decomposing the Legendre polynomials $\mathcal{L}_\ell$ into *spherical harmonics* $Y_{\ell m}$:

$$\mathcal{L}_\ell(\hat{\mathbf{x}} \cdot \hat{\mathbf{k}}) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{m=\ell} Y_{\ell m}(\hat{\mathbf{x}}) Y^\star_{\ell m}(\hat{\mathbf{k}}), \tag{5.13}$$

(5.12) becomes

$$F_\ell(\mathbf{k}) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{m=\ell} Y^\star_{\ell m}(\hat{\mathbf{k}}) \int d^3x e^{-i\mathbf{k}\cdot\mathbf{x}} w_{\text{FKP}}(\mathbf{x}) F(\mathbf{x}) Y_{\ell m}(\hat{\mathbf{x}}), \tag{5.14}$$

and requires the computation of only $2\ell + 1$ Fast Fourier Transforms for each multipole ℓ.

► **Aliasing and interlacing** To perform the computation of (5.12), the FKP field $\mathcal{F}$ is interpolated on mesh creating *aliasing* effects. In general, the interpolation scheme is chosen to be the $(p-1)$-th order convolution of the top hat function with itself:

- p=1, nearest-grid-point (NGR):

$$W^{(1)}(s) = \begin{cases} 1 & \text{si } |s| < \frac{T_c}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

- p=2, cloud-in-cell (CIC):

$$W^{(2)}(s) = \begin{cases} 1 - |s| & \text{si } |s| < T_c \\ 0 & \text{otherwise} \end{cases} \tag{5.16}$$

- p=3, triangular shaped cloud (TSC):

$$W^{(3)}(s) = \begin{cases} \frac{3}{4} - s^2 & \text{si } |s| < \frac{T_c}{2} \\ \frac{1}{2}\left(\frac{3}{2} - s^2\right)^2 & \text{si } \frac{T_c}{2} < |s| < \frac{3T_c}{2} \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

where s is the distance between the object and the node of the mesh, and p is the order of the interpolation. This order corresponds to the number of nodes in which the object is distributed.

The aliasing effects can be corrected by the *interlacing* method, as proposed by Sefusatti et al. (2016). The interlacing at order n consists of painting the density field shifted by $[0, 1, \ldots, n-1]$ mesh cell size in x, y and z directions and then averaging the FFT multiplied by the appropriate phase terms. The resulting Fourier transform must still be divided by the Fourier transform of the interpolation scheme $W^p(k)$.

With the TSC interpolation scheme and the interlacing method, the power spectrum is measured with relative errors of $\sim 10^{-4}$ up to the Nyquist frequency defined as

$$k_{\rm nyq} = \pi \frac{N_{\rm mesh}}{L_{\rm box}}, \tag{5.18}$$

where $L_{\rm box}$ is the size of the grid and $N_{\rm mesh}$ is the number of nodes along one direction in which we perform the Fourier transform.

Remark: In the following, all the power spectrum will be computed with `pypower`[3] using the TSC sampling and interlacing at order $n = 3$.

5.1.1.2 Optimal quadratic estimator for the scale-dependent bias

The above section gives, with the FKP weights, the optimal estimator for measuring the power spectrum. However, it does not give the optimal way to extract the scale-dependent bias signal in the power spectrum. This can be achieved by using an optimal quadratic estimator (OQE) as in Mueller et al. (2019), Castorina et al. (2019), Mueller et al. (2022). In the following, we will follow Castorina et al. (2019) that propose to weight each galaxy instead of weighting pairs of galaxies as in Mueller et al. (2022), which is more natural to compute FKP field $\mathcal{F}$.

The optimal estimator for extracting $f_{\rm NL}^{\rm loc}$ has the same form as (5.11) but with a different weighting scheme:

$$\hat{P}_\ell\left(k_\mu\right) = \frac{2\ell+1}{A_\ell V_{k_\mu}} \int_{V_{k\mu}} {\rm d}\mathbf{k} \tilde{F}(\mathbf{k}) F_\ell(-\mathbf{k}) - P_\ell^{\rm noise}\left(k_\mu\right), \tag{5.19}$$

where

$$\begin{aligned} \tilde{F}(\mathbf{k}) &= \int {\rm d}\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{w}\left(\mathbf{x}\right) \mathcal{F}(\mathbf{x}) \\ F_\ell(\mathbf{k}) &= \int {\rm d}\mathbf{x} e^{i\mathbf{k}\cdot\mathbf{x}} w_\ell\left(\mathbf{x}\right) \mathcal{F}(\mathbf{x}) \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) \end{aligned}, \tag{5.20}$$

and the shot noise contribution is

$$P_\ell^{\rm noise}\left(k_\mu\right) = \frac{\delta_{\ell 0}}{A} \int {\rm d}\mathbf{x} \bar{n}(\mathbf{x}) \tilde{w}\left(\mathbf{x}\right) w_\ell\left(\mathbf{x}\right).$$

Similarly, the normalization factor (5.2) becomes

$$A_\ell = \int {\rm d}\mathbf{x} \bar{n}(\mathbf{x})^2 \tilde{w}\left(\mathbf{x}\right) w_\ell\left(\mathbf{x}\right).$$

The optimal weights[4] $\tilde{w}$, w_0 and w_2 for the quadratic estimator are

$$\begin{aligned} \tilde{w}(z) &= w_{\rm FKP}(z)\left[b(z) - p\right] \\ w_0(z) &= w_{\rm FKP}(z) \times D(z)\left[b(z) + f(z)/3\right], \\ w_2(z) &= w_{\rm FKP}(z) \times 2/3 D(z) f(z) \end{aligned} \tag{5.21}$$

[3] https://github.com/cosmodesi/pypower

[4] Here, we assume that the density distribution is isotropic and only depends on the redshift.

where p is the parametrization used in b_ϕ (2.120), f is the growth rate (1.108) and D is the growth factor (1.105).

These weights are displayed in Fig. 5.1 for the main QSO sample of DESI, where we supposed that the QSO linear bias follows the relation found in Laurent et al. (2017). The weights will overweight the objects at high redshift where the bias is higher, and so, where the measurement of $f_{\rm NL}^{\rm loc}$ will be easier since at first order in $f_{\rm NL}^{\rm loc}$, the scale-dependent bias (2.105) reads as $P(k,z) = \left(b(z)^2 + b(z) \times b_\phi(z) f_{\rm NL}^{\rm loc}/\alpha(k,z)\right) P_{lin}(k,z)$. Besides, b_ϕ increases also with the redshift in the case of the universal mass relation.

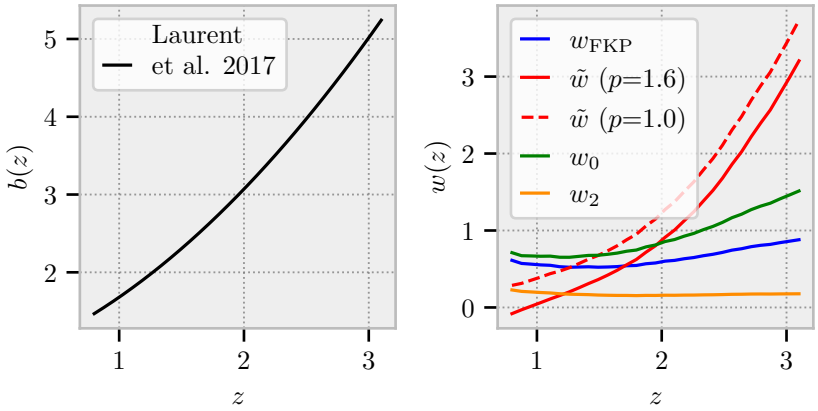


Figure 5.1: Left: Linear bias of QSO measured by Laurent et al. (2017) with the SDSS DR14 quasars. Right: Weights for the optimal quadradic estimator. We use $P_0 = 3e4$ for the FKP weights.

Practically speaking, this is just the cross-power spectrum between one FKP field weighted by $\tilde{w}$ and another one weighted by w_ℓ. Therefore, the computation of the power spectrum can be carried out without any hassle!

Remark: Since the bias vanishes in the hexadecapole ($\ell = 4$), no specific weights are needed for this multipole. Including the hexadecapole can only improve the statistical error on $f_{\rm NL}^{\rm loc}$ by improving the measurement of the other fitting parameters. In the case of QSOs, the large-scale hexadecapole will be far too noisy to be used and will therefore be neglected in the rest of the analysis.

5.1.1.3 Correcting observational effects

The data can be weighted, and so the FKP field, in order to correct some observational effects[5].

▶ **Fiber assignement** One big concern with a multi-fiber spectroscopic survey is the impact of the choice of the targets during the observation known as the *fiber assignement* (F.A.). Fig. 5.2 illustrates this process with one and two passes in the same regions. Each gray patch is one tile that can observe about 5000 objects simultaneously. The number of available targets is higher than the number of fibers for each exposure, and several passes are needed to observe all the targets.

[5]Since the randoms should reproduce exactly the same distribution as the data, they should also be weighted in the same manner. To avoid any problems, one can simply draw randomly the redshifts and associated weights from the data for the randoms (Ross et al. 2020)

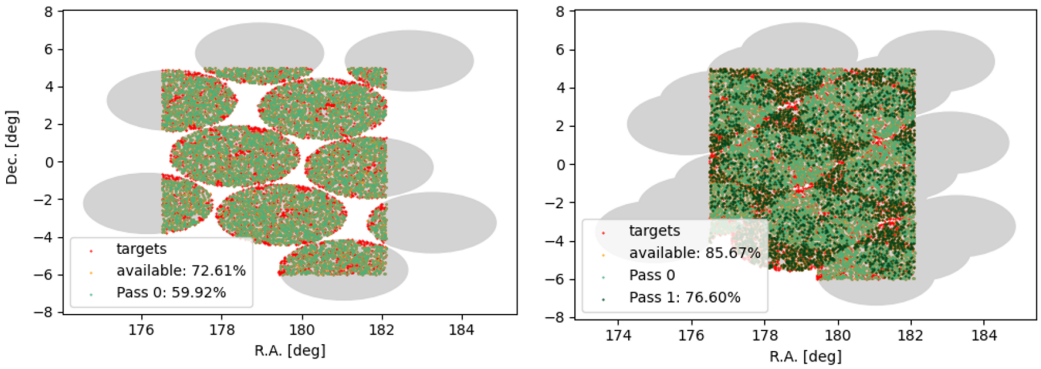


Figure 5.2: Fiber assignment with one pass (left) and two passes (right). Red points are the targets that are inside the geometrical region observed by the considered tiles. Yellow points are the available targets *i.e.* targets that could be observed by one fiber. Due to the particular shape of the focal plane, or due to broken fibers, some areas do not have any fiber, and so all the targets cannot be assigned. Green points are the targets that are observed. Since they are too numerous, they cannot be observed in a single pass. DESI is expected to have at least five passes over a given region to observe all the targets.

The complex geometry of the survey will be taken into account via the randoms and does not represent any difficulties. On the other hand, targets unobserved because they are no free fibers to observe them, will lead to bad estimation of the FKP field. These missing targets cannot be taken into account by the randoms since the associated geometrical regions, where these targets are, are observed by the survey. This will impact both large and small scales. Fig. 5.3 shows the impact at large scales for two different regions of the Y1 data: SGC (poorly complete) and the restricted 3-pass area in the NGC (almost complete), see Fig. 4.40. The green lines are not corrected of this effect and present an excess of power on large scales. The impact at small scales is a tiny effect that is not visible in this log-log plot but can be visualized by plotting $kP(k)$ as a function of k.

The large scales can be easily corrected by *completeness* weights $w_{\rm comp}$. Fig. 5.4 illustrates how we can weight an object according to nearby unobserved targets. The completeness weights are then the number of unobserved targets plus one in the control radius of a fiber[6]. Each unobserved object can only be counted once. The number of unobserved targets is decreased with the number of passes, so the completeness weights will become almost one everywhere once the survey is done. Red lines in Fig. 5.3 are the power spectrum computed with the completeness weights and match perfectly the power spectrum without fiber assignment at large scales.

The impact at small scales can be corrected with more sophistical methods (see, for instance, Bianchi et al. 2018, Mohammad et al. 2020) but will be neglected in the following since the completeness weights correct exactly the scales of interest used for the scale-dependent bias measurement.

► **Imaging weights** Imaging systematics occur during the target selection and will create, unfortunately, an excess of correlation at large scales. These effects will be corrected by a set of weights $w_{\rm sys}$. One of the main results of this dissertation is the development of these weights for DESI, see Section 4.3. This also explains the special attention I paid to the quasar selection in DESI, see Section 4.2, to reduce as much as possible the imaging systematics in the spectroscopic sample, see Section 4.4.2.

[6]In DESI, this number is 1 /FRACZ_TILELOCID

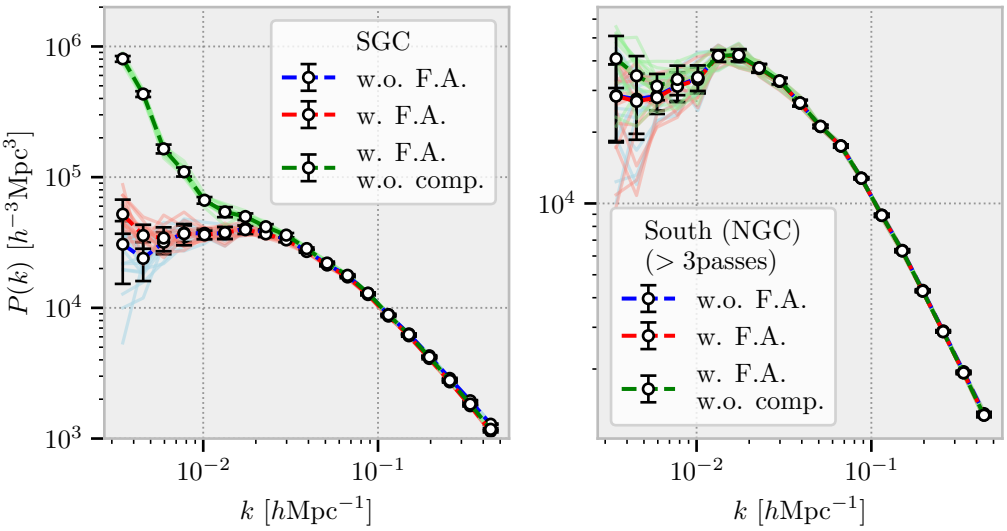


Figure 5.3: Mean of 16 sub-realizations of my fastpm simulation with $f_{\rm NL}^{\rm loc} = 0$ (Fastpm simulation will be described in Section 5.2.1) for SGC (left) and the 3-pass area of the South NGC (right). Errors are the standard deviation of these 15 sub-realizations. Blue lines are the power spectrum measured from the simulations without the fiber assignment applied. Green ones are from the simulations with fiber assignment applied, and red ones are after the correction with completeness weights.

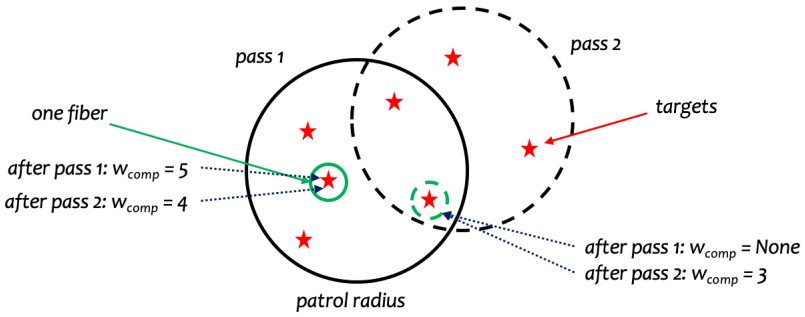


Figure 5.4: Illustration of the completeness weights. Red stars are targets in a nearby region. Black circles are the control radius of a fiber *i.e.* the area reachable by the fiber. In the first (*resp.* second) pass, five (*resp.* four) targets are reachable by the fiber. The completeness weights are computed to be egal of the number of unobserved targets plus one in the control radius. Note that an object can only be counted once.

The major concern about these weights is the possibility to remove too much, or not enough, power at large scales in the power spectrum, thus biasing the measurement of $f_{\rm NL}^{\rm loc}$. Hence, Section 5.2 will study this impact carefully.

► **Redshift efficiency weights** Classification and redshift determination will depend on the quality of the observation. Poor weather, noise in the CDDs or dust in the sky can indeed have an impact on the spectra collected. At this stage of DESI, no impact has been shown on a large scale, and this effect will not be corrected in the following. Although it will naturally be taken into account in our theoretical model.

▶ **Total weights** Finally, the final weights used for each galaxy are

$$w_{\rm tot} = w_{\rm comp} \times w_{\rm wsys} \times w_{\rm FKP/OQE} \tag{5.22}$$

5.1.2 Theory versus Data

In DESI, we will use the Planck18 cosmology (Planck Collaboration et al. 2020), given in Table 1.2, as the fiducial cosmology to transform redshifts in distances.

5.1.2.1 Power spectrum model

Multipoles of the power spectrum measured from the data will be compared to

$$P_{{\rm theo},\ell}(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu P_{\rm theo}(k,\mu)\mathcal{L}_\ell(\mu), \tag{5.23}$$

with P_{theo} our model. For the measurement of the primordial non-gaussianity, we will use the linear theory with the Kaiser term (1.127), a damping factor (1.126) and a scale-dependent bias (2.105):

$$P_{\rm theo}(k,\mu) = \frac{\left[b + \dfrac{b_\Phi}{\alpha(k,z_{\rm eff})} f_{\rm NL}^{\rm loc} + f\mu^2\right]^2}{\left[1+\frac{1}{2}\left(k\mu\Sigma_s\right)^2\right]^2} \times P_{lin}(k,z_{\rm eff}) + s_{n,0} \tag{5.24}$$

where P_{lin} is the linear power spectrum computed with `CLASS` at the effective reshift of the data, $s_{n,0}$ accounts for the residual shot noise and Σ_s represents the typical damping velocity dispersion. Although initially introduced to explain the Finger-of-God effect, the damping term will also naturally account for the redshift determination errors. For instance, in DESI, these errors are about 150 kms^{-1} for quasars with $z < 4.0$ (Brodzeller et al. 2023) or see Fig. 4.20.

Remark: During all the parameter estimation, we will evaluate the linear power spectrum at the fiducial cosmology, keeping the shape of the linear power spectrum fixed. This means that we will neglect the uncertainty on the shape of the power spectrum given by Planck Collaboration et al. (2020) compared to the uncertainty of our measurement, by not propagating uncertainties about Ω_m.

5.1.2.2 Effective Redshift

The model (5.24) requires the evaluation of P_{lin} at the *effective redshift* of the data. Under the local plane parallel approximation, with infinite dΩ_k, and assuming the survey selection function varies slowly with respect to the correlation function, the Yamamoto estimator (5.9) gives, in average (similarly to (5.7)),

$$\left\langle \hat{P}_\ell(k) \right\rangle = \frac{\int {\rm d}\mathbf{r}\bar{n}^2(\mathbf{r}) P_\ell(k, z(\mathbf{r}))}{\int {\rm d}\mathbf{r}\bar{n}^2(\mathbf{r})}, \tag{5.25}$$

where $\bar{n}$ is the expected mean density of the data (weighted) without clustering. $P_\ell(k, z(\mathbf{r}))$ is the true power spectrum of the data evaluated at the redshift z. The Taylor expansion of the power spectrum around the effective redshift $z_{\rm eff}$ gives

$$P_\ell(k,z) = P_\ell\left(k, z_{\rm eff}\right) + P'_\ell\left(k, z_{\rm eff}\right)\left(z - z_{\rm eff}\right) + \cdots \tag{5.26}$$

Finally, injecting (5.26) into (5.25) leads, at first order in the Taylor expansion, to

$$\tilde{P}_\ell(k) = \frac{\int \mathrm{d}\mathbf{r}\bar{n}^2(\mathbf{r})P_\ell\left(k, z_{\mathrm{eff}}\right)}{\int \mathrm{d}\mathbf{r}\bar{n}^2(\mathbf{r})}, \tag{5.27}$$

using the effective redshift defined by

$$z_{\mathrm{eff}} = \frac{\int \mathrm{d}\mathbf{r}\bar{n}^2(\mathbf{r})z(\mathbf{r})}{\int \mathrm{d}\mathbf{r}\bar{n}^2(\mathbf{r})}. \tag{5.28}$$

This definition can be easily extended in the case of two different sets of weights, as required for the OQE weights, by

$$z_{\mathrm{eff}} = \frac{\int dzn(z)^2 w_1(z)w_2(z)z}{\int dzn(z)^2 w_1(z)w_2(z)}. \tag{5.29}$$

Table 5.1 gives the effective redshift under the different sets of weights that will be used in the following. The redshift distribution of the DESI quasars is given in Fig. 4.42. Note the significant increase in effective redshift when including quasars with high redshift ($z > 2.1$), especially when using OQE weights.

In the following, for the OQE weights, we will only generate the model at the effective redshift given by the monopole weights. This simplified calculation will be removed in the near future.

Table 5.1: Effective redshift for the DESI quasars with different redshift ranges. For comparison, the mean redshift of the sample is displayed in the first line.

Redshift range	0.8 – 2.1	0.8 – 3.1	0.8 – 3.5
$\bar{z}$	1.491	1.768	1.803
z_{eff}	1.435	1.570	1.575
z_{eff} (FKP)	1.441	1.649	1.663
z_{eff} (OQE $\ell = 0$, $p = 1.0$)	1.597	1.964	2.005
z_{eff} (OQE $\ell = 2$, $p = 1.0$)	1.546	1.837	1.862
z_{eff} (OQE $\ell = 0$, $p = 1.6$)	1.715	2.114	2.162
z_{eff} (OQE $\ell = 2$, $p = 1.6$)	1.684	2.015	2.048

Remark: Many other definitions of effective redshift exist, see, for instance, De Mattia et al. (2021).

5.1.2.3 Geometrical effects

Some effects cannot be corrected by weighting the data and we need to change the theory to match correctly the observation.

► **Window function** Due to the stars, Milky Way dust, or simply the focal plane, some parts of the footprint are masked or unobserved. Hence, we do not exactly observe the full density field $\delta(\mathbf{x})$, but only a fraction of it given by $W(\mathbf{x})\delta(\mathbf{x})$ where W is the survey function, as defined in Section 5.1.1.1, $W(\mathbf{x}) = \langle n_d(\mathbf{x})\rangle$. Hence, the resulting multipole of the power spectrum estimator reads as

$$\left\langle \hat{P}_\ell(k)\right\rangle = \frac{(2\ell+1)}{A}\int\frac{\mathrm{d}\Omega_k}{4\pi}\int \mathrm{d}\mathbf{s_1}\int \mathrm{d}\mathbf{s_2}e^{i\mathbf{k}(\mathbf{s_2}-\mathbf{s_1})}\xi(\mathbf{s_1},\mathbf{s_2})W\left(\mathbf{s_1}\right)W\left(\mathbf{s_2}\right)\mathcal{L}_\ell\left(\hat{\mathbf{k}}\cdot\hat{\mathbf{s_1}}\right) \tag{5.30}$$

Following Beutler et al. (2019), the correlation function can be also expanded into Legendre multipoles. Let $\mathbf{s} = \mathbf{s_1} - \mathbf{s_2}$, and under the local plane-parallel approximation limit ($\mathbf{s} \ll \mathbf{s_1}, \mathbf{s_2}$),

$$\xi\left(\mathbf{s}_1, \mathbf{s}_2\right) = \sum_p \xi_p(s) \mathcal{L}_p\left(\hat{\mathbf{s_1}} \cdot \hat{\mathbf{s}}\right). \tag{5.31}$$

The estimator becomes

$$\begin{aligned}\left\langle \hat{P}_\ell(k) \right\rangle =& \frac{(2\ell+1)}{A} \sum_p \int \frac{\mathrm{d}\Omega_k}{4\pi} \int \mathrm{d}\mathbf{x} \int \mathrm{d}\mathbf{s} e^{-i\mathbf{k}\cdot\mathbf{s}} W(\mathbf{x}) W(\mathbf{x}-\mathbf{s}) \xi_p(s) \mathcal{L}_p(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}) \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}). \\ =& 4\pi(-i)^\ell(2\ell+1) \sum_{\ell_1,\ell_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 0 & 0 & 0 \end{pmatrix}^2 \int \mathrm{d}s s^2 j_\ell(ks) \xi_{\ell_1}(s) \mathcal{W}_{\ell_2}(s)\end{aligned} \tag{5.32}$$

where we have successively used the usual tools for the multipole manipulation[7], and we have introduced the real space *window matrix*

$$\mathcal{W}_\ell(s) \equiv \frac{(2\ell+1)}{4\pi \times A} \int \mathrm{d}\Omega_s \int \mathrm{d}\mathbf{x} W(\mathbf{x}) W(\mathbf{x}-\mathbf{s}) \mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}). \tag{5.33}$$

The multipoles of the correlation function can be computed from the theoretical power spectrum (5.24) using (1.121):

$$\xi_\ell(r) = \frac{i^\ell}{2\pi^2} \int k^2 dk j_\ell(kr) P_\ell(k). \tag{5.34}$$

In practice, the window matrix is computed from the randoms that fully described the geometry of the data. In the following, we will use the implementation of the window matrix computation from `pypower`[8] and will convolve the theoretical model with the window function thanks to `desilike`[9]. As in Beutler et al. (2017), we will only use the multipoles up to $\ell = 4$ in (5.32) such that we will only consider the window matrix up to $\ell = 8$[10].

The impact of the window function for the DESI QSO Y1 is shown in Fig. 5.5. The impact on large scales of the window matrix cannot be neglected when constraining the primordial non-gaussianity.

Remark: As indicated by De Mattia and Ruhlmann-Kleider (2019), the real space window matrix $\mathcal{W}_0$ is not normalized to 1 when $s \to 0$ since small scales are not completely well sampled by the randoms. The normalization factor is the same as the one used in Yamamoto's estimator (5.11) and does not introduce a bias during the parameter estimation.

[7] The Rayleigh expansion of a plane wave is

$$e^{-i\mathbf{k}\cdot\mathbf{s}} = \sum_{q=0}^{+\infty} (-i)^q (2q+1) j_q(ks) \mathcal{L}_q(\hat{\mathbf{k}}\cdot\hat{\mathbf{s}}),$$

orthogonality of Legendre polynomials gives

$$\int \frac{d\Omega_k}{4\pi} \mathcal{L}_\ell(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) \mathcal{L}_q(\hat{\mathbf{k}}\cdot\hat{\mathbf{s}}) = \frac{\delta_{\ell q}}{2\ell+1} \mathcal{L}_\ell(\hat{\mathbf{x}}\cdot\hat{\mathbf{s}}),$$

and finally the Wigner 3-j symbols (alternative to the Clebsch–Gordan coefficients) are defined by

$$\mathcal{L}_{\ell_1}(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) \mathcal{L}_{\ell_2}(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}) = \sum_{\ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2 (2\ell_3+1) \mathcal{L}_{\ell_3}(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}}).$$

[8] https://github.com/cosmodesi/pypower/blob/main/nb/window_examples.ipynb

[9] https://github.com/cosmodesi/desilike

[10] Non-zero Wigner 3-j symbols must respect: $|l_1 - l_2| < l \leq l_1 + l_2$

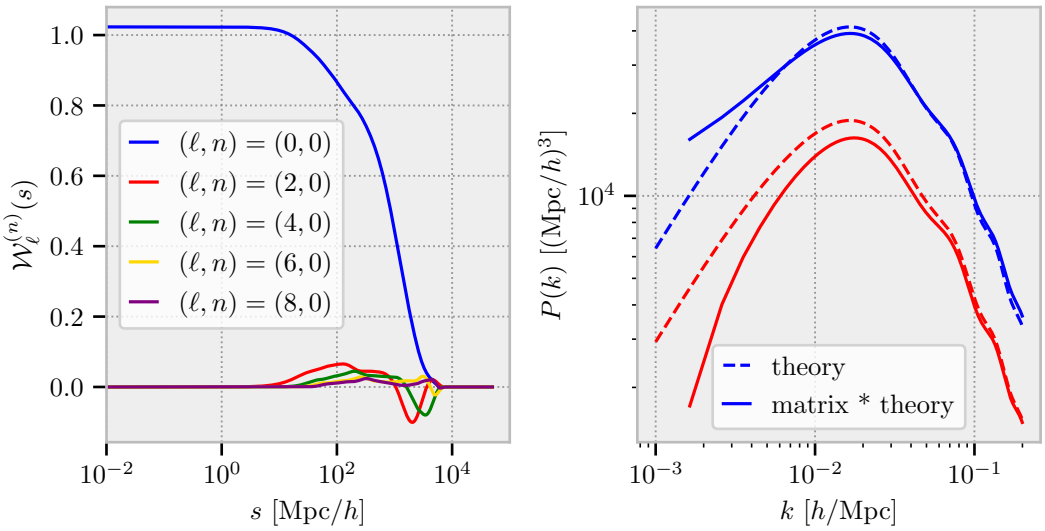


Figure 5.5: Left: Real space window matrix up to $\ell = 8$. Right: Monopole (blue) and quadrupole (ref) for a simple Kaiser model with expected parameters for the QSO Y1 (dashed lines) and the associated convolved model (full lines) with the window matrix computed for the geometry of DESI QSO Y1.

▶ **Wide-angle effect** In order to speed up the computation, we chose one galaxy as the line-of-sight in the Yamamoto estimator (5.25). This is a good choice under the local plane-parallel approximation ($\mathbf{s} \ll \mathbf{x_1}, \mathbf{x_2}$). As shown in Castorina and White (2018), this choice creates *wide-angle effect* compare to the measurement using the midpoint line-of-sight when the approximation of the local plane-parallel does not hold. Note that this choice will also create non-vanishing odd multipoles.

Following Beutler et al. (2019), the wide-angle effect can be taken into account by expanding the theoretical correlation function such as

$$\xi\left(\mathbf{x}_1, \mathbf{x}_2\right) = \sum_{p,n} \left(\frac{s}{d}\right)^n \xi_p^{(n)}(s) \mathcal{L}_p(\hat{\mathbf{d}} \cdot \hat{\mathbf{s}}), \tag{5.35}$$

where $\mathbf{s} = \mathbf{x_2} - \mathbf{x_1}$ is the pair separation and $\mathbf{d}$ a generic line-of-sight (not normalized). This expression, under the local plane-parallel approximation ($s/d \to 0$), gives the usual Legendre multipole decomposition given in (5.31).

One can show (Beutler et al. 2019) that at first order ($n = 1$), choosing one galaxy as the line-of-sight leads to

$$\begin{aligned} \xi_1^{(1)}(s) &= -\tfrac{3}{5}\xi_2^{(0)}(s) \\ \xi_3^{(1)}(s) &= \tfrac{3}{5}\xi_2^{(0)}(s) - \tfrac{10}{9}\xi_4^{(0)}(s) \end{aligned} . \tag{5.36}$$

This correction can be easily added in the above window matrix formalism, introducing the $(s/d)^n$ expansion in (5.32). Hence, additional window matrices, of order 1, have to be computed:

$$\mathcal{W}_\ell^{(n)}(s) = \frac{2\ell + 1}{4\pi \times A} \int \mathrm{d}\Omega_s \int \mathrm{d}\mathbf{x} d^{-n} W(\mathbf{x}) W(\mathbf{x} - \mathbf{s}) \mathcal{L}_\ell(\hat{\mathbf{d}} \cdot \hat{\mathbf{s}}). \tag{5.37}$$

The impact of the wide-angle effect is shown in Fig. 5.6. As expected, no strong effects are visible for the DESI QSO Y1, since the quasars are far enough away from us for the local plane-parallel approximation to hold. The effect is only visible in the very large scales *i.e.* for small k.

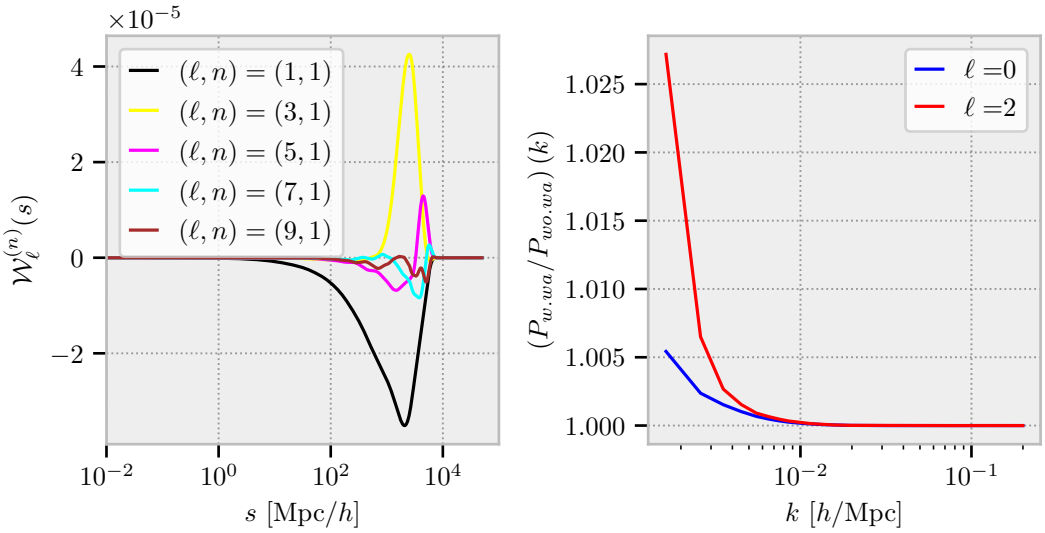


Figure 5.6: Left: Real space window matrix at order $n = 1$ for the DESI QSO Y1. Right: Ratio of convolved power spectrum with window matrix including wide-angle effect at order 1 to the one with window matrix not including the wide-angle effect.

► **Integral constraint** De Mattia and Ruhlmann-Kleider (2019) noted that the real selection function is unknown and will strongly depend on the luminosity function of the considered tracer or of the efficiency to determine the redshift. These effects cannot be, unfortunately, handled correctly by the randoms, and this leads to an additional correction in the theoretical correlation function or power spectrum.

In the following, these effects are expected to be small enough compared to the parameter statistical errors and will be neglected.

5.1.2.4 Computing the covariance matrix

The covariance associated with the power spectrum measurement will be computed as the covariance between measurements done in a large set of realistic simulations, commonly called *mocks*, that emulate as faithfully as possible the observations.

Real N-body simulations, see Section 5.2.1.1, are unfortunately too time-consuming to be produced in sufficient numbers. As in eBOSS (Zhao et al. 2021), we will use the approximate method known as *EZmocks* (Chuang et al. 2015). This method will generate, at an effective redshift, a galaxy field with position and velocity that follows an input power spectrum, thanks to the *Zel'dovich approximation* (Zeldovich 1970). See Zhang et al. (2023) for an assessment of the covariance matrix from EZMocks.

The EZmocks generated for DESI are similar to what is described in Zhao et al. (2021). We use 2000 boxes of 6 Gpc h^{-1} side generated at $z = 1.4$, 1000 for the NGC and 1000 for the SGC. These two regions are completely disconnected during the power spectrum measurement and therefore can be treated independently.

Remark: Despite the large size of the boxes, quasars, that are too dispersed in redshift, cannot be fully emulated without repeating the box. To avoid any problems, we will stop with the maximum redshift range possible with these simulations without repeating the box *i.e.* $z_{max} = 3.1$. This leads the nominal choice for the rest of the analysis.

Below, we describe the steps involved in creating the most realistic simulation possible. All these steps were performed with `mockfactory` [11] an MPI-based code to generate cutsky mocks from box simulations, which I have helped to develop. Although the EZmocks are already in *cutsky* form *i.e.* transformed from the cartesian coordinates to the (R.A., Dec., z) observational coordinates, these steps will be still described. These steps will be applied for both data and randoms, in the exact same way. Starting from box simulation with position and velocity of galaxies in cartesian coordinates:

- **From cartesian to sky coordinates:** First, the box, which is generally periodic, is remapped according to Carlson and White (2010) in order to increase the potential sky coverage of the simulation. Then, the remapped box is moved along an axis and it is cut out all the objects which are not in the desired redshift range and the sky coverage. Hence, there is a sweet spot to find between the simulated redshift range and the available sky coverage at a fixed box size. Fig. 5.7 shows the remapping method and how the box is then transformed into a cutsky simulation.

- **Redshift space distortion:** Once the simulation is transformed into a cutsky, one needs to add the redshift space distortion effect along the line-of-sight. The position $\mathbf{r}$ in real space will be translated to the redshift space $\mathbf{s}$, following Kaiser (1987),

$$\mathbf{s} = \mathbf{r} + \frac{\mathbf{v} \cdot \mathbf{l}}{aH(a)}\mathbf{l} \tag{5.38}$$

  where $\mathbf{l}$ is the line of sight and the velocity is given in km.s^{-1}. The conformal Hubble parameter $\mathcal{H} = aH(a)$ is sometimes called also the *RSD factor* and converts the velocity into h^{-1}Mpc.

- **Match density and $n(z)$ distribution:** Since the photometric survey used for the target selection, see Section 3.2.3.1, exhibits three distinct photometric regions (North, South, DES), we will match independently the density and the redshift distribution in these distinct regions. See Fig. 3.14 for their positions in the sky. In addition, we will split the region South in two: South (NGC) and South (SGC) because the NGC and SGC part of the mocks are generated independently.

  Fig. 5.8 shows the relative fluctuation of the redshift distribution in the different photometric regions and Table 5.2 gives the the density of the quasars in these regions. In particular, DES has more high-z quasars than the others because the imaging is deeper and DES and North have less low-z quasars than in the South since the PSF is better resolved.

  Since the fiber assignment or the completeness is applied after, we match the density computed from the data weighted by the completeness weights.

Table 5.2: Quasar density in the different photometric region computed with the completeness weights.

region	density [deg^{-2}]
North	186.63
South (NGC)	188.73
South (SGC)	188.73
DES	192.73

[11] https://github.com/cosmodesi/mockfactory

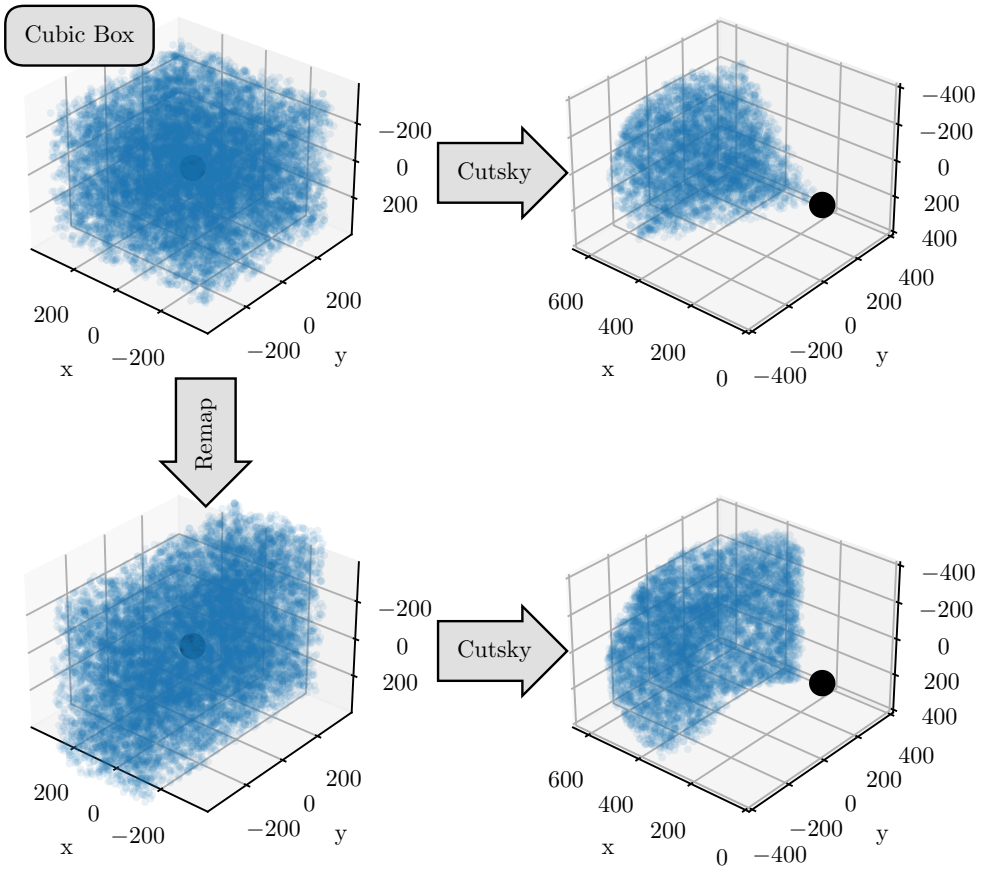


Figure 5.7: Top left: a standard periodic cubic box. Top right: the black point is the observer. The cubic box is moved along x such that the redshift coverage starts at 0, then the box is cut to match a real survey. Bottom left: the periodic box is remapped according to Carlson and White (2010) to increase the box's sky coverage. Bottom right: box is moved along x. Due to the remapping, the box is not deep enough to have similar redshift coverage to the unmapped box. However, the sky coverage is broader.

- **Match DESI footprint:** To reproduce exactly the DESI footprint, one needs to apply the fiber assignment which was used during the observation *i.e.* selecting one data only on each working fiber, or all the randoms reachable by the working fibers. Unfortunately, the fiber assignment process currently takes too long to be applied 1,000 times over[12]. As shown in Section 5.1.1.3, large scales are mainly impacted by the completeness of the observation.

  Thus, we first only keep objects which are located in a DESI tile and then, we simply downsample the data and the randoms via an Healpix map at $N_{side} = 256$ representing the fraction of the pixel that was observed. We finally remove objects that are in bad imaging regions *i.e.* that are located in the imaging maskbits 1, 7, 8, 11, 12 and 13 of the legacy survey.

  Note that, maskbits 1, 12 and 13 are from the target selection and should be removed

[12] I developed an MPI-based version that considerably reduces runtime, but this effect is negligible for the large-scale studies.

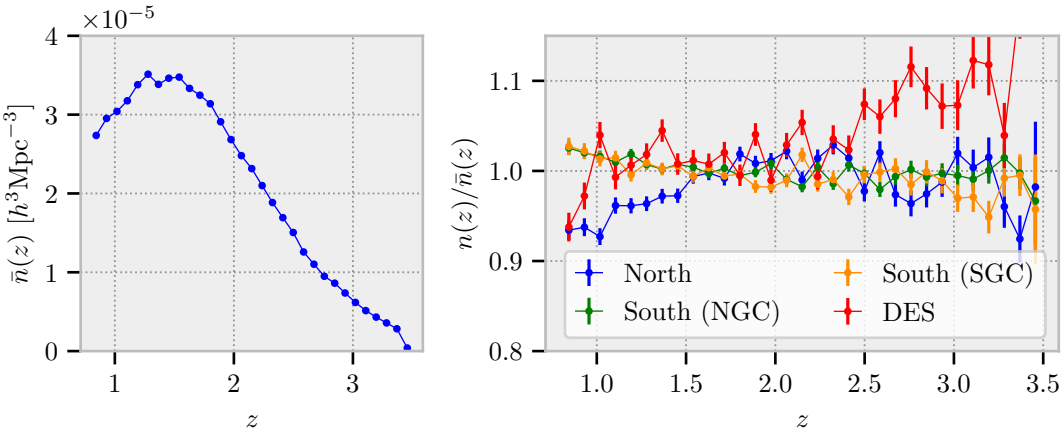


Figure 5.8: Left: redshift distribution of DESI QSO Y1 weighted by the completeness on all the footprint. Right: the ratio between the redshift distribution computed in the specific photometric region and all the footprint. In particular, there are fewer quasars in the North at low redshift and more quasars at high redshift in DES.

before the fiber assignment, while maskbits 7, 8 and 11 are quality masks added after the observation. Whether or not to include the latter maskbits will have to be studied at a later date.

Once each cutsky is matched to the DESI Y1 observations, we compute the power spectrum with different sets of weights, such as FKP or OQE weights. Fig. 5.9 shows the power spectrum of the 1000 EZmocks (NGC+SGC) matching the Y1 observations without FKP or OQE weights.

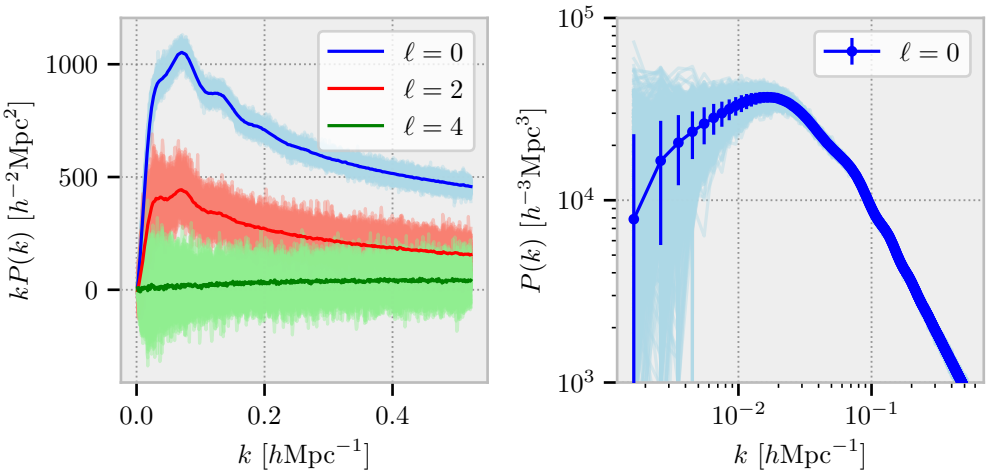


Figure 5.9: Multipoles (full lines) of the power spectrum are the average on 1000 EZmocks emulating the DESI QSO Y1, without any FKP or OQE weights. Light lines beyond show the individual measurement and will give the covariance. The errors on the right are the standard deviation (diagonal of the covariance matrix) of the 1000 EZmocks.

Remark: The power spectra of the EZmocks, although they were not generated at the corrected effective redshift, have the same amplitude as the data, see Section 5.4.2.1, and therefore could be used to compute the covariance matrix.

Finally, the covariance C_{ij}, that will be used for the parameter estimation, will be simply

the covariance between these 1000 measured power spectra. Since the covariance is estimated from a finite number of mocks, C_{ij}^{-1} provides a bias estimation of the inverse of the covariance due to the skewed nature of the inverse Wishart distribution. As proposed in Hartlap et al. (2007), this bias can be corrected by re-scaling the inverse of the covariance as

$$C_{ij}^{-1} \equiv \frac{N_m - n - 2}{N_m - 1} C_{ij}^{-1}, \tag{5.39}$$

where N_m is the number of mocks and n is the number of data points. Typically, in the following $N_m = 1000$ and $n = 116$ such that the Hartlap factor corresponds to ~ 0.88. For the moment, we do not include the extra correction from Percival et al. (2014) that is expected to have only a minor contribution to the final errors.

Caveat: When I match the redshift distribution of the data and randoms, I remove all the objects in the first half bin and in the last half bin of the redshift histogram, due to an interpolation problem. In practice, the objects are generated with $0.818 < z < 3.081$ instead of $0.8 < z < 3.1$, and this effect is completely negligible for the accuracy of our measurement since the density used is the one for $0.8 < z < 3.1$.

5.1.2.5 Parameter estimation

In Section 5.1.1.1, we derive how the power spectrum will be measured from the spectroscopic data. Then, in Section 5.1.2.1, we described the theoretical model used to estimate the amount of primordial non-gaussianity. Finally, in Section 5.1.2.4, we have explained how to generate a covariance matrix to describe the statistical errors for each measured point and the correlation between them. These different steps are summed up in Fig. 5.10.

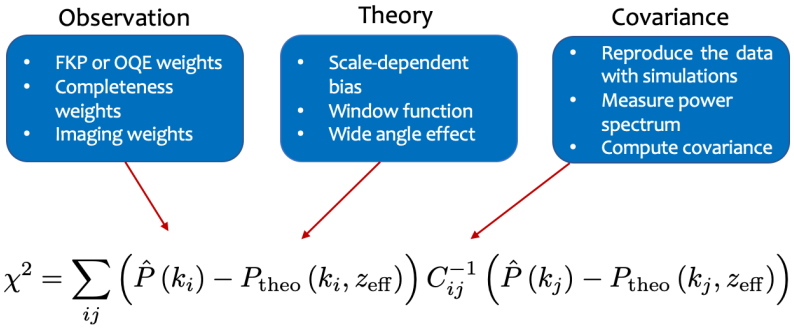


Figure 5.10: Summary of the different steps to constrain cosmological parameters with a galaxy survey.

The parameter estimation will be done in two steps:

- The parameter value will be the best-fit value from the χ^2 minimization. In the following, we will use `minuit`[13] to perform the minimization (James and Roos 1975).
- The errors will be estimated as the 68% of the confidence level of the posterior computed via the Markov chain Monte Carlo (MCMC) method. In particular, we will use `emcee`[14] (Foreman-Mackey et al. 2013).

To process each of these steps, we will use the coherent and user-friendly framework `desilike`[15].

[13] https://iminuit.readthedocs.io/en/stable/
[14] https://emcee.readthedocs.io/en/stable/
[15] https://github.com/cosmodesi/desilike

5.2 Validation with quasi N-body simulations

A critical part of the measurement of primordial non-gaussianity in the power spectrum is the correction of the large-scale contaminant. In particular, the imaging systematics presented in Section 4.3 which, if corrected, could result in the removal of excess power on a large scale, thus biasing this measurement.

In order to quantify the efficiency and safety of the method presented in Chaussidon et al. (2022), we will test it on realistic quasi-Nbody simulations that emulate the real impact of the primordial non-gaussianity.

Here, we will only briefly describe the N-body and quasi N-body simulations used in cosmology; see Angulo and Hahn (2022) for an excellent, comprehensive and up-to-date review of the subject!

5.2.1 Simulate the Universe

5.2.1.1 N-body simulation

Simulating the full range of interactions between all the constituents of the universe, from large-scale gravity to small-scale electromagnetic interactions, requires extraordinarily high resolution and astronomical computing times. While such simulations are necessary for the study of, for instance, Ly-α forests, they cannot be produced with volumes large enough to simulate the portion of the universe we will be observing with DESI.

One solution is to neglect all interactions other than gravity. As we saw in Section 1.2.3.5, gravity, at scales of interest, in a matter-dominated Universe, is well described by Newtonian gravity in comoving coordinates. Matter will be described by dark matter particles corresponding to a given mass and they will evolve step by step by computing the Newtonian gravity force between them. Such simulations are called *N-body simulations*.

To speed up execution time, particle displacement under the force of gravity is separated into two steps at each time step. First, the force of gravity is calculated on a large scale, and then the particles are moved according to this force. To do this, the large-scale Φ gravity field is determined using the Poisson equation in Fourier space (1.102). This step involves calculating the Fourier transform of the density field and it is known as a *particle mesh* computation. Then, to account for small scales, the force of gravity is calculated for each particle in its traditional $1/r^2$ form, taking into account only the particles in a reduced sphere to reduce computation time. The particles are then moved to their final state for this time iteration.

The first step can be performed with a Fast Fourier Transform (FFT) algorithm, which has a complexity of $\mathcal{O}(Nln(N))$, whereas the traditional calculation of the force of gravity experienced by the N particles has a complexity of $\mathcal{O}(N^2)$. For more details, see the reference code `GADGET` (Springel 2005) or the more recent implementation designed for DESI (Garrison et al. 2021).

The second stage, when the number of particles is sufficiently large, enables us to model the gravitational collapse that occurs on small scales. However, when we're only interested in larger scales, the first step may be more than adequate. These are known as *quasi N-body* simulations. In the following, we will use `FastPM`[16] (Feng et al. 2016) for particle mesh solver, and we will follow the parametrization done in Ding et al. (2022).

[16] I use the pythonic version: https://github.com/echaussidon/fastpm-python

5.2.1.2 Initial conditions

► **Generate field from a power spectrum** We will need to generate an initial density field following the desired initial power spectrum. Let's generate the field δ such that

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3}\int_{\mathbb{R}^3} \mathrm{d}\mathbf{k} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \varphi(k) W(\mathbf{k}), \tag{5.40}$$

where W is a *white noise i.e.* a complex-valued 3D random field[17] satisfying

$$W(\mathbf{k}) = W(-\mathbf{k}), \quad \langle W(\mathbf{k})\rangle = 0, \quad \left\langle W(\mathbf{k}) W^\star\left(\mathbf{k}'\right)\right\rangle = \delta_D\left(\mathbf{k}-\mathbf{k}'\right), \tag{5.41}$$

and φ is chosen in function of the desired power spectrum $P(k)$:

$$\varphi(k) = (2\pi)^{3/2}\sqrt{P}(k). \tag{5.42}$$

Hence, δ satisfies

$$\left\langle \phi(\mathbf{k})\phi^\star\left(\mathbf{k}'\right)\right\rangle = (2\pi)^3 P(k)\delta_D\left(\mathbf{k}-\mathbf{k}'\right). \tag{5.43}$$

► **Reducing cosmic variance** The white noise can be polar decomposed as $W(\mathbf{k}) \equiv A(\mathbf{k})\mathrm{e}^{\mathrm{i}\theta(\mathbf{k})}$ where A follows a *Rayleigh distribution* and θ a uniform distribution in $[0, 2\pi)$. The power spectrum is independent of the phase θ and

$$\delta_D\left(\mathbf{k}-\mathbf{k}'\right) = \left\langle W(\mathbf{k}) W^\star\left(\mathbf{k}'\right)\right\rangle = \left\langle A(\mathbf{k}) A\left(\mathbf{k}'\right)\right\rangle. \tag{5.44}$$

For one realization of $W(\mathbf{k})$, we expect fluctuation in $\langle W(\mathbf{k})W^\star(\mathbf{k}')\rangle$ around the ensemble average of the order of $1/sqrtN_k$ where N_k is the number of independent modes, in a finite interval $k, k+dk$, available in the simulation. This is the cosmic variance, and it is related to the volume of the simulation.

It can be drastically reduced by sampling A from a Dirac distribution such that $A(\mathbf{k}) = \delta_D(\mathbf{k})$ (Angulo and Pontzen 2016). This is commonly referred to 'fixing' or 'unitary' method, and it allows us to generate large simulations with low cosmic variance *i.e.* with only the desired physical signal without noise. The cosmic variance can be also reduced by using pairs of simulations where the second simulation is initialized with $W^\star(\mathbf{k})$ instead of $W(\mathbf{k})$, see Avila and Adame (2023) for a recent utilization with PNG simulations.

In our case, we want to probe the impact of the imaging systematics on the PNG signal. Hence, we will use the 'unitary' method to reduce the cosmic variance.

► **Generate initial particle position** The particle position and velocity are set via the 1 order of the Lagrangian perturbation theory, known as the *Zel'dovich approximation* (Zeldovich 1970)[18].

First, we generate a random uniform distribution of particles, and then, we update their position and give their velocity following

$$\mathbf{x} \to \mathbf{x} - \nabla\nabla^{-2}\delta(\mathbf{x}), \tag{5.45}$$

and set the velocity $\mathbf{v} = a\mathbf{v}_p$ with

$$\mathbf{v} \to -aHf\nabla\nabla^{-2}\delta(\mathbf{x}). \tag{5.46}$$

[17] It is called white noise since its power spectrum does not depend on $\mathbf{k}$

[18] See White (2014) for a modern review.

► **Initial conditions with $f_{\rm NL}$** In `FastPM`, the initial density field is generated at $z = 0$ and this field is then *rescale* with the growth factor at the initial redshift used for the particle generation. The initial density field is generated following these steps:

- Generate density field $\delta(\mathbf{k}, z = 0)$ according to the matter power spectrum.
- δ is linked to the primordial potential ϕ_{prim} via the transfer function (2.106)

$$\delta(\mathbf{k}, z) = \alpha(\mathbf{k}, z)\phi_{prim}(\mathbf{k}). \tag{5.47}$$

- Add primordial non-gaussianity in real space:

$$\phi^{NG}(\mathbf{x}) = \phi_{prim}(\mathbf{x}) + f_{\rm NL}\left(\phi^2_{prim}(\mathbf{x}) - \langle\phi^2_{prim}\rangle\right). \tag{5.48}$$

- Create initial non Gaussian density field $\delta^{NG}(\mathbf{k}, z = 0)$ thanks to the transfer function again:

$$\delta^{NG}(\mathbf{k}, z = 0) = \alpha(\mathbf{k}, z = 0)\Phi^{NG}(\mathbf{k}) \tag{5.49}$$

Remark: In the case of EZmocks, mentioned in Section 5.1.2.4, we just emulate the power spectrum at the desired effective redshift in the same way as described above, using the second order in Lagrangian perturbation theory. The particle field does not evolve with time, it is generated at a specific redshift with a specific shape of power spectrum. EZmocks do not emulate the gravity, they just reproduce a theoretical model. In our case, we use quasi N-body simulation to reproduce the behavior of the gravity into the halo formation. Hence, with these simulations, we can validate the theoritical model (Slosar et al. 2008, see, for instance,).

5.2.1.3 From box to realistic simulation

Once the dark matter particle field has been generated at redshift $z = 19$ (in our case), we evolve the particles in 40-time steps to $z = 1.5$, saving the particle field at different redshifts (*snapshot*) $z \in [2.5, 2.25, 2.0, 1.75, 1.5]$.

► **From dark matter particles to quasars** Despite the complexities of galaxy formation, dark matter represents 80% of the matter in the Universe and so there is a robust, well-established fact: galaxies and quasars reside in massive, gravitationally bound structures called halos. This can be emulated by carrying out the following two steps:

- Halos finder: First, we need to identify halos *i.e.* dark matter particle clusters in which galaxies and quasars are in the real Universe. This step is achieved with an MPI version (Feng and Modi 2017)[19] of the popular *friends-of-friends* (FOF) algorithm (Huchra and Geller 1982, Press and Davis 1982, Davis et al. 1985). It builds groups of particles with a linking length b for the mean interparticle distance. The larger the linking length, the more particles will constitute each group. We will use, the adopted value, $b = 0.2$ corresponding to form halos with density $\Delta = 180$ times higher than the average "global" density, which is typically the overdensity used for another popular halo finder, as the so-called spherical overdensity algorithm (Warren et al. 1992).

  See Biagetti et al. (2017) for a comparison between FoF, spherical overdensity, and `Rockstar`[20] algorithms with real N-body simulations in the particular scope of primordial non-gaussianity.

[19] I improve this version, see here, to make it executable on many number of processors simultaneously.

[20] `Rockstar` (Behroozi et al. 2013) is generally considered the best method for extracting the physical limit of dark matter halos.

- Halos Distribution Function (HOD): Then, we need to populate the halos with galaxies or quasars. Several methods exist as abundance matching or the halos distribution function. We will use the latter with the simplest distribution function by selecting all the halos above a specific mass. For more complex models, one can have a look at the DESI paper on this topic (see, for instance, Yuan et al. 2023, Rocher et al. 2023). Note that with N-body simulations, we only have access to the mass of the halos, so we cannot check whether the tracer we are using verifies the universal mass relation (2.119).

▶ **Calibrate the Mass cut** To find the expected bias from Laurent et al. (2017), we need to calibrate the mass cut for the HOD step. Fig. 5.11a shows the bias measured as a function of the redshift for different mass cuts. We find that $M_c = 2.25e12$ $M_\odot$ gives the expected bias. Note that this minimal mass cut is in agreement with the one find in Table 4 of Laurent et al. (2017).

To increase the size of the box, without increasing the number of meshes and so the RAM needed to run the simulation, we test the dark matter particle mass dependence. Fig 5.11b shows the power spectrum and the ratio for two simulations with the same volume, but one has better resolution, see Table 5.3. Keeping the same mass cut ($M_c = 2.25e12$ $M_\odot$) for HOD step, the high-resolution simulation features a minimum of 178 dark matter particles per halo, while the low-resolution simulation a minimum of 22. The residual $\sim 10\%$ effect in the ratio is scale-independent and has no impact on the behavior of the scale-dependent bias. Due to the better resolution, the mass distribution function of the halos is shifted towards the lower masses, so the effective bias in the high-resolution simulation is lower than in the low-resolution simulation. Finally, using $1e11$ $M_\odot$ as dark matter particle mass seems enough to reproduce the desired scale-dependent bias.

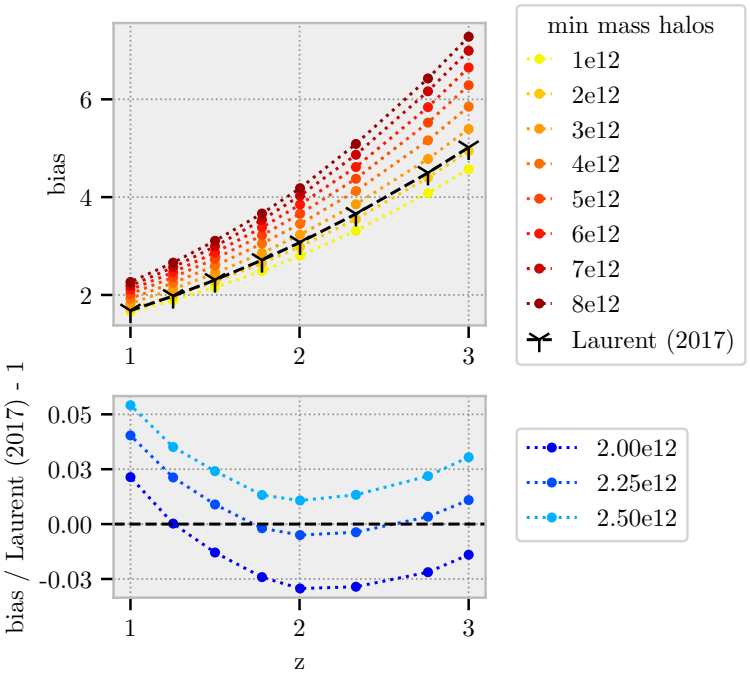
▶ **Final run** I ran these simulations on the National Energy Research Scientific Computing Center (NERSC) hosted by the Lawrence Berkeley National Laboratory (LBNL) with the KNL partition of Cori. Fig. 5.12 shows a part of Cori in LBNL dedicated building.

The three sets of simulations are summarized in Table 5.3. The first one (run-knl-1) was the initial setup to calibrate the minimum halos' mass and validate the initial condition with primordial non-gaussianity that I wrote. The second set (run-knl-2) was used to validate the resolution of run-knl-1. Finally, the last set (run-knl-3) was the largest box with the resolution of run-knl-1 that I could run easily. The power spectrum for redshift $z = 2.5, 2.0, 1.5$ are displayed in Fig. 5.13, in which we can clearly identify the scale-dependent bias generated by non-zero $f_{\rm NL}^{\rm loc}$.

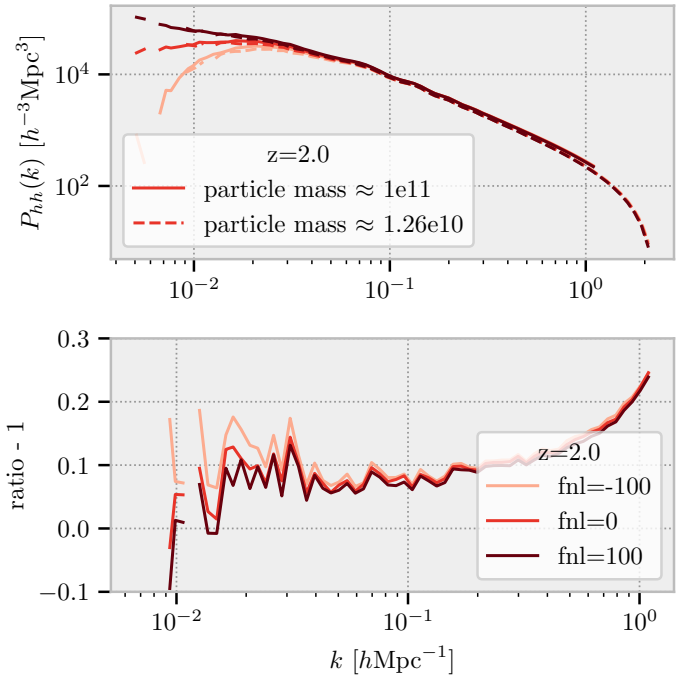
Table 5.3: Parameters for the three sets of simulations that I ran on NERSC. All the simulation are run with the same seed and uses the unitary method to reduce the cosmic variance.

	run-knl-1	run-knl-2	run-knl-3
f_{NL}	-100/0/100	-100/0/100	-25/0/12/25
Boxsize [Gpc/h]	2760	5520	5520
Nmesh	3000	3000	6000
Particle Mass [$M_\odot$]	$\approx 1e11$	$\approx 1.26e10$	$\approx 1e11$

Remark: I was limited by the memory required to build a grid with enough meshes inside. For instance, increasing by a factor of two the number of meshes on each side *i.e.* by a factor of 8 in run-knl-3, requires too much memory for the entire KNL partition. The largest boxes took on 1600 KNL nodes, 5 hours to be generated and required about 153TB of random access memory (RAM).



(a) QSO bias in fastpm simulations.



(b) Impact of particle mass in fastpm simulation.

Figure 5.11: (a) Top: Linear bias as a function of the redshift computed in one of our FastPM simulations with different mass cuts during the HOD step. The expected bias from Laurent et al. (2017) is plotted in black. Bottom: Ratio between the measured bias and the expected bias from Laurent et al. (2017) for a more restrictive choice of mass cut. (b) Top: Power spectrum measured at $z = 2$ with $M_c = 2.25e12$ $M_\odot$ for high (particle mass $1.26e10$ $M_\odot$) and low (particle mass $1e11$ $M_\odot$) resolution simulations (see Table 5.3) and for $f_{\rm NL}^{\rm loc} \in [-100, 0, 100]$. Bottom: Ratio between the two.

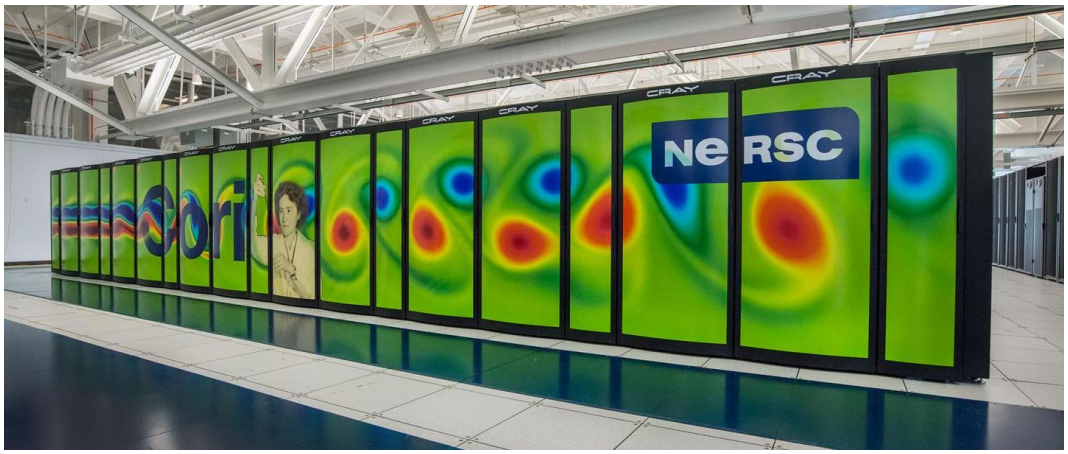


Figure 5.12: Cori is the 8^{th} supercomputer of the National Energy Research Scientific Computing Center (NERSC[21]) hosted by the Lawrence Berkeley National Laboratory (LBNL). Cori is made of 9,688 *Intel Xeon Phi Processor 7250* [22](KNL nodes) and 2,388 *Intel Xeon Processor E5-2698 v3* [23](Haswell nodes).

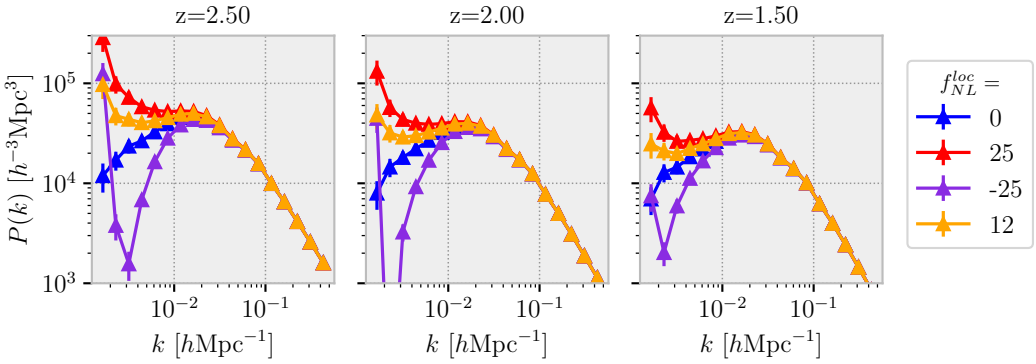


Figure 5.13: Power spectrum measured on my 5.5 Gpc/h boxes with $f_{\rm NL}^{\rm loc} \in [-25, 0, 12, 25]$ for three snapshots $z = 2.5, 2.0, 1.5$. Error bars are theoretical predictions with respect to the number of observed independent modes in the box.

► **Mimicking the DESI survey** Finally, we follow the step describe in Section 5.1.2.4 to generate realistic cutsky from the run-knl-3 boxes. Unfortunately, even with a 5.5 Gpc/h box and the remapping algorithm, I am just able to reproduce either North, South (NGC) or SGC (South and DES) with a redshift range of $0.8 < z < 2.65$ for the first two and $0.8 < z < 3.1$ for the SGC.

Additionally, from on box, I can extract 16 different subsamples that match the DESI QSO survey with a density of 280 deg^{-2}. The cutskys are generated with a density higher than the one observed in the data, see Table 5.2, in order to contaminate these cutskys with realistic imaging systematics. Although they are not completely independent, we will use the mean of these 16 samples to reduce the cosmic variance.

Table 5.4: Best fit values from the mean of the 16 subsamples with realistic DESI QSO Y1 covariance matrix rescaled by 16. Errors are the 1σ (68% CL) from the posteriors shown in Fig. 5.15a and in Fig. 5.15b.

		$f_{NL}^{\rm loc}$	b	$s_{n,0}$	Σ_s	$\chi^2/(115-4)$
$f_{NL}^{\rm loc}=0$	uncont.	$2.3^{+3.4}_{-2.9}$	$2.604^{+0.022}_{-0.017}$	550^{+230}_{-200}	$1.95^{+0.72}_{-0.36}$	4.23
	cont. + corr.	$12.9^{+2.8}_{-2.6}$	$2.603^{+0.019}_{-0.017}$	550^{+210}_{-200}	$2.37^{+0.50}_{-0.28}$	4.07
$f_{NL}^{\rm loc}=12$	uncont.	$10.1^{+3.2}_{-2.5}$	$2.606^{+0.020}_{-0.018}$	470^{+190}_{-230}	$1.85^{+0.69}_{-0.37}$	3.94
	cont. + corr.	$22.0^{+2.9}_{-2.5}$	$2.575^{+0.021}_{-0.016}$	850^{+210}_{-210}	$2.55^{+0.38}_{-0.33}$	3.89
$f_{NL}^{\rm loc}=25$	uncont.	$20.4^{+2.4}_{-2.5}$	$2.616^{+0.019}_{-0.017}$	380^{+210}_{-200}	$2.04^{+0.59}_{-0.35}$	3.88
	cont. + corr.	$30.9^{+2.4}_{-2.1}$	$2.587^{+0.017}_{-0.017}$	700^{+180}_{-230}	$2.11^{+0.64}_{-0.28}$	3.97
$f_{NL}^{\rm loc}=-25$	uncont.	$-15,7^{+3.8}_{-3.6}$	$2.625^{+0.021}_{-0.021}$	480^{+210}_{-230}	$1.98^{+0.67}_{-0.31}$	3.27
	c ont. + corr.	$-1.9^{+3.3}_{-3.3}$	$2.620^{+0.020}_{-0.020}$	450^{+230}_{-210}	$2.41^{+0.45}_{-0.33}$	3.93

5.2.2 Validate imaging systematic mitigation

5.2.2.1 Impact of imaging systematic mitigation

Exactly as performed for the angular cases, see Appendix 5.1, we will test the mitigation method proposed in Chaussidon et al. (2022) by comparing the impact of the mitigation on contaminated cutskys to uncontaminated cutskys. We will, in particular, follow the colors used in the pipeline shown in Fig. A5.1, where uncontaminated cutsky is in blue, contaminated cutsky is in green, and the correction of contaminated cutsky is in red.

In this test, we consider the SGC part of the footprint, and we contaminate the cutsky with realistic contamination matching the DESI QSO Y1 imaging systematics displayed in Fig. 4.44.

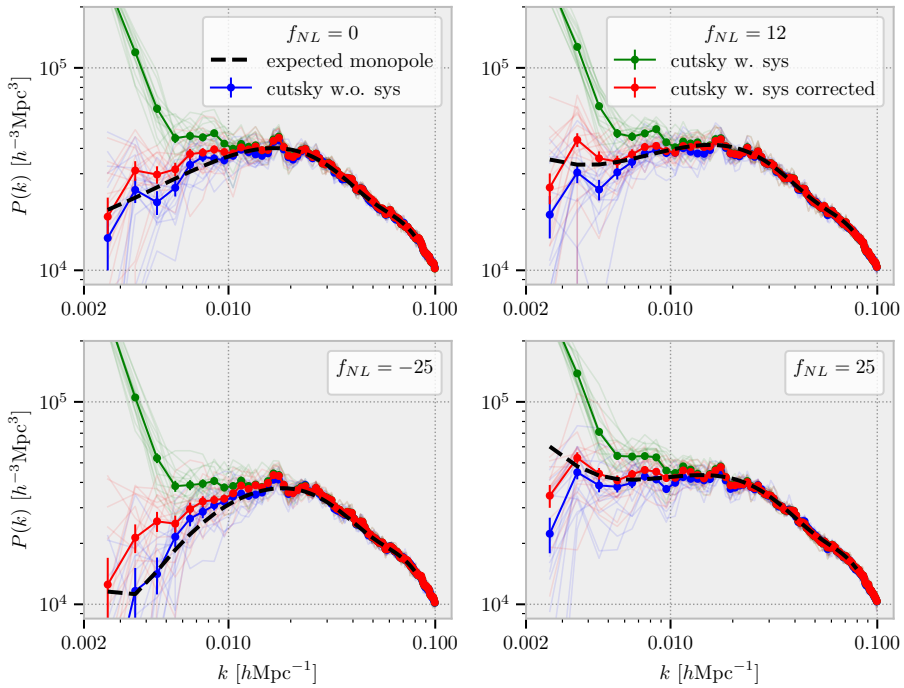
Fig. 5.14 shows the monopole and the quadrupole of the mean on the 16 subsamples extracted from the same box for four different values of $f_{\rm NL}^{\rm loc}$. The expected window convolved model (5.32) is shown in black.

5.2.2.2 Systematic bias from imaging systematic mitigation

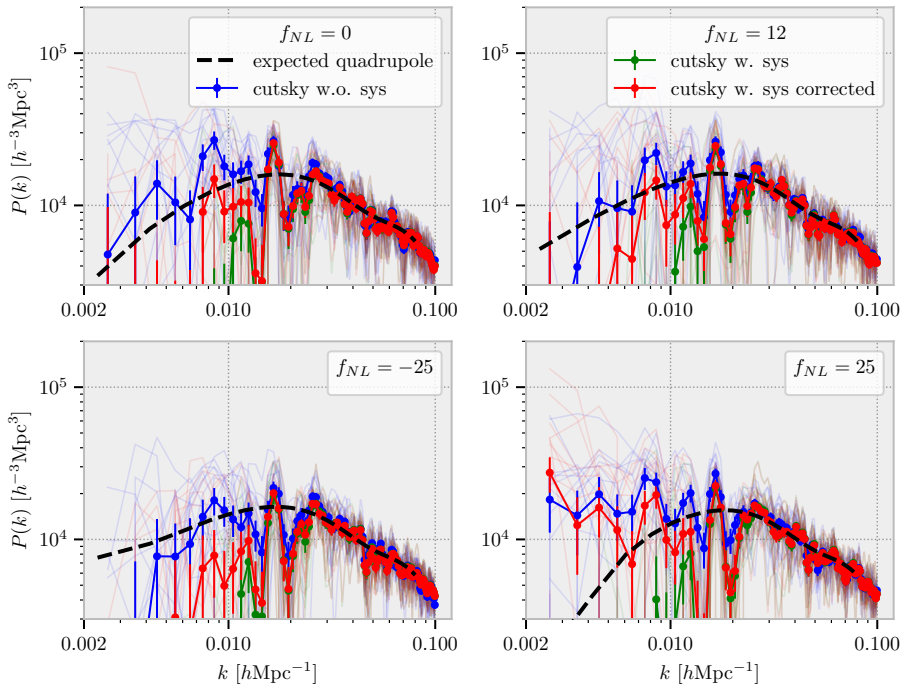
To assess any systematic bias in the correction, we fit the mean of 16 subsamples of the monopole and the quadrupole in the uncontaminated and contaminated with `regressis` correction cases with $0.003 < k < 0.08$, $k_{step} = 0.001$ $[h\ \mathrm{Mpc}^{-1}]$ for the monopole, and $0.003 < k < 0.08$, $k_{step} = 0.002$ $[h\ \mathrm{Mpc}^{-1}]$ for the quadrupole. We use the 1000 EZmocks (only SGC part) of DESI Y1 to build the covariance, and we rescale it by 16. Note that, here, the universal mass relation (2.119) ensures $p = 1$.

The result of these fits is given in Table 5.4 for the four different values of $f_{\rm NL}^{\rm loc}$. Posteriors are given in Fig. 5.15a for the uncontaminated mocks, following the color code from Fig. 5.13, and in Fig. 5.15b for the contaminated mocks with correction.

First, our cutsky fastpm simulations reproduce correctly the expected scale-dependent bias for the given $f_{\rm NL}^{\rm loc}$ with the exception that we find, as in Biagetti et al. (2017), a lower amplitude for the scale-dependent bias about 80%. Note that this effect is more important for our $f_{\rm NL}^{\rm loc} = -25$ simulation.



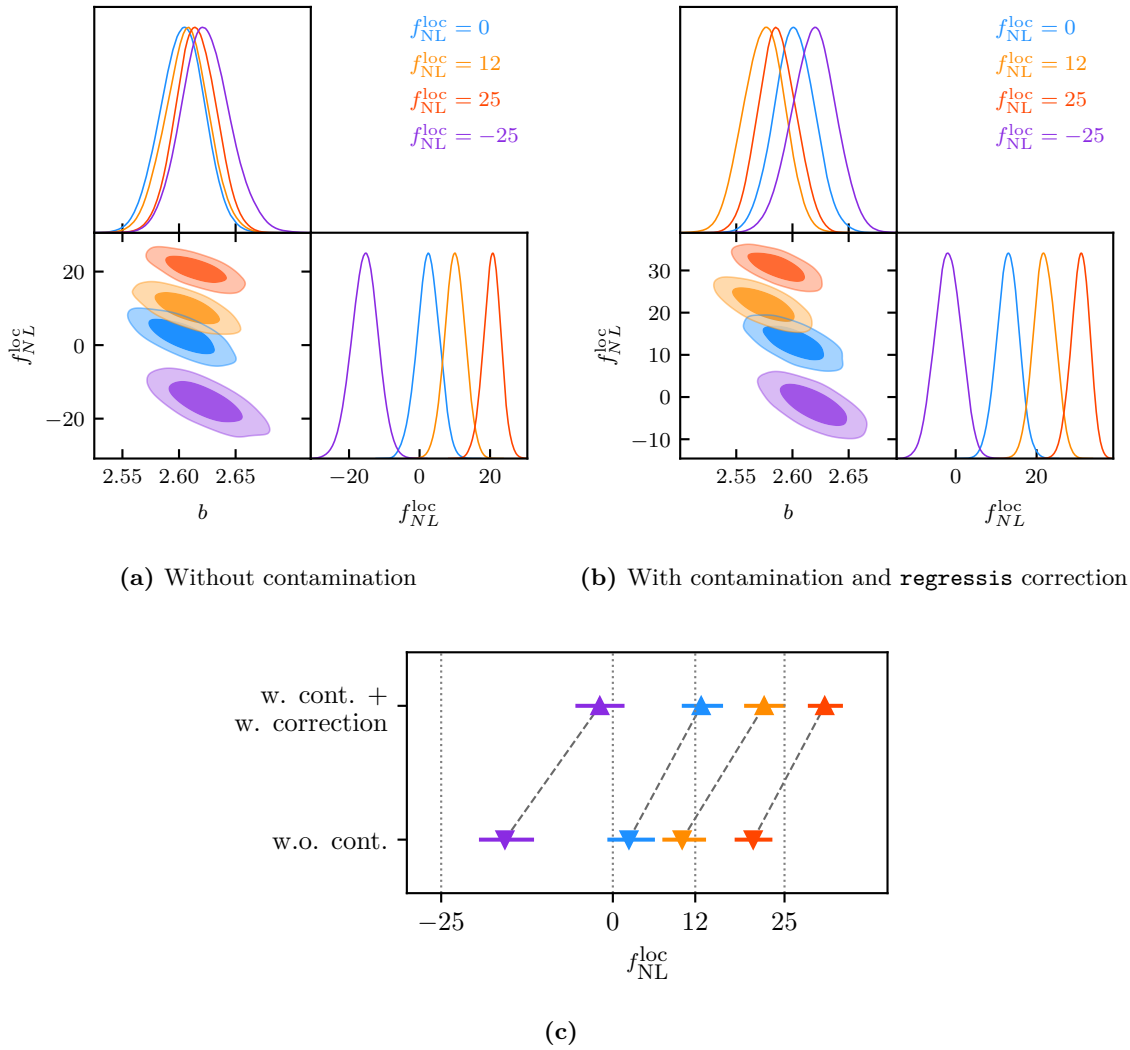
(a) Monopole ($\ell = 0$)



(b) Quadrupole ($\ell = 2$)

Figure 5.14: (a) Monopole (b) Quadrupole of the mean on 16 subsamples extracted from the same box with $f_{\rm NL}^{\rm loc} \in [-25, 0, 12, 25]$ at $z = 1.75$. The power spectrum of each subsample is displayed in the background. The uncontaminated cutskys are in blue, the contaminated cutskys are in green, and the contaminated cutskys with `regressis` mitigation are in red. The black line is the predicted convolved model. The errors are from the 1000 EZmocks (only SGC) part renormalized by $\sqrt{16}$.

As noticed by Desjacques et al. (2009), Biagetti et al. (2017), this difference arises from the fact that the halos in our fastpm simulation are extracted with a Friend-of-Friend (FoF) halo finder which is known to not exactly reproduce the correct behavior of b_ϕ. One explanation could be that, compared with what was assumed in Section 2.2.3, the formation of halos is not strictly spherical.



(a) Without contamination

(b) With contamination and `regressis` correction

(c)

Figure 5.15: (a) Posteriors for the mean of the 16 subsamples from the different `fastpm` simulations with $f_{\rm NL}^{\rm loc} \in [-25, 0, 12, 25]$ in the uncontaminated case. (b) Same as (a) but for the contaminated with `regressis` correction case. (c) Best fit value for $f_{\rm NL}^{\rm loc}$, with errors as the 1σ (68% CL) from the posteriors, for the different simulations and the uncontaminated and the contaminated with correction cases.

Although it removes most of the excess power produced by systematics imaging, our mitigation method does not precisely reproduce the power spectrum measured in the uncontaminated case; see blue lines versus red lines in Fig.5.14. These residuals excess of power bias, unfortunately, our $f_{\rm NL}^{\rm loc}$ measurement. The best-fit value for the different simulations and cases are shown in Fig. 5.15c, illustrating a clear systematic bias after the corrections.

This systematic bias can be the sign of a non-perfect correction and could be reduced by fine-tuning `regressis`. However, the contamination used here was extracted from the data with `regressis`, so the mitigation can produce the expected correction. The systematic bias may therefore be due to how the correction is performed. In particular, it seems that this systematic

bias is almost the same for the different $f_{\rm NL}^{\rm loc}$. This bias was already observed in the angular case in Rezaie et al. (2023), and we may also need to correct it similarly.

This is the critical part of the scale-dependent bias measurement. In the following, we'll assume that this systematic bias is under control, although we will need to study this further to ensure an unbiased measurement.

Remark: Here, we compute the power spectra without FKP or OQE weights, and we need to test also the impact of the systematic mitigation in the presence of these optimal weights.

5.3 Blinding of the data set

As described in Section 5.2, the imaging systematics that occurs during the target selection, strongly bias the large-scale measurement. The method proposed in Section 4.3 learns unfortunately directly from the data and is known to be able to overfit that also biasing the large-scale measurement. This prompted us to validate the method using simulation in the previous Section, but this does not allow us to know whether or not all the imaging systematics are corrected.

Following the seminal work of Brieden et al. (2020) on the RSD and BAO, and in order to investigate unknown observational systematics without any confirmation bias ($f_{\rm NL}^{\rm loc}$ is expected to be closed to 0), we propose a method to blind the value of $f_{\rm NL}^{\rm loc}$ *i.e.* we will simulate the impact on the scale-dependent bias of an unknwon, randomly selected, $f_{\rm NL}^{blind}$ value.

In this section, we describe a method to simulate a fake scale-dependent bias that could be produced by $f_{\rm NL}^{\rm loc}$ and validate this method, in particular, in presence of $f_{\rm NL}^{\rm loc} \neq 0$ and with the combination of imaging systematic mitigation what we want to test without confirmation bias. This method is implemented in `mockfactory`[24] and is currently used in the DESI clustering analysis.

5.3.1 Theoretical description

5.3.1.1 Model

The main idea is to generate weights for the data catalogs to mimic the power spectrum with the scale-dependent bias (2.104) given any value of $f_{\rm NL}^{\rm loc}$. The **proposed weights** are:

$$w^{d}{}_{blind}(\mathbf{x}) = 1 + w_{blind}(\mathbf{x}), \tag{5.50}$$

where $w_{blind}(\mathbf{x})$ is the Fourier transform of

$$w_{blind}(\mathbf{k}) = \frac{b_\phi f_{\rm NL}^{blind}}{\alpha(k)} \times \frac{1}{b} \times \widehat{\delta}^r(\mathbf{k}) = a(k) \times \widehat{\delta}^r(\mathbf{k}), \tag{5.51}$$

b_ϕ and $\alpha(k)$ are computed at the effective redshift $z_{\rm eff}$ of the data. $\widehat{\delta}^r(k)$ is the estimation of the reconstructed field in real space. The reconstruction is performed with the iterative FFT reconstruction algorithm (Burden et al. 2015) implemented in `pyrecon`[25].

To perform the reconstruction, the first step is to paint the density field on the grid on which the density field is smoothed by a window function

$$W^s(k) = \exp\left(-\frac{1}{2}k^2\sigma^2\right)$$

[24] https://github.com/cosmodesi/mockfactory/blob/main/mockfactory/blinding/catalog.py
[25] https://github.com/cosmodesi/pyrecon/tree/mpi

where σ is the chosen smoothing radius. Since we only need to reconstruct the real-space density field at large scales (where PNG matters), we take $\sigma = 30$ Mpc h^{-1}. This window function for several values of σ is displayed in Fig. 5.16.

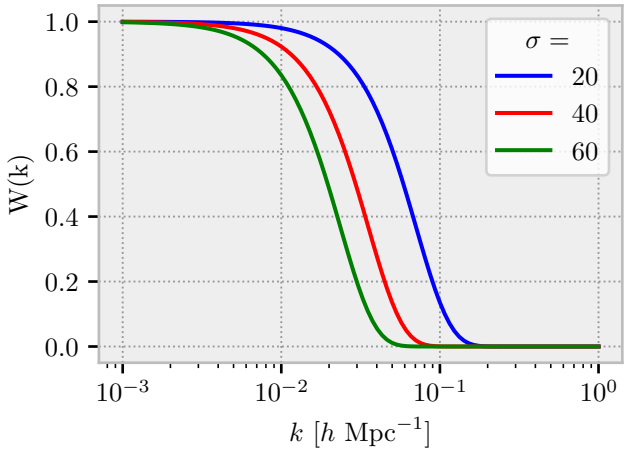


Figure 5.16: Reconstruction window function for several values of smoothing radius parameter σ.

Additionally, the density field is smoothed by the painting and the reading of weights on the grid at particle positions. For both steps, we use the CIC (cloud-in-cell) assignment scheme. The total window function which is applied to the density field is:

$$W^{\rm tot}(k) = W^{s}(k) \times W^{cic}(k)^2, \tag{5.52}$$

where

$$W^{cic}(k) = \mathrm{sinc}^2\left(\frac{\pi k}{2k_N}\right)$$

with $k_N = \pi/H$ is the Nyquist frequency corresponding to the cell size H. Here, we take the window function to be *isotropic*, as an approximation for the 3D:

$$W^{cic}(\mathbf{k}) = W^{cic}(k_x)\ W^{cic}(k_y) W^{cic}(k_z).$$

Thus, the real field that we evaluate is such that:

$$\langle|\widehat{\delta}^{r}(k)|^2\rangle \propto W^{\rm tot}(k)^2 \times \left[b^2 P_{lin}(k) + 1/n\right], \tag{5.53}$$

where $1/n$ is an undesired shot noise term, that if not corrected for will bias the blinding procedure as discussed below.

5.3.1.2 From galaxy catalog to FKP field

With this blinding scheme applied to the data, the FKP field (5.3) is

$$\begin{aligned} F(x) &= n_g(x) - \alpha n_s(x) \\ &= W(x)\left[(1+\delta_g(x)) \times (1 + w_{blind}(x))\right] - W(x) \\ &= W(x)\delta_g(x) + W(x) w_{blind}(x), \end{aligned}$$

where $\delta_g(x) \times w_{blind}(x)$ is neglected since it is a term of order 2 in δ_g.

This assumption can be avoided by weighting the randoms instead of the data with

$$w^r{}_{blind}(\mathbf{x}) = 1 - w_{blind}(\mathbf{x}). \tag{5.54}$$

The FKP field becomes then,

$$\begin{aligned} F(x) &= W(x)\,(1+\delta_g(x)) - W(x) \times (1 - w_{blind}(x)) \\ &= W(x)\delta_g(x) + W(x)w_{blind}(x), \end{aligned}$$

see Section 5.3.2.1 for a comparison between applying weights to the data and to the randoms.

Remark: The expected value of n_g will be unchanged with the blinding scheme applied in the data. Indeed, the expected value reads:

$$\begin{aligned} \langle n_g(x)\rangle &= W(x)\langle 1 + \delta_g(x) + w_{blind}(x) + \delta_g(x)w_{blind}(x)\rangle \\ &\simeq W(x)\langle 1 + \delta_g(x) + w_{blind}(x)\rangle \\ &= W(x) = \alpha n_s(x), \end{aligned}$$

where, again, $\delta_g(x)w_{blind}(x)$ was neglected. A simple way to test this approximation is to compute the change in $\langle n_g(x)\rangle$, which just corresponds to the change in mean data weights. This change is very small. For instance, the data weights with a blinding scheme are displayed in Fig. 5.24 and the change in the mean is about 10^{-4}.

5.3.1.3 Shot noise correction

With the blinding weights, the new density field is $\delta^{s'}(\mathbf{k}) = \delta^s(\mathbf{k}) + a(k)\widehat{\delta}^r(\mathbf{k})$. δ^s is the density field in redshift space and $\widehat{\delta}^r$ is the estimation of the reconstructed field in real space. The associated power spectrum is:

$$\begin{aligned} \langle|\delta^{s'}(\mathbf{k})|^2\rangle = &\underbrace{\langle|\delta^s(\mathbf{k})|^2\rangle}_{=(b+f\mu^2)^2P_{lin}(k)+1/n} + 2a(k)Re(\underbrace{\langle\delta^s(\mathbf{k})\widehat{\delta}^{r\star}(\mathbf{k})\rangle}_{\equiv\widetilde{X}(\mathbf{k})}) + a(k)^2\underbrace{\langle|\widehat{\delta}^r(\mathbf{k})|^2\rangle}_{\equiv\widetilde{Y}(\mathbf{k})}, \end{aligned} \tag{5.55}$$

with the low-k limits:

$$\begin{cases} \widetilde{X}(\mathbf{k}) = b(b+f\mu^2)W^{\rm tot}(k)P_{lin}(k) + \dfrac{1}{n}W^{\rm tot}(k)\exp\left(-\frac{1}{2}k^2\mu^2f^2\sigma_d^2\right) \\ \widetilde{Y}(\mathbf{k}) = b^2W^{\rm tot}(k)^2P_{lin}(k) + W^{\rm tot}(k)^2\dfrac{1}{n} \end{cases}, \tag{5.56}$$

Some brief remarks on the above equations:

- The shot-noise can be supposed similar between the redshift space distribution and the real space distribution since the size of fluctuations of $n(x)$ are much smaller than the displacement field.
- The exponential term in $\widetilde{X}(k)$ is from section 5.2 of Smith et al. (2021), where the variance of the displacement σ_d^2 is defined as

  $$\sigma_d^2 = \frac{1}{6\pi}\int \mathrm{d}k P_{lin}(k). \tag{5.57}$$

  For the same reason as the window $W^{\rm tot}(k)$ appears in 5.53, the displacement field ψ is also computed from the smoothed density field and σ_d must be rescaled as (with only one CIC-reading)

  $$\sigma_d^2 = \frac{1}{6\pi}\int \mathrm{d}k (W^sW^{cic})^2(k)P_{lin}(k). \tag{5.58}$$

- The shot noise terms in (5.56) are undesired but unfortunately cannot be removed. The shot noise evolves with the survey selection function, so will depend on the redshift. The power spectrum computed in different redshift slices would lead to different apparent $f_{\rm NL}^{\rm loc}$. It is therefore crucial to correct this shot noise effect.

Let us find $a'(\mathbf{k})$ to be used instead of $a(k)$ in (5.51) to recover the expected value of the additional PNG power without the undesired shot noise term. Hence, we want

$$2a'(\mathbf{k})\widetilde{X}(\mathbf{k}) + a'^2(\mathbf{k})\widetilde{Y}(\mathbf{k}) = 2a(k)X(\mathbf{k}) + a^2(k)Y(\mathbf{k}) \equiv A(\mathbf{k}), \tag{5.59}$$

where $X(\mathbf{k})$ (*resp.* $Y(\mathbf{k})$) corresponds to the non shot noise part ($1/n \to 0$) of $\widetilde{X}(\mathbf{k})$ (*resp.* $\widetilde{Y}(\mathbf{k})$). Solving the above second-degree polynomial gives:

$$a'(\mathbf{k}, f_{\rm NL}^{blind}, n) = \frac{-\widetilde{X}(\mathbf{k}, n) \pm \sqrt{\widetilde{X}(\mathbf{k}, n)^2 + \widetilde{Y}(\mathbf{k}, n)A(\mathbf{k})}}{\widetilde{Y}(\mathbf{k}, n)} \tag{5.60}$$

$\widetilde{X}(\mathbf{k}, n)$ and $\widetilde{Y}(\mathbf{k}, n)$ depend on both $\mathbf{k}$ and $\mathbf{x}$ via the shot noise term. To ensure that the amplitude of the PNG signal in the power spectrum does not depend on the local survey selection function, we want to retain the dependence on $\mathbf{x}$ and then, for simplicity, remove the dependence on $\mathbf{k}$. We, therefore, choose a value of $\mathbf{k}$ defined as (k_{pivot}, μ_{pivot}) where the equality will be respected and normalize the currently proposed weights:

$$w_{blind}(\mathbf{k}) = \underbrace{\frac{a(k_{pivot}, \mu_{pivot}, n)}{a(k_{pivot}, \mu_{pivot})}}_{\text{shot noise correction factor}} \times a(k) \times \widehat{\delta}^r(\mathbf{k}) \tag{5.61}$$

5.3.1.4 Positive or negative solution?

From (5.50) with $f_{\rm NL}^{\rm loc} > 0$, it is clear that we need to choose the positive solution in (5.60) requiring to use $+$. For $f_{\rm NL}^{\rm loc} < 0$, we choose also the $+$ solution in (5.60) motivated by the fact that we want a corrective factor (a'/a) which should be $\in [0, 1]$. This ratio is plotted in Fig. 5.17 for different values of $f_{\rm NL}^{\rm loc}$ and for different shot noises, where the $+$ solutions are in blue/violet whereas the $-$ solutions are in orange/red.

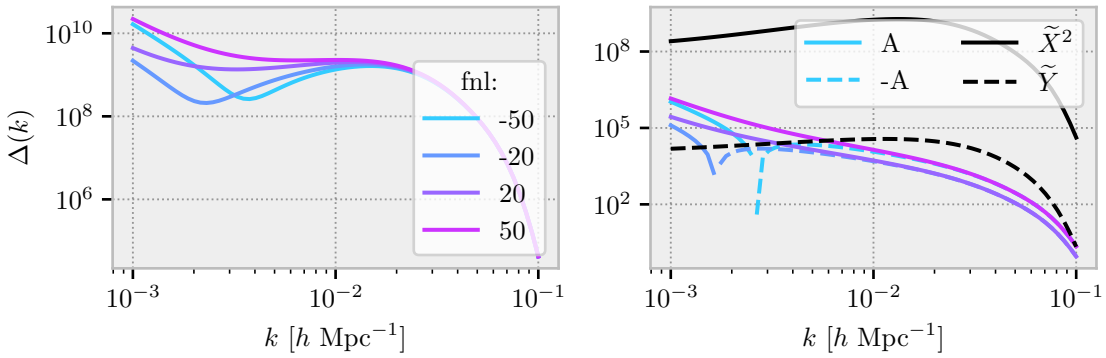
To avoid any troubles with the behavior where $f_{\rm NL}^{\rm loc}$ is negative ($b + \Delta b$ can be zero), we choose a value of $k_{pivot} = 8 \cdot 10^{-3}\ h\ \mathrm{Mpc}^{-1}$. For $f_{\rm NL}^{\rm loc} > 0$ we choose a lower value in order to de-bias the PNG signal as best as we can, fixing: $k_{pivot} = 4 \cdot 10^{-3}\ h\ \mathrm{Mpc}^{-1}$. In all cases, we use $\mu_{pivot} = 0.6$. The impact of the value of k_{pivot} is shown in Fig 5.18.

5.3.1.5 Shot noise with data weights

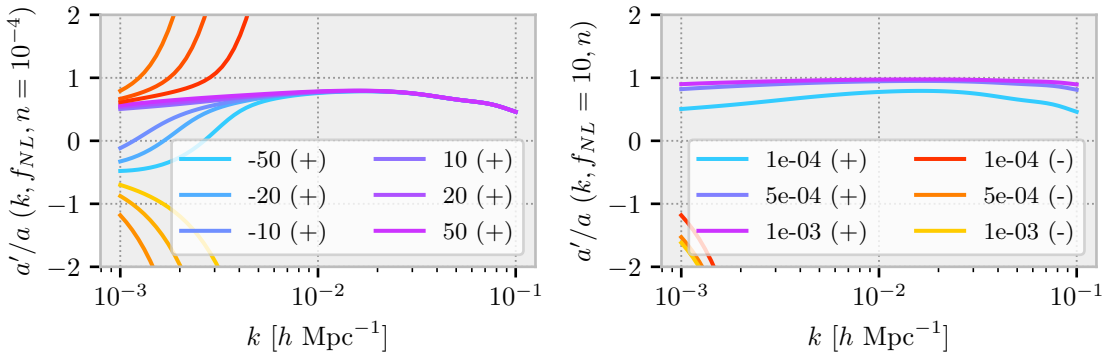
For the shot noise correction to work, we need a decent estimate of the shot noise. We suggest using the same estimation as used in the power spectrum estimation, that is, in each cell of volume dV:

$$N = \frac{\sum_{dV} w_g^2}{\alpha(\sum_{dV} w_g)(\sum_{dV} w_s)/dV} \tag{5.62}$$

The denominator is an estimation of the integral $\int d^3x \bar{n}^2(x)$. We use $\alpha(\sum_{dV} w_g)(\sum_{dV} w_s)$ instead of $(\sum_{dV} w_g)^2$ as the latter estimate is biased due to shot noise, which the former avoids in the assumption that data and randoms are uncorrelated.



(a)



(b)

Figure 5.17: (a) Left: Reduced discriminant $\Delta(\mathbf{k}) = \widetilde{X}(\mathbf{k}, n)^2 + \widetilde{Y}(\mathbf{k}, n)A(\mathbf{k})$ is always positive. (5.59) has always two real solutions. Right: Different terms from which the reduced discriminant is computed. In the two cases, the shot noise is $1/n = 10^4$ Mpc^3 h^{-3}. (b) Left: Corrective factor for different values of $f_{\rm NL}^{\rm loc}$ with $n = 10^{-4}$ Mpc^{-3} h^3. The + solutions are in blue and the − ones are in orange. For $f_{\rm NL}^{\rm loc} < 0$, we choose the + solution since we will fix the value of k_{pivot} to $8 \cdot 10^{-3}$ h Mpc^{-1} (see Fig. 5.18). Right: Similar to left but for different values of the shot noise with $f_{\rm NL}^{\rm loc} = 10$.

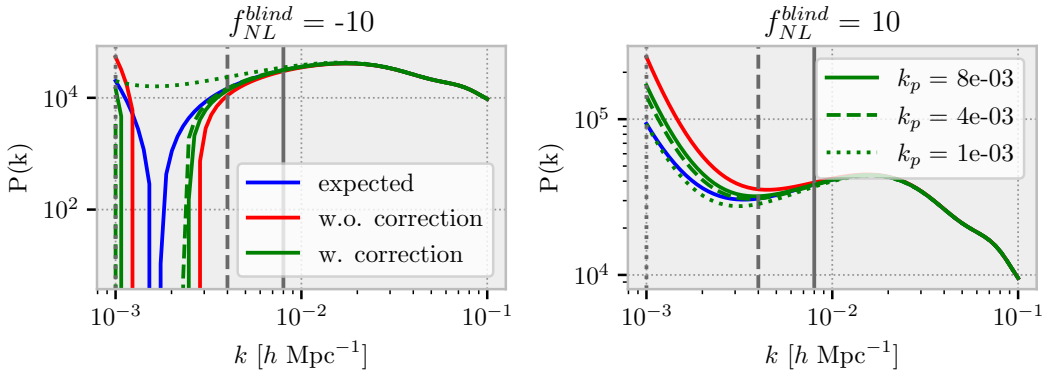


Figure 5.18: Impact of the value of k_{pivot} for $f_{\rm NL}^{blind} = -10$ (left) and for $f_{\rm NL}^{blind} = 10$ (right).

5.3.2 Validation with mocks

In the following, most of the tests are performed on the fastpm cubic box with 5.5 Gpc h^{-1} size presented in Section 5.2.1, while the real DESI cutsky Y1 (SGC) will be tested in Section 5.3.2.2.

5.3.2.1 Mocks vs. model

► **Mocks without shot noise corrective factor** Let us first try to apply the blinding scheme without adding the shot noise correction factor. Fig. 5.19 shows the power spectra blinded with different values of $f_{\rm NL}^{\rm loc}$. The blue dashed lines are the expected power spectra for the corresponding $f_{\rm NL}^{blind}$ values (with hand-tuned linear bias) without any contribution from the shot noise. This is what we want to reproduce. In the left panel, the blinding scheme is applied on a high-density sample where the shot noise is almost negligible. The blinding scheme recovers correctly the desired blind value. In the right, the blinding scheme is applied on a low-density sample and the red full lines are the predicted models from (5.55) (i.e.) including the shot noise. Thus, the predicted model agrees with the measured power spectra. As described above, the blinding weights without the shot noise correction factor yield a larger PNG signal than desired.

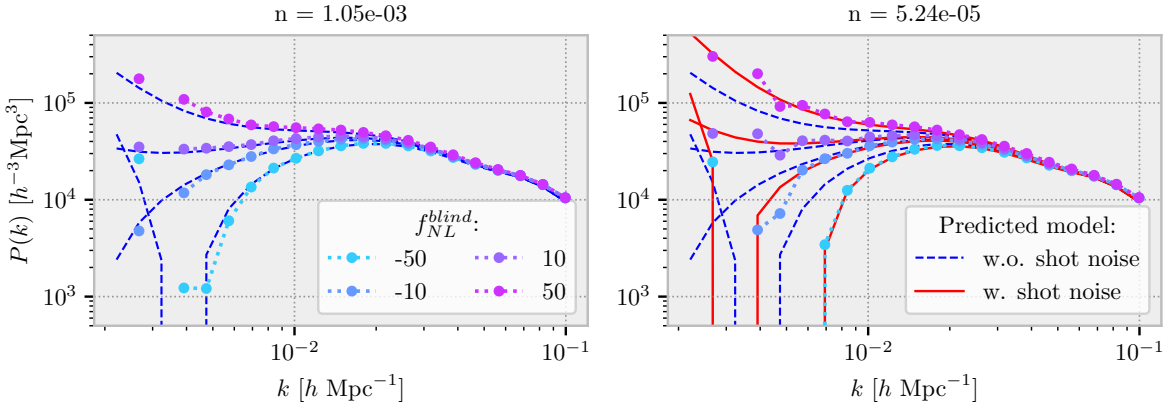


Figure 5.19: Blinding scheme **without the shot noise correction factor**, where the contribution of the shot noise is added to the prediction. Dots are the measured power spectrum monopoles. Left (*resp.* Right) is for high (*resp.* low) density sample. Dashed lines are the desired power spectra with the corresponding values of $f_{\rm NL}^{\rm loc}$ without shot noise contribution and the red ones are the prediction from (5.55) including the shot noise contribution.

► **Mocks with shot noise corrective factor** Let us test the final blinding scheme including the shot noise correction factor. The blinding power spectra are shown in Fig. 5.20 for high and low-density samples. The predicted models do not have any shot noise contribution and are exactly what we want to achieve. The shot noise correction seems to reproduce the desired signal up to the typical scale of the fitting range $k_{min} \sim 4 \cdot 10^{-3}$ h Mpc^{-1}. The shot noise correction factor is required especially for the low-density samples to recover the desired PNG signal.

Let us try to measure $f_{\rm NL}^{\rm loc}$ from these blinded power spectra. The fits are performed with `iminuit`, and posterior samples are drawn with `zeus`[26] wrapped in the `desilike` framework. The power spectrum monopole is fit with a damped Kaiser model, with the usual parametrization[27] $(f_{\rm NL}^{\rm loc}, b_1, s_n, \Sigma_s)$ (5.24), where we fix b_ϕ with the universal mass function assumption (2.119). The power spectrum covariance is computed analytically in the Gaussian approximation, without the window effect.

The fits for $f_{\rm NL}^{\rm loc}$ and b_1 with the associated errors are presented in Fig. 5.21. The high-density sample with $n = 1.05 \cdot 10^{-3}$ Mpc^{-3} h^3 (called in the figures "full sample") is in blue.

[26] https://github.com/minaskar/zeus
[27] https://github.com/echaussidon/desilike/blob/main/nb/png_examples.ipynb

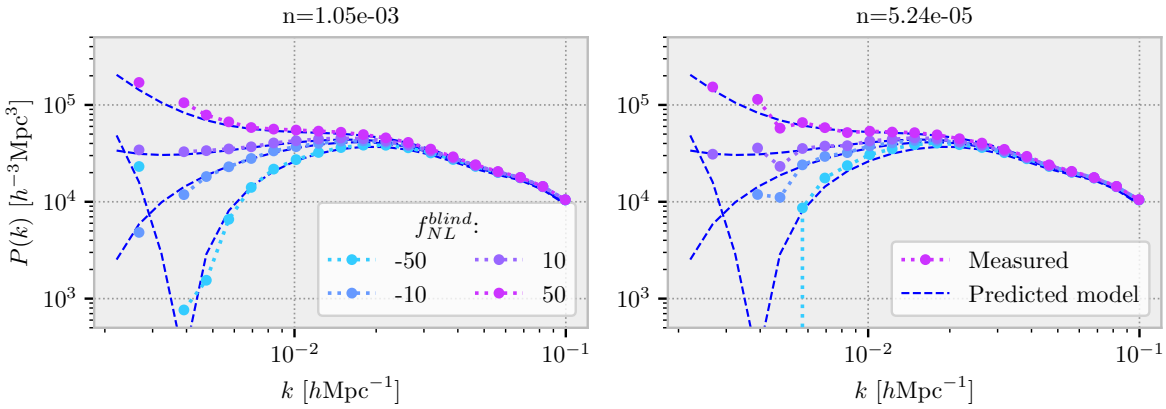


Figure 5.20: Blinding scheme **with the shot noise correction factor**, where the predicted models **do not have** the shot noise contribution. On the left, it is for a very dense sample where the shot noise is almost negligible in the scales of interest. Whereas, on the right, it is for low-density samples, typically DESI QSO at high redshift, where the shot noise matters. The blinding scheme with the shot noise correction recovers the desired power spectra.

The subsample with $n = 1.05 \cdot 10^{-4}$ Mpc^{-3} h^3 (called 10× subsample) is in red. The low-density sample with $n = 5.24 \cdot 10^{-5}$ Mpc^{-3} h^3 (called 20× subsample) is in green or in grey in the posterior plots.

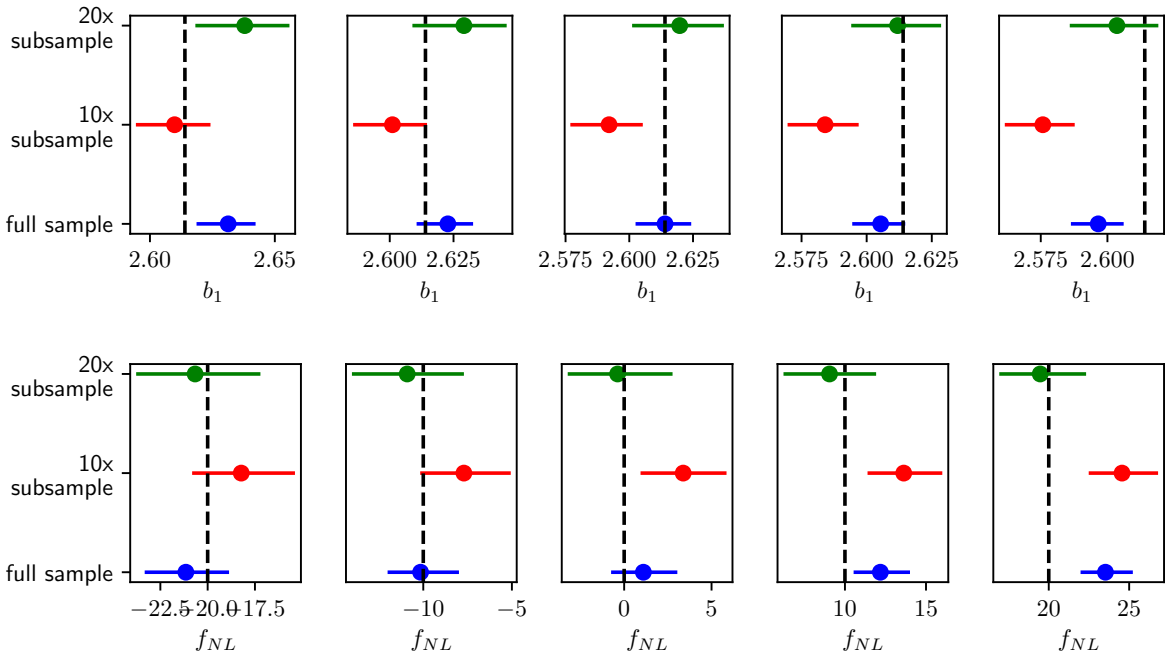


Figure 5.21: b_1 and $f_{\rm NL}^{loc}$ best fit values for the non-blinded case ($f_{\rm NL}^{loc} = 0$) and different input values of $f_{\rm NL}^{blind} \in \{-20, -10, 10, 20\}$, with 1σ posterior errors. Black dashed lines are the expected value of the linear bias and the expected blinded value $f_{\rm NL}$.

► **Blinded range of scales** Fig. 5.22 shows the ratio between the blinded and the non-blinded power spectrum monopoles. As explained in Section 5.3.1.1, the scale-dependent bias created in the blinding scheme is $W^{\rm tot}(k)\Delta b(k)$: the measured 'high' k PNG signal is damped

with the window function $W^{\rm tot}(k)$.

The dashed lines are the theoretical prediction without $W^{\rm tot}(k)$ and the black dotted lines are the theoretical predictions with only the $W^{\rm tot}(k)$ order 1 term[28] (leading the small discrepancy at very high k).

Yet, this $W^{\rm tot}(k)$ term is not included in the fitted model, thereby explaining the different bias values displayed in Fig. 5.21. Since we recover the correct (at least in the expected error bars) value of $f_{\rm NL}^{\rm loc}$, this is not a problem, just something to be aware of.

Important: With this window function the impact of the $f_{\rm NL}^{\rm loc}$ blinding is then negligible beyond $k \sim 0.06\ h\ {\rm Mpc}^{-1}$ and should not impact BAO or RSD measurements.

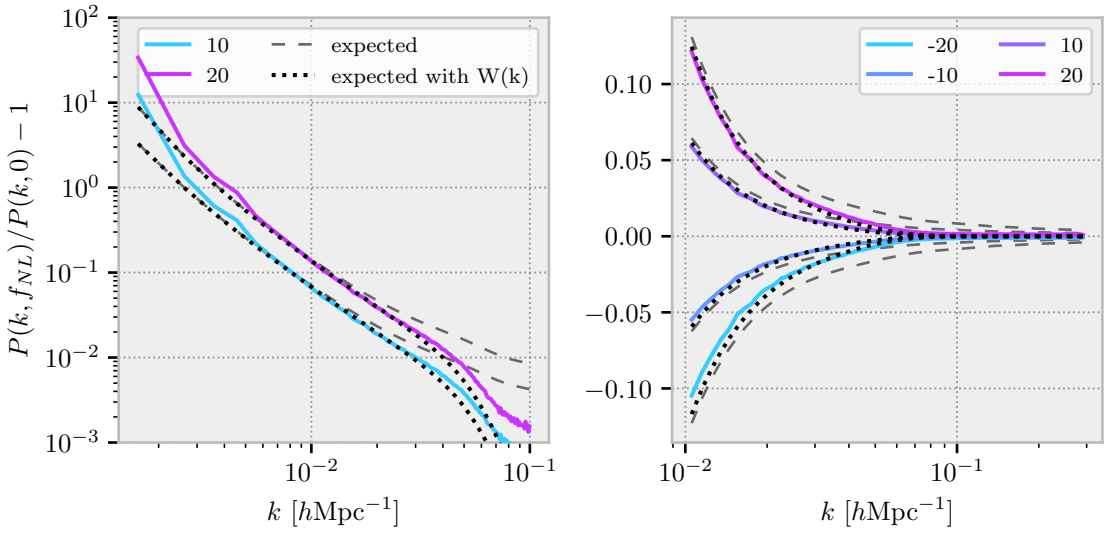


Figure 5.22: Both figures represent the ratio between the blinded power spectrum and the non-blinded power spectrum for different values of $f_{\rm NL}^{\rm loc}$. Dashed (*resp.* dotted) lines are the model predictions without (*resp.* with) the window function $W^{\rm tot}(k)$ from the reconstruction smoothing and particle assignment. The window function damps the PNG signal, such that it goes to zero faster than the theory.

► **Weights applied on data or randoms?** We can either compute the weights for the data: $1 + w_{blind}(\mathbf{x})$ or for the randoms: $1 - w_{blind}(\mathbf{x})$. The difference between the two cases is negligible, as shown in Fig. 5.23 for the 20× subsample.

It is more convenient to add weights to the data, to keep the definition of randoms as sampling the survey selection function (without – or as less as possible – correlation with the data).

[28] (ie) $P(k) = \left(b^2 + 2W^{\rm tot}(k)b\Delta b + W^{\rm tot}(k)\Delta b^2\right) P_{lin}(k)$

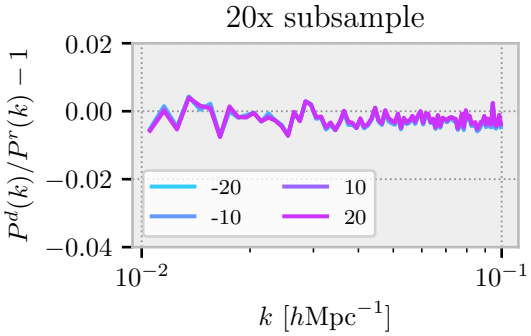


Figure 5.23: Ratio between the measured power spectrum after the blinding scheme with weights applied to the data and with the weights applied to the randoms.

► **Distribution of blinding weights** Let us compute the expected standard deviation of the proposed weights (5.50) and (5.61):

$$\begin{aligned}
\sigma^2_{w^d_{blind}} &= \langle (1 + w_{blind} - 1)^2 \rangle \\
&= \langle w_{blind}(\mathbf{x}) w_{blind}(\mathbf{x}) \rangle \\
&= \frac{1}{(2\pi)^6} \int \mathrm{d}\mathbf{k} \int \mathrm{d}\mathbf{q}\ e^{i(\mathbf{k}+\mathbf{q})\mathbf{x}} \langle w_{blind}(\mathbf{k}) w_{blind}(\mathbf{q}) \rangle \\
&= \frac{1}{(2\pi)^6} \int \mathrm{d}\mathbf{k} \int \mathrm{d}\mathbf{q}\ e^{i(\mathbf{k}+\mathbf{q})\mathbf{x}} a(k,n) a(q,n) \langle \widehat{\delta}^r(\mathbf{k}) \widehat{\delta}^r(\mathbf{q}) \rangle \\
&= \frac{1}{(2\pi)^6} \int \mathrm{d}\mathbf{k} \int \mathrm{d}\mathbf{q}\ e^{i(\mathbf{k}+\mathbf{q})\mathbf{x}} a(k,n) a(q,n) (2\pi)^3 \delta_D(\mathbf{k}+\mathbf{q}) P_{\widehat{\delta}^r}(k) \\
&= \frac{1}{(2\pi)^3} \int \mathrm{d}\mathbf{k}\ a(k,n)^2 W(k)^2 \left(b^2 P_{lin}(k) + 1/n \right) \\
&= \frac{1}{2\pi^2} \int \mathrm{d}k\ k^2 a(k,n)^2 W(k)^2 \left(b^2 P_{lin}(k) + 1/n \right)
\end{aligned} \tag{5.63}$$

Remark: The simulation box size is of finite size, so we have to integrate only from $k_{min} = k_f = 2\pi/L$ with L the box size to $k_{max} = k_N$ the Nyquist frequency. In fact, k_{max} does not really matter since the integrand goes to 0 really quickly for high k. In principle, the best would be to perform the above integral on the $\mathbf{k}$-grid actually used in the blinding routines.

Fig. 5.24 gives the integrand (left) for two values of $f_{\rm NL}^{loc} \in \{-10, 20\}$ and for two values of $n \in \{10^{-4}, 5 \cdot 10^{-5}\}$ Mpc^{-3} h^3. The integrand peaks at low k, such that the minimum scale used in the integral indeed has great importance. The two other panels show the dispersion of the blinding weights for the corresponding values of $f_{\rm NL}^{\rm loc}$ for the different samples used above full sample, 10× subsample and 20× subsample.

Remark: Importantly, the terms of the weights are very small, less than 10^{-2}, such that they can discreetly be mixed with other weights, e.g. photometric systematic weights.

► **Sanity check with hexadecapole** The hexadecapole ($\ell = 4$) in the linear regime is $P_4^{(z)}(k) = \frac{8}{35} f^2 P_{lin}(k)$, hence does not depend on the bias $b + \Delta b(f_{\rm NL}^{loc}, k)$ and then on $f_{\rm NL}^{\rm loc}$. The hexadecapole, for multiple values of $f_{\rm NL}^{\rm loc}$, is shown in Fig. 5.25. It is not impacted by the blinding weights as expected.

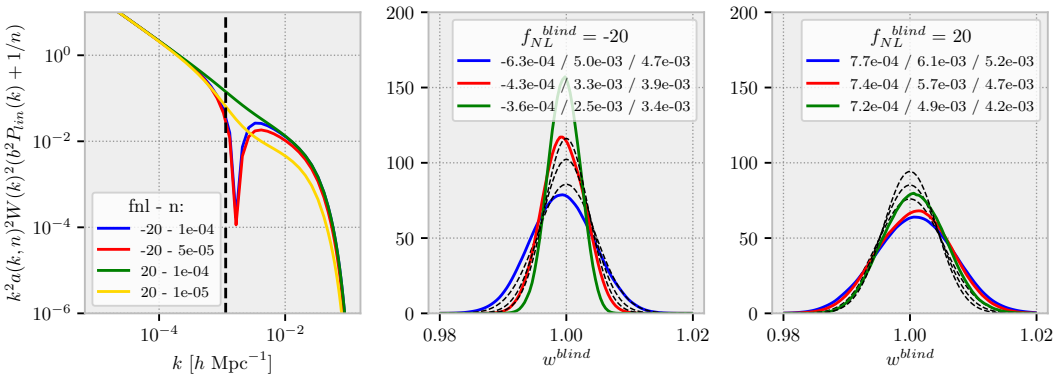


Figure 5.24: Left panel: integrand of (5.63). Middle/Right panels: distribution of the blinding weights in colored lines, and Gaussian with dispersion predicted with (5.63) in dashed black lines. The RMS is given in the legend (measured/computed).

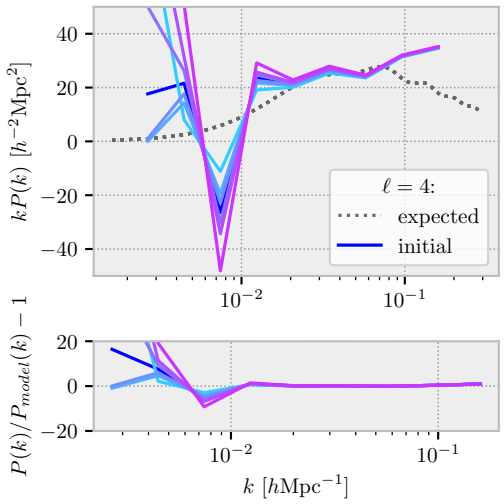


Figure 5.25: Hexadecapole for multiple values of $f_{\mathrm{NL}}^{blind} \in \{-50, -20, -10, 10, 20, 50\}$ for the high-density sample $n = 1.05 \cdot 10^{-3}$ Mpc^{-3} h^3.

5.3.2.2 Blinding with a misspecified survey selection function (remaining systematics)

In order to test the blinding procedure in realistic conditions, we use the 16 subsamples matching the DESI QSO Y1 SGC with the expected imaging contamination from the current Y1 data, described in Section 5.2.1. On each subsample, contaminated by the imaging systematics, we apply the blinding scheme, and then we correct the imaging systematics either with the inverse of the contamination or with `regressis`[29] (Chaussidon et al. 2022). We then measure the output value of $f_{\mathrm{NL}}^{\mathrm{loc}}$, hoping to recover the input blinding value.

To reduce sampling noise, we use the mean of the 16 realizations. The measured power spectrum for the different cases is shown in Fig. 5.26. The errors are from 1000 EZ mocks matching the DESI QSO Y1 SGC sample **not** divided by 16.

$f_{\mathrm{NL}}^{\mathrm{loc}}$ posteriors, for the mean of the 16 subsamples, are displayed in Fig. 5.27. For the Y1 case, we use the covariance matrix from the Y1 footprint instead of the one from the SGC

[29] https://github.com/echaussidon/regressis

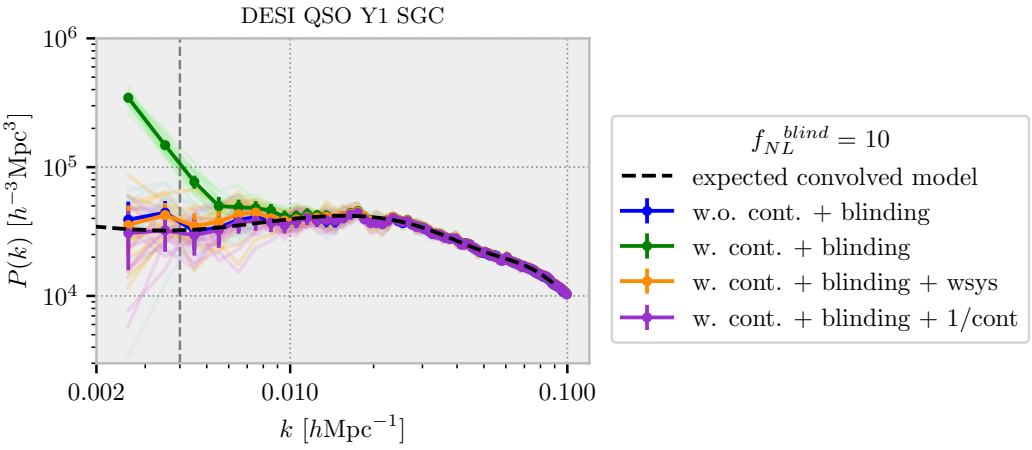
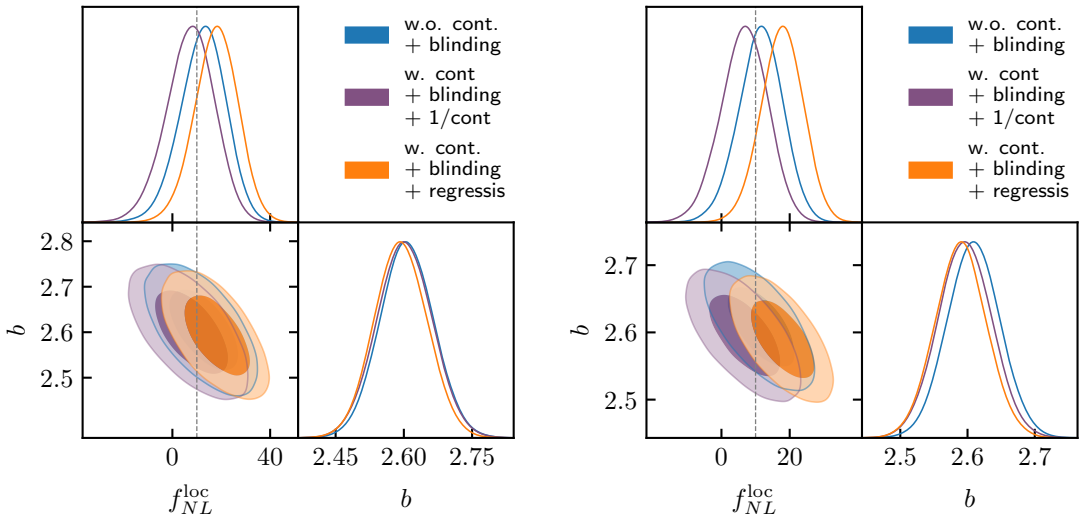


Figure 5.26: Mean of 16 subsamples mimicking the DESI QSO Y1 SGC sample. The errors are from the standard deviation of 1000 EZ mocks. The dashed vertical line is about the minimum scale used in the fit. The blue line is the power spectrum measured in simulations without imaging systematics and with blinding. Green is for contaminated simulations with blinding. Orange (*resp.* violet) is for contaminated simulations with blinded and correction with `regressis` (*resp.* inverse of the contamination).

footprint, just to reduce the errors. We recover the input blinding value $f_{\mathrm{NL}}^{blind} = 10$ well within the Y1 uncertainty both if no contamination is applied and including the imaging systematics, with the exact inverse correction. In the latter case, we obtain a slightly lower value of $f_{\mathrm{NL}}^{\mathrm{loc}}$. The correction with `regressis` is a bit farther off (1.4σ), but no tuning has been performed yet.



(a) SGC footprint

(b) Y1 (SGC + NGC) footprint

Figure 5.27: We almost recover the input blinding value: $f_{\mathrm{NL}}^{blind} = 10$ in the three cases. Blue contours correspond to the blinding applied on the uncontaminated sample. Violet (*resp.* orange) contours correspond to blinding applied on contaminated sample and with the inverse of the contamination (*resp.* with `regressis`). (a) uses the SGC covariance matrix, (b) the Y1 covariance matrix (notice the change in uncertainties).

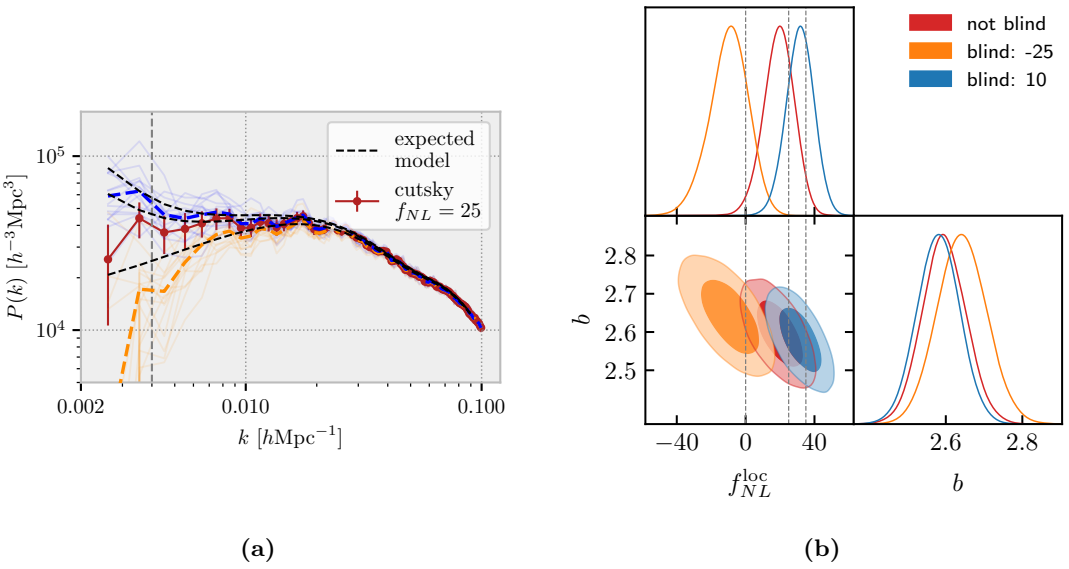
5.3.2.3 Blinding with non-zero $f_{\rm NL}^{\rm loc}$ simulations

Although Planck constraints are compatible with $f_{\rm NL}^{\rm loc} = 0$ (Planck Collaboration et al. 2020), we check how the blinding pipeline responds to an initial non-zero value of $f_{\rm NL}^{\rm loc}$. We use similar fastpm simulations to the one with $f_{\rm NL}^{\rm loc} = 0$ but with $f_{\rm NL}^{\rm loc} = 25$, creating the same 16 subsamples which match the DESI QSO Y1 SGC sample.

Fig. 5.28 shows the measured Y1 SGC power spectra for the non-blinded case, and blinding values of $f_{\rm NL}^{\rm loc} \in \{-25, 10\}$. The best-fit values are given in Table 5.5. The errors are smaller than those we'll obtain in the next section because we have a larger bias and use p=1. Up to the scale of interest ($k_{min} = 4 \cdot 10^{-3}\ h\ \mathrm{Mpc}^{-1}$), we mostly recover the expected signal. Yet, a discrepancy between the model and the measured power spectrum is already observable in the non-blinded case, see Section 5.2.2.2, which propagates to the blinded cases.

Table 5.5: Best fit values for non-blinded simulation with $f_{\rm NL}^{\rm loc} = 25$ and for blinded cases (+10, -25)

	non-blinded	blinded + 10	blinded - 25
$f_{\rm NL}^{\rm loc}$	19^{+9}_{-9}	31^{+9}_{-8}	-9^{+12}_{-10}



(a) **(b)**

Figure 5.28: (a) The brick line is the measured power spectrum without blinding for the Y1 SGC footprint, based on fastpm simulations with $f_{\rm NL}^{\rm loc} = 25$, with error bars from 1000 EZ mocks. Orange and blue lines show the power spectra blinded with $f_{\rm NL}^{\rm loc} \in \{-25, 10\}$ respectively. The dashed vertical line is about the minimum scale used in the fit. (b) Posteriors for the three power spectra in (a). The posterior of the fastpm (non-blinded) power spectrum is slightly shifted to a lower $f_{\rm NL}^{\rm loc}$ value than input, which propagates to the blinded cases.

Remark: For the purposes of this dissertation, we are only interested in analyzing the large scales of the power spectrum, and we will not be looking at the scales at which the BAO peak and RSD are measured. We, therefore, don't need to look at the impact of this blinding scheme on the other scales of the power spectrum. However, we have verified (not shown here) that the addition of this blinding does not alter the measurement of the BAO and RSD parameters at the statistical accuracy of Y1 measurements, nor does it alter the blinding setup for BAO and RSD.

As a result, this blinding was introduced into DESI's clustering analysis pipeline, and since

simple forecast predicts a sensitivity on $f_{\rm NL}^{\rm loc}$ about 15 with the Y1 data, we draw a random value for the blinding in $[-15, 15]$.

5.4 Preliminary analysis: PNG with DESI QSO Y1 data

The blinding scheme for BAO, RSD, and $f_{\rm NL}^{\rm loc}$ were effective at the end of May 2022, and the analysis on real data could start. As explained above, we have blind the data with $f_{\rm NL}^{blind} \in [-15, 15]$.

5.4.1 Clustering catalog validation

A preliminary description of the clustering catalogs is available in DESI Collaboration et al. (2023b). The case of quasars is the easiest. For the clustering catalog, we consider only the quasar targets classified as a quasar with $0.8 < z < 3.5$ by the pipeline presented in Fig. 4.11. In addition, we mask the region with imaging maskbits 7, 8 (bright star masks from WISE) and 11 (star mask from Gaia), although the need for these additional masks has not yet been studied.

The density of the quasars used for the clustering is given in Table 5.2, and their redshift distribution is displayed in Fig.5.8. The overall imaging systematic study is presented in Section 4.4.2, and the summary plots are given in Fig. 4.43 for the North and South (NGC) and Fig. 4.44 for the South (SGC) and DES.

5.4.1.1 Analysis in redshift bins

As described in Section 4.2.3.3, at low z, the target selection depends notably on the definition of the stellar morphology ('PSF'), see Fig. 4.13. However, the morphology of the objects is determined by the images of the legacy survey such that a better image quality means better object resolution and, in the case of quasars, precise determination of the extended galaxy host in addition to the quasar at its center. Hence, one can expect different imaging effects depending on the redshift.

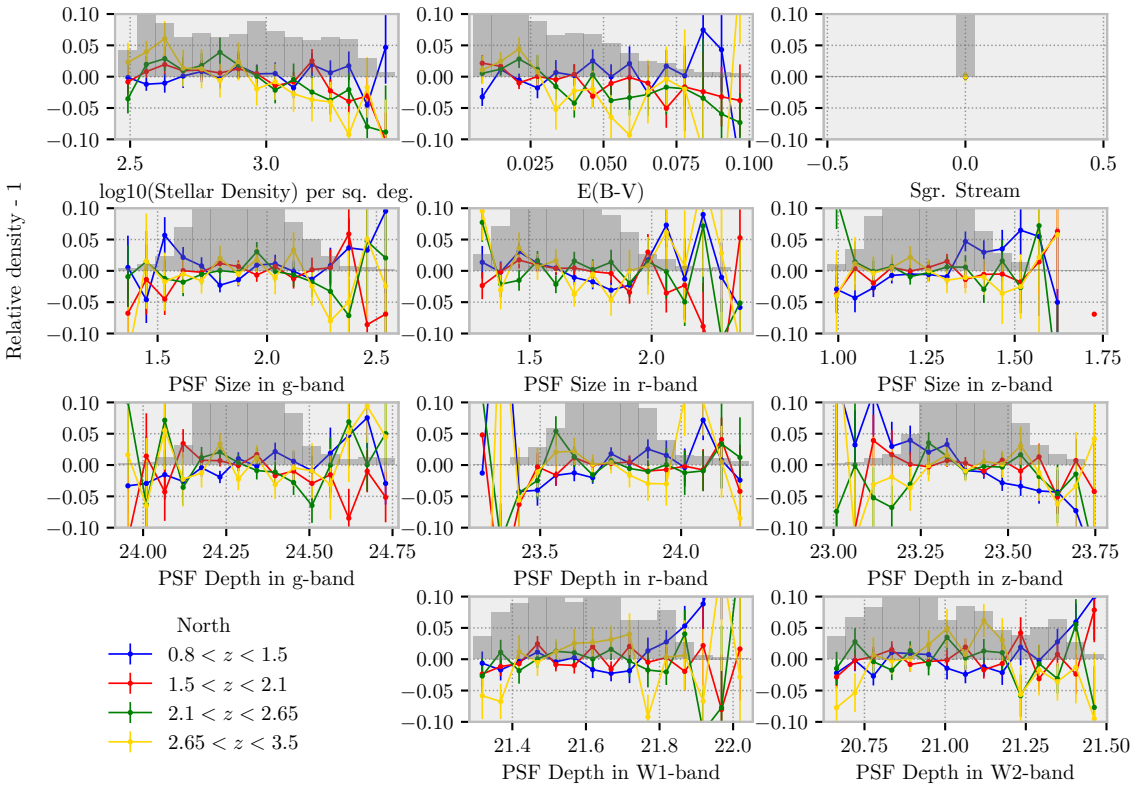
The relative QSO density in the North and South, as a function of each observational parameter for different redshift bins, is given in Fig. 5.29. Here, and in the following, we have not displayed the figure for DES due to the lack of statistics. One can clearly identify a different behavior between the low redshift sample ($0.8 < z < 1.5$) shown in blue and the higher redshift sample, particularly in the sub-panel for the PSF size and PSF depth of z.

This behavior is consistent with what we have seen using the Morphology criterion. When image quality is good[30], it is easier to observe the quasar and its host galaxy, classifying it as an extended source and not selecting it as a target quasar. The closer the objects are, the greater the effect. Conversely, when image quality is poor, it is impossible to detect extended sources, and most objects are considered point sources.

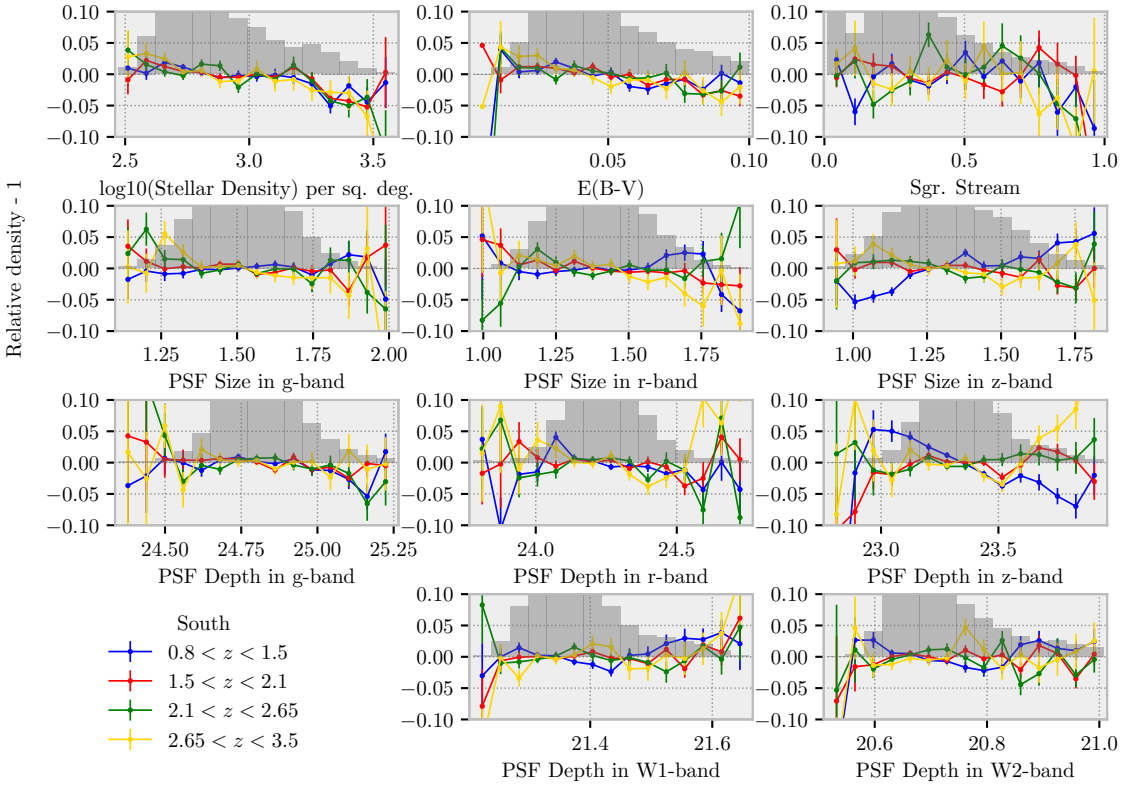
Morphology determination of the sources in the Legacy Imaging Surveys is done by combining the different optical bands. In particular, this effect should be accentuated in the z-band since z has the best quality image in the different regions, see Table 3.3.

One can zoom in the first redshift bin $0.8 < z < 1.5$, to understand the dependence of the effect as a function of the redshift. Thinner redshift bins for the South region are shown in

[30]Reminder: good quality is high PSF Depth and low PSF size.



(a) North



(b) South

Figure 5.29: Same as Fig. 4.28, but for the clustering catalog with completeness weights in different redshift bins. (a) is for North and (b) for South footprint.

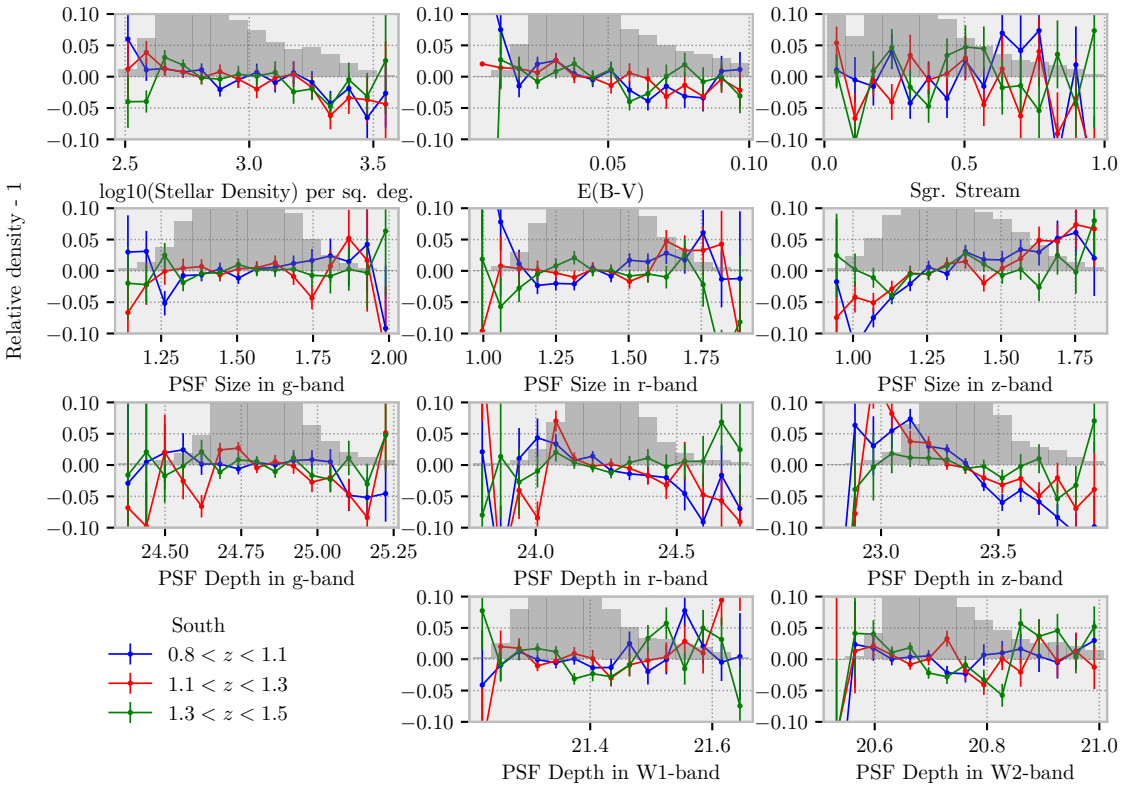


Figure 5.30: Same as Fig. 5.29, but for thinner redshift bins in the South region.

Fig. 5.30. The redshift bin $1.3 < z < 1.5$ seems to have similar trends to the higher redshift bins shown in Fig.5.29, in particular for the PSF size and PSF depth of z.

5.4.1.2 Two weights for the imaging systematics

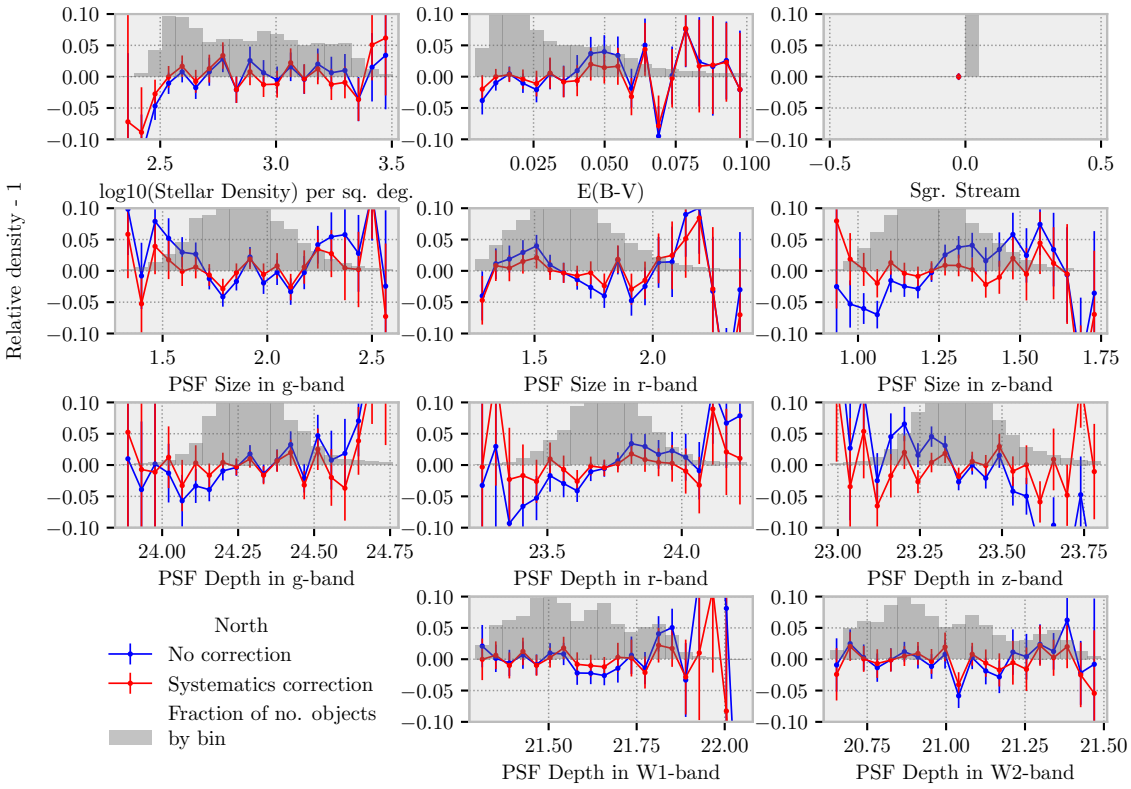
The two distinctive trends displayed in Fig.5.29 cannot be corrected with a single weight computed on the entire redshift range since no redshift information is currently used in `regressis`. The simple solution is then to generate two sets of weights independently. Based on Fig. 5.30, we decided to split the sample into two subsamples:

- low-z sample: $0.8 < z < 1.3$
- high-z sample: $1.3 < z < 3.5$

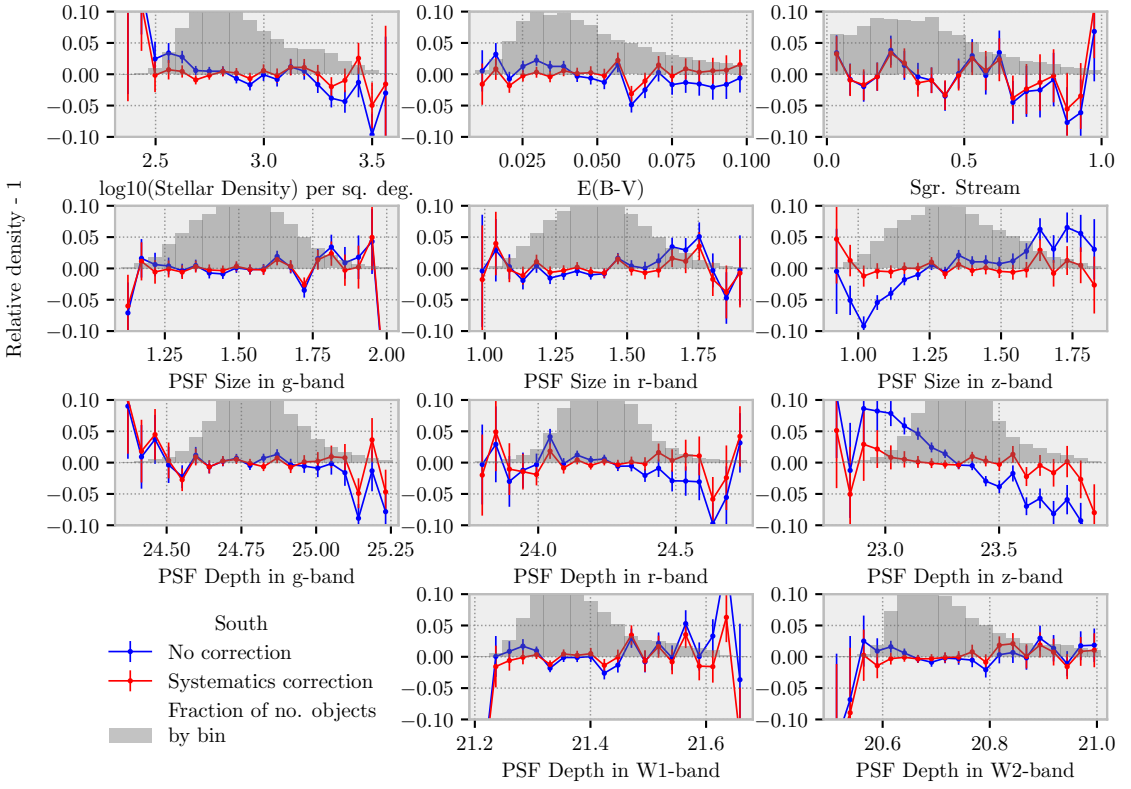
We used lower pixelization order to keep the number of objects sufficiently high in each pixel during the regression, from $N_{side} = 256$ to $N_{side} = 128$[31]. The impact of N_{side} was relatively small at large scales during our angular study in Section 4.3. However, this change must be tested more thoroughly for the power spectrum.

The efficiency of the correction is shown in Fig.5.31, where mainly the trends in PSF size and PSF depth z are perfectly corrected. Plots for the high-z sample correction are similar to Fig.4.28 and are well corrected.

[31] At $N_{side} = 256$ (*resp.* $N_{side} = 128$), the typical size of a pixel, at $z_{\rm eff} = 1.7$, is $\sqrt{\rm Area} = 4.8\ h^{-1}$Mpc (*resp.* $= 9.6\ h^{-1}$Mpc) corresponding to a Fourier scale $\sim 0.2\ h$Mpc^{-1} (*resp.* $\sim 0.1\ h$Mpc^{-1}) greater than the scales used for the scale-dependent bias measurement.



(a) North



(b) South

Figure 5.31: Same as Fig. 5.29, but blue is for the clustering catalog with completeness weights and $0.8 < z < 1.3$, and red are after the new correction computed specifically in this redshift range.

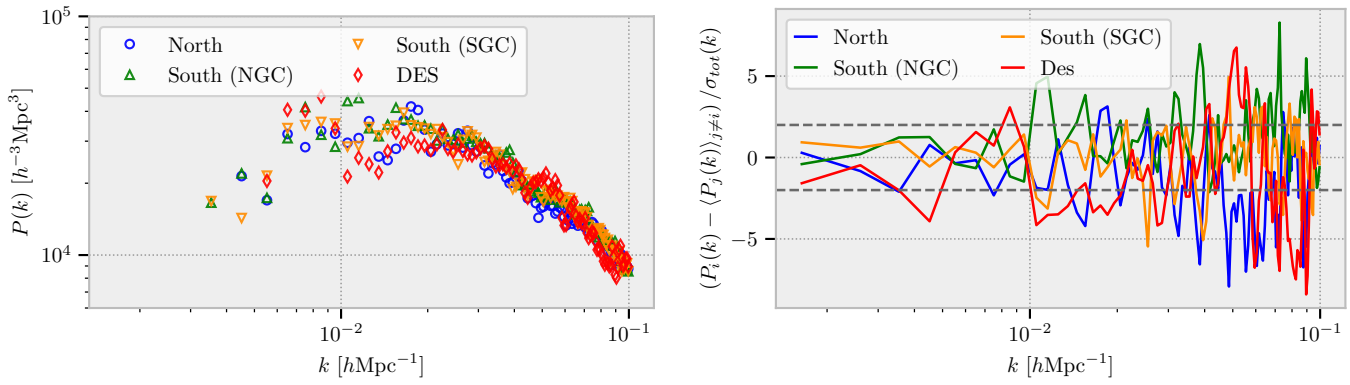
Remark: At present, the z=1.3 separation value has not been tested further. We can also imagine having more than two different weights by cutting the sample into more redshift slices to improve the correction.

5.4.1.3 QSO Y1 power spectrum measurement

Now that the imaging effects seem to have all been corrected correctly, we can compare the measured power spectrum with different sub-samples. Our aim here is not to find out whether the shape of the power spectrum is the expected one but to check the consistency between the different photometric regions.

▶ **Four photometric regions** The first step is to compare the power spectrum computed independently in the four photometric regions. The four power spectra are displayed in Fig. 5.32a. To test the compatibility between the four photometric regions, we compute the difference between the power spectrum measured in one region and the average over the effective coverage of the power spectra measured in the other regions, divided by the expected final errors, as displayed in Fig. 5.32b. The final errors take into account the modes between the different regions, and, therefore, are weaker than the errors using only the four regions independently. Due to the coverage, the power spectrum measured in DES is expected to be noisier. Note the excellent agreement between these regions, in particular at large scales.

As a result, we can measure the power spectrum over the entire NGC and SGC parts of the footprint simultaneously, using modes between North/South and South/Des, increasing our statistical precision.



(a) **(b)**

Figure 5.32: (a) Power spectrum measured in the four photometric regions using two weights to mitigate imaging systematics, one for low-z sample ($z < 1.3$) and one for the high-z sample ($z > 1.3$). (b) Compatibility test between the four photometric regions. The solid lines are the difference between the power spectrum measured in one region and the average over the effective coverage of the power spectra measured in the other regions, divided by the expected final errors. The final errors take into account the modes between the different regions. The four regions are in complete agreement, in particular at large scales.

For sanity check, one can compare the measured power spectrum with the two weights for the imaging systematic and the one with a single weight generated for the entire sample. Fig. 5.33 shows for the four photometric regions the power spectra with two weights (faceless marker) and with a single weight (full marker). Note the strong impact on large scales for the different photometric regions.

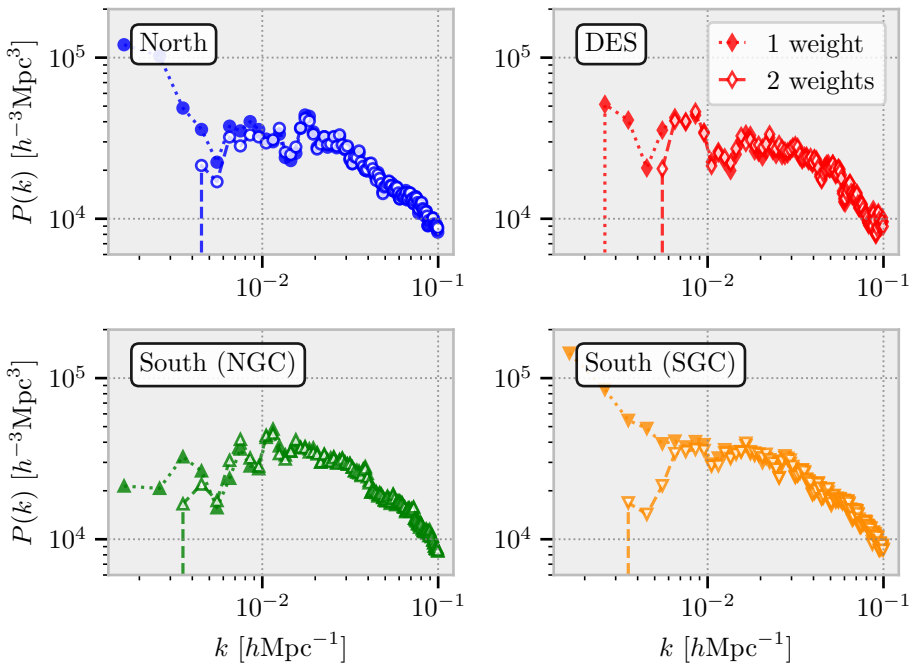


Figure 5.33: Impact of using two weights, one for low-z and one for high-z sample (faceless marker), versus a single weight generated with the entire sample (full marker).

► **Normalize South (SGC) to DES** In the current DESI pipeline, the South (NGC) is normalized to the North when the power spectrum is computed simultaneously on these two regions. Here, by normalized, we mean that the randoms redshift distribution is matched independently in the North and the South (NGC), and the randoms weights in the South (NGC) are multiplied by the normalization factor:

$$f_{\mathrm{norm}} = \frac{\sum_{i\in\mathrm{North}} w_{r,i}}{\sum_{i\in\mathrm{North}} w_{d,i}} \times \frac{\sum_{i\in\mathrm{South\ (NGC)}} w_{d,i}}{\sum_{i\in\mathrm{South\ (NGC)}} w_{r,i}}. \tag{5.64}$$

This normalization step takes into account the fact that tracer density and redshift distribution are not the same between the two photometric regions, see for instance the quasar case in Table 5.2 and in Fig. 5.8. This may be due to either a slightly different target selection in these regions, or because the photometry is deeper in one of the two regions and more targets are visible.

For the exact same reason, the South (SGC) region has to be normalized to the DES region by generating independently the randoms redshift and matching the weights between these two regions. For quasars, this is crucial, given the difference in density shown in Table 5.2 and on the redshift distribution shown in Fig. 5.8. Fig. 5.34 shows the result of this normalization. Without this normalization (brown), we measure an excess of power at large scales, when you consider objects in South (SGC) and DES simultaneously, compared to the separate region South (SGC) (orange) and DES (red) which will bias the measurement of primordial non-gaussianity. We recover an expected behavior using normalization (black).

► **Compare NGC vs. SGC** Since we can compute the power spectrum correctly in all the NGC and SGC, one can compare these two regions and finally combine the power spectrum from these two independent measurements. Fig. 5.35 shows the power spectrum for these two regions, showing again an excellent agreement.

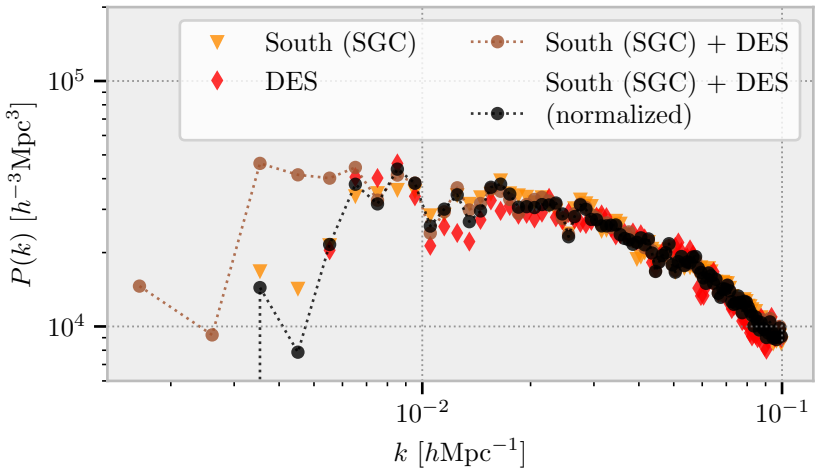


Figure 5.34: Comparison between South (SGC) (orange), DES (red), South (SGC) + DES where we do not use the normalization (brown) and when we use the normalization (black).

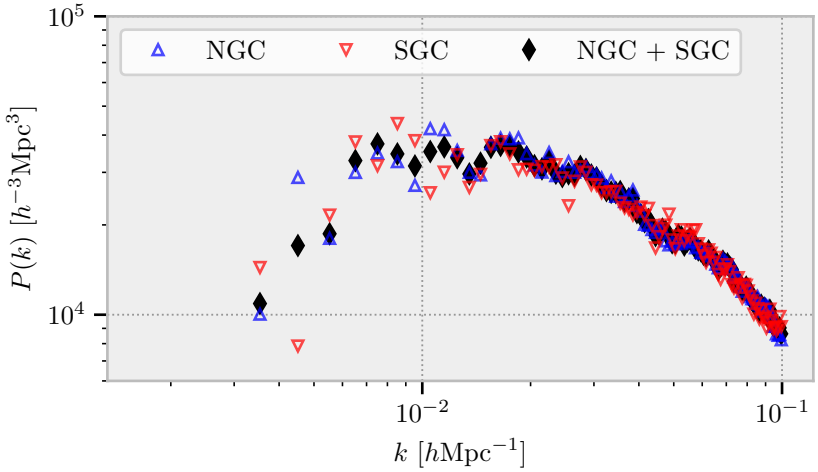


Figure 5.35: Power spectrum for the NGC (blue), SGC (red) and NGC combined with SGC (black). NGC and SGC are showing a perfect agreement.

► **Redshift evolution** To validate the imaging systematic mitigation, we can also look at the power spectrum in different redshift bins. Fig. 5.36 shows the power spectrum for different redshift bins. Note that only the second redshift bin mixes the two weights to mitigate the imaging systematics, otherwise, each redshift bin is using a single weight.

As expected, one can see an amplitude evolution as a function of the redshift. However, the global shape of the power spectrum shows no difference between the different redshift bins.

Finally, with these two weights to mitigate the imaging systematics, we have shown that

- The four different photometric regions lead to a similar power spectrum.
- No discrepancy between the different redshift bins.
- One can compute the power spectrum on all the NGC and SGC.

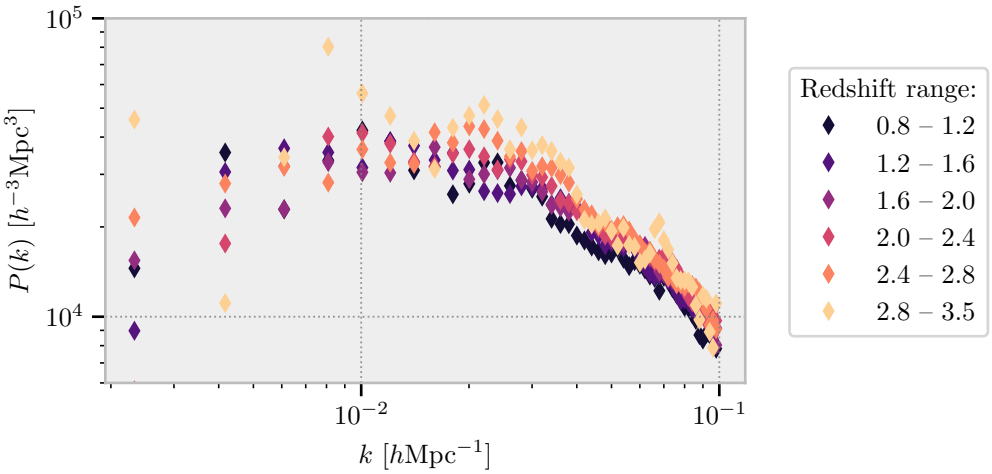


Figure 5.36: Power spectrum (NGC + SGC) computed for different redshift bins. Only the second redshift bin ($1.2 < z < 1.6$) uses the two weights simultaneously, otherwise, each redshift bin is using a single weight generated either for the low-z or high-z sample.

5.4.2 Constraints on local PNG

Due to the limitation on the maximal redshift in the EZmocks, we will limit here the analysis for the QSO with $0.8 < z < 3.1$.

5.4.2.1 Compare DESI QSO Y1 vs. EZMocks

The covariance matrix, see Section 5.1.2.4, that will be used for the parameter estimation step, will be computed from 1000 EZmocks. Unfortunately, these EZmocks are generated at $z = 1.4$ *i.e.* at a lower redshift than the effective redshift of the DESI QSO Y1 sample given in Table 5.1.

Since we are measuring large scales, where linear theory applies, it does not really matter. This is only a question of rescaling. Indeed, the linear power spectrum evolves only as a function of the growth factor that does not depend on k, see late time evolution given by (1.106).

Hence, we only need the power spectrum from the EZmocks to have a similar amplitude as the power spectrum from the data, such that we will have the correct errors on the power spectrum. Fig. 5.37 shows the mean power spectrum of the EZmocks (blue line) compared to the power spectrum from the DESI QSO Y1. The expected errors are the dispersion of the 1000 EZmocks shown in light blue. It is worth noting that the catalog presented here is blind and does not show any large-scale excess or deficiency of power.

5.4.2.2 Analysis with EZmocks

Before fitting the data, we test the pipeline and evaluate the sensitivity of our sample by fitting the mean of the 1000 EZmocks without normalizing the covariance matrix *i.e.* we use the covariance of the 1000 EZmocks. First, for testing purposes and to be conservative, we rescale the mean of the EZmocks to find the expected bias from Laurent et al. (2017) at the effective redshift given in Table 5.1. We fit the monopole ($\ell = 0$) and the quadrupole ($\ell = 2$) with

$$\begin{cases} \ell_0 : 0.003 < k < 0.08 \text{ with } k_{step} = 0.001 \quad [h \text{ Mpc}^{-1}] \\ \ell_2 : 0.003 < k < 0.08 \text{ with } k_{step} = 0.002 \quad [h \text{ Mpc}^{-1}] \end{cases}, \tag{5.65}$$

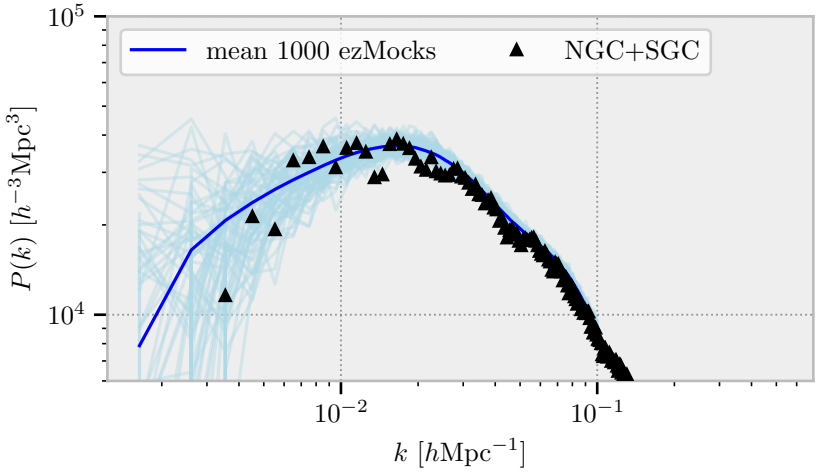


Figure 5.37: Comparison between the mean (blue line) of the 1000 EZmocks (light blue lines) used to compute the covariance and the power spectrum (NGC + SGC) computed from the data (black triangles) with $0.8 < z < 3.5$. Here, the data are blinded with $f_{\rm NL}^{blind} \in [-15, 15]$.

where we have rebin the quadrupole to avoid any problem related to the power spectrum estimation since the EZmock quadrupoles are highly scattered.

Unless otherwise stated, we will use $p = 1.6$ to parameterize the relation (2.120), as it was proposed by Slosar et al. (2008) and was done recently by Castorina et al. (2019), Mueller et al. (2022).

The result of the fit is given in Table 5.6 when we use no weights, FKP weights (5.8) and OQE weights (5.21), and the posteriors are shown in Fig. 5.38.

Table 5.6: Best fit values from the mean of 1000 EZmocks with the realistic DESI QSO Y1 covariance matrix. Errors are the 1σ (68% CL) from the posteriors shown in Fig. 5.38.

	$f_{NL}^{\rm loc}$	b	$s_{n,0}$	Σ_s
no weights	3^{+15}_{-15}	$2.405^{+0.041}_{-0.043}$	280^{+530}_{-430}	$3.13^{+1.08}_{-0.59}$
no weights (new maskbits)	2^{+15}_{-14}	$2.405^{+0.040}_{-0.041}$	290^{+440}_{-480}	$3.13^{+1.05}_{-0.56}$
FKP weights	2^{+14}_{-12}	$2.519^{+0.041}_{-0.045}$	140^{+520}_{-460}	$2.64^{+1.22}_{-0.72}$
OQE weights	$-2.5^{+9.4}_{-8.7}$	$3.255^{+0.063}_{-0.064}$	-700^{+680}_{-680}	$0^{+0.26}_{-0.89}$

The different sets of weights modify the effective redshift of the sample as we overweight high redshift quasars, and so the linear bias b_1 used to fit the theory to the data increases reducing the error on the desired parameter: $f_{\rm NL}^{\rm loc}$.

Although the measured value of $f_{\rm NL}^{\rm loc}$ is not strictly 0, one can assume that the systematic bias in our measurement is low enough compared to the statistical Y1 errors. In particular, the geometric effects from the window function are sufficiently well taken into account in our theoretical model, see Fig. 5.5. Note also, that the difference between the FKP weights and the OQE weights $(2-(-2.5) = 4.5)$ is about 1/3 (*resp.* 1/2) of the statistical error from FKP (*resp.* OQE) weights, such that it corresponds[32] to an increase about 5% (*resp.* 10%) of the errors.

[32]Systematic and statistical errors are added in quadrature: $\sigma_{\rm tot} = \sqrt{\sigma^2_{\rm sys} + \sigma^2_{\rm stat}}$

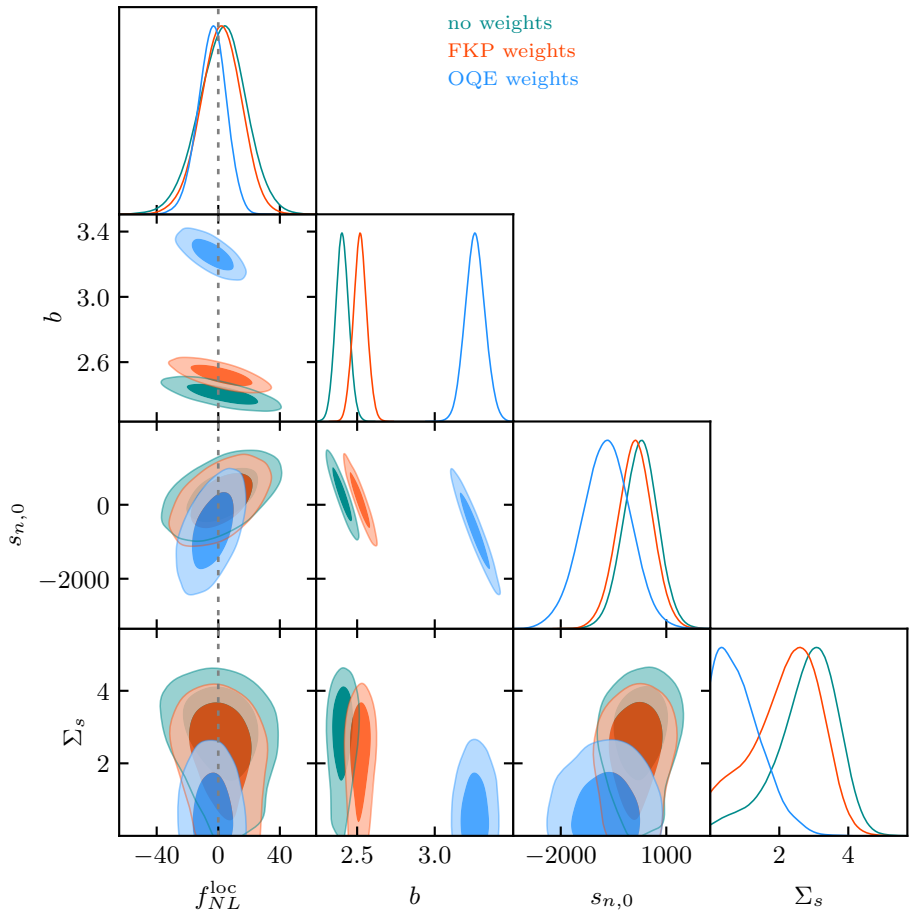


Figure 5.38: Posteriors for the mean of the 1000 EZmocks without weights (green), with FKP weights (orange) and OQE weights (blue), calibrated with the expected bias from Laurent et al. (2017) at the different effective redshifts given in Table 5.1 using the DESI QSO Y1 covariance matrix.

Therefore, it could be, for the moment, neglected.

► **Error evolution** As described in Section 5.2, the imaging systematic mitigation can overfit the data and remove too much power at a very large scale. This can be limited by not considering these scales during the fitting procedure.

Fig.5.39 shows the error evolution as a function of the minimal (*resp.* maximal) k value used during the fit. In particular, it is interesting to note that reducing k_{min} to 0.001 does not drastically improve error compared to 0.003. We need to find the best trade-off between statistical and systematic errors. Hence, to avoid any imaging contamination or overfitting consideration, we choose $k_{min} = 0.003$. On the other hand, the value k_{max} is fixed by the validity of our theoretical model, and we choose $k_{max} = 0.08$, see Fig. 1.14.

► **New maskbits** For safety, the current DESI clustering pipeline adds three additional imaging masks, removing all the quasars with `maskbits` 7, 8, and 11. This corresponds to removing about 6% of the final clustering catalogs.

Since the quasars are shot noise limited, we would certainly like to keep as many objects as possible. Fig. 5.40 shows the posterior for the nominal case (blue) and when we do not use the

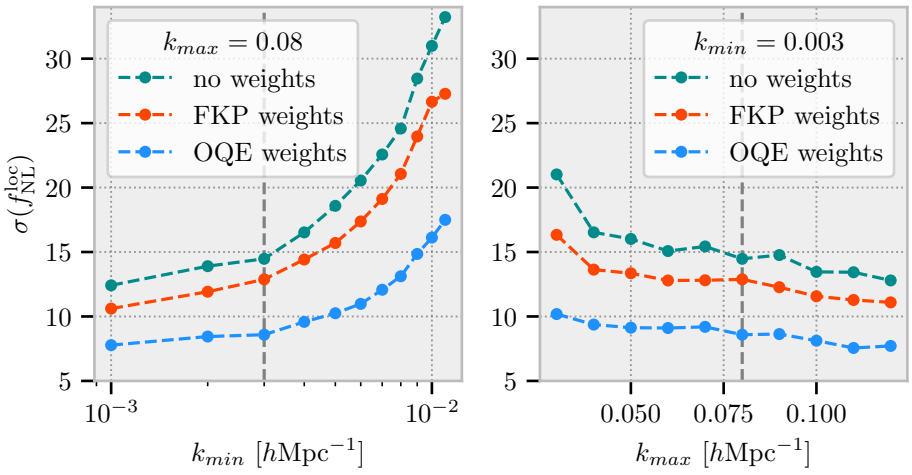


Figure 5.39: Left: Evolution of the errors for the different weighting schemes as a function of the minimal k value used during the fit at $k_{max} = 0.08$. Right: Evolution of the error as a function of the maximal k value at $k_{min} = 0.003$. Errors are estimated as the 1σ interval given by the posterior computed with MCMC chains from a fit of the mean of the 1000 EZmocks calibrated with the expected bias from Laurent et al. (2017) at the different effective redshifts, using the DESI QSO Y1 covariance matrix.

three additional `maskbits`. The best-fit parameters are given in Table. 5.6. The expected gain on the error of $f_{\rm NL}^{\rm loc}$ is also about 6%, as expected (4.2).

Further work will be required to ensure that the addition of these quasars is not problematic from an imaging systematics point of view so that they can be included in this analysis.

5.4.2.3 Analysis with DESI QSO Y1

Since our analysis pipeline has been approved on simulations, we can measure the blinded data without anxiety. The best-fit parameters are given in Table 5.7 and the posteriors for the power spectrum estimation without weights, with FKP weights and OQE weights are shown in Fig. 5.41.

Table 5.7: Best fit values for the DESI QSO Y1 data. Errors are the 1σ (68% CL) from the posteriors shown in Fig. 5.41.

	$f_{NL}^{\rm loc}$	b	$s_{n,0}$	Σ_s	$\chi^2/(115-4)$
no weights	8^{+25}_{-26}	$2.110^{+0.043}_{-0.045}$	1050^{+460}_{-470}	$7.95^{+0.47}_{-0.47}$	1.53
FKP weights	6^{+22}_{-19}	$2.206^{+0.041}_{-0.048}$	810^{+440}_{-480}	$7.97^{+0.52}_{-0.41}$	1.38
OQE weights	-17^{+15}_{-12}	$2.833^{+0.074}_{-0.069}$	-480^{+630}_{-670}	$8.71^{+0.52}_{-0.53}$	1.54

As expected, the linear bias is higher at higher effective redshift, but the value found here is lower than the prediction from Laurent et al. (2017) that worked with the eBOSS QSO. Additionally, the difference between the FKP and the OQE weights is higher than with the EZmocks. These two effects will be discussed later.

The errors on $f_{\rm NL}^{\rm loc}$ are bigger than in the case of EZmocks because we measure a smaller bias in the data. To avoid this discussion, we should instead measure $b_\phi f_{\rm NL}^{\rm loc}$ directly, and we should

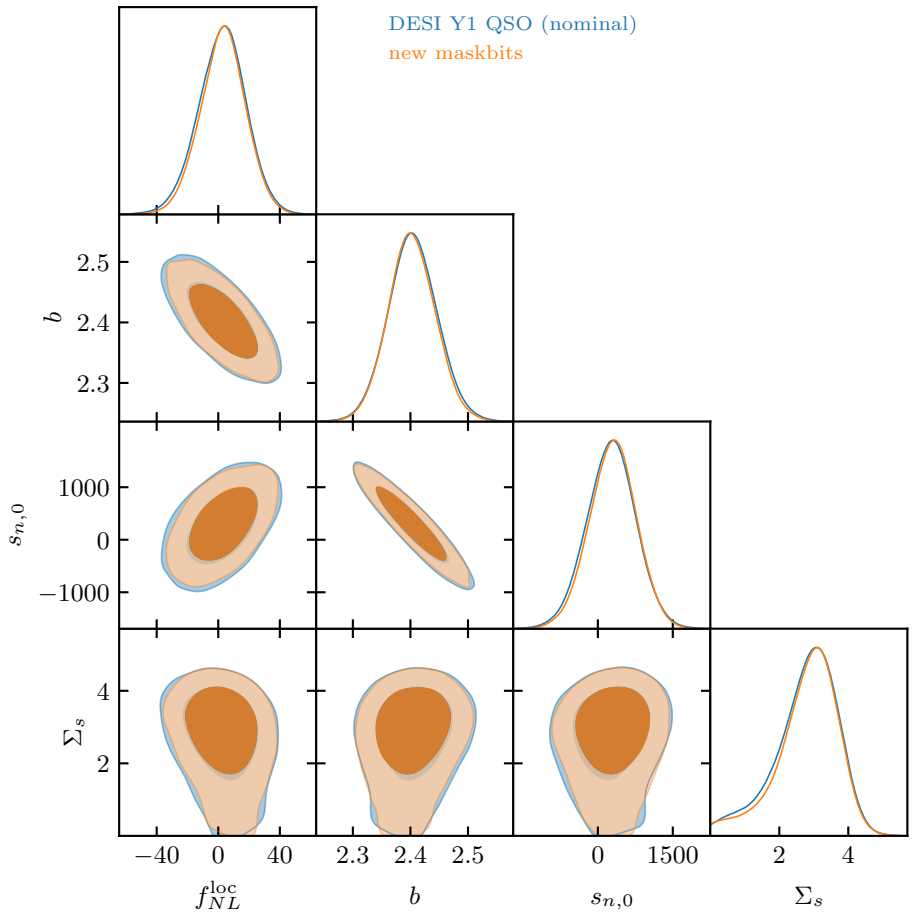


Figure 5.40: Similar to Fig. 5.38, but the posterior is for power spectrum estimated without weights for the nominal case (blue) and when we do not use the three additional `maskbits`.

then find similar errors. Note, however, that the error ratio is roughly similar between EZmocks and data using the same covariance matrix.

Fig. 5.42 shows the best-fit model compared to the data for the OQE weights case, and the corresponding residuals. We also plot the model at the best value where we change the value of $f_{\rm NL}^{\rm loc}$ to best-fit value $\pm 1\sigma$. Note that the quadrupole does not bring too many constraints on $f_{\rm NL}^{\rm loc}$, but it breaks down degeneracies between the various parameters of the model.

Remark: Compared to Castorina et al. (2019), we find Gaussian posterior for $f_{\rm NL}^{\rm loc}$, as in Mueller et al. (2022). This is because we are not probing sufficiently negative values of $f_{\rm NL}^{\rm loc}$ for the shape of the scale-dependent bias to be different.

► **Discrepancy between FKP and OQE weights** In the EZmocks, the small difference was noted in the fit of the mean of the 1000 EZmocks. To investigate the effect, we fit individually the 1000 EZmocks with different weights. To have similar errors to the data, we rescale the EZmocks at the effective redshift with the measured bias from Table 5.7.

The normalized distribution of the difference between the best-fit value obtained with the different sets of weights on each EZmock is shown in Fig. 5.43. First, the dispersion on 1000 EZmocks of the best value for each set of weights is compatible with the errors given in Table 5.7. We still have a coherent shift ($(\mu, \sigma) = (5.9, 9.1)$) in the distribution of the difference between

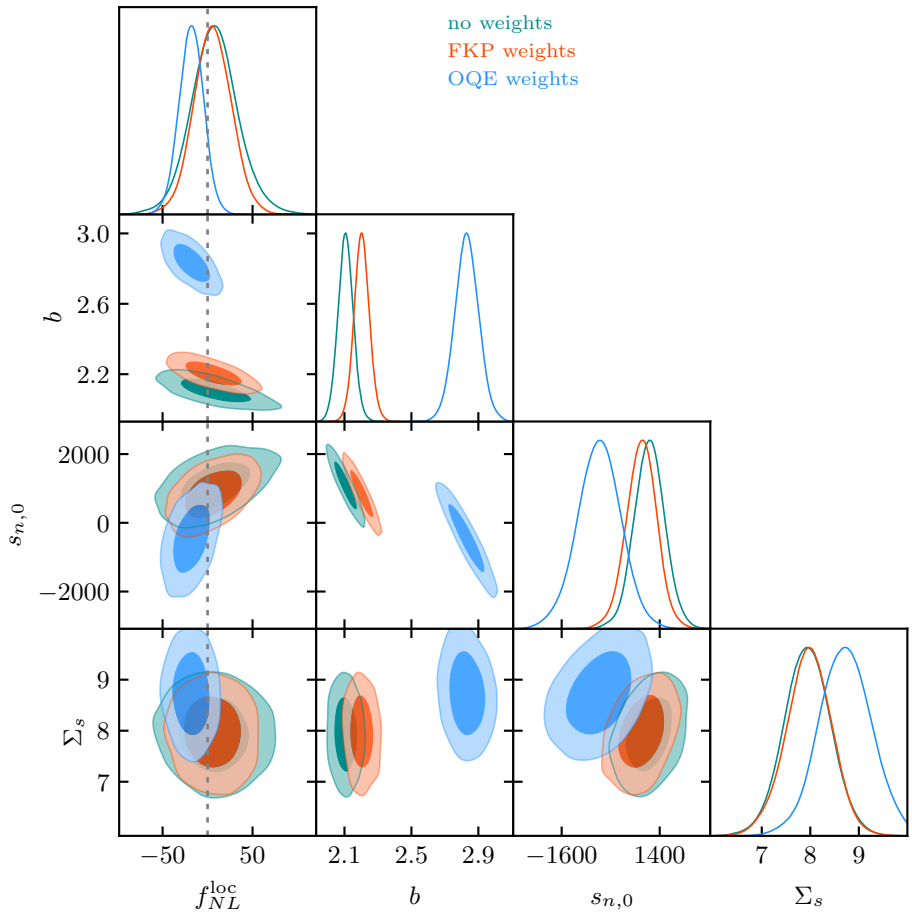


Figure 5.41: Posteriors for data without weights (green), with FKP weights (orange) and OQE weights (blue) at the different effective redshifts given in Table 5.1. Note that the linear bias b_1 is lower than the prediction at the effective redshift done by Laurent et al. (2017).

FKP and OQE cases, when we renormalize the dispersion, with the previous EZmocks case.

Finally, the difference that we observe $(6-(-17)=23)$ in the data (single realization), is included in the 2σ interval $[-12, 24]$ for the distribution of the difference between FKP and OQE cases shown in Fig. 5.43.

Remark: Although this shift also exists in Castorina et al. (2019), Mueller et al. (2022), it is not discussed in these papers since the effect was not large enough compared to the statistical errors. In particular, Mueller et al. (2022) found a shift of 24 $(13-(-11))$. However, we will need to study this shift in more detail, especially for the final release of DESI, to avoid biasing our measurement.

► **Bias QSO** To investigate the lower bias found in Table. 5.7, we compute the power spectrum in different redshift slices with the blinded data of the Y1 sample and with the *non−blinded* data of the first two months of observation, called hereafter DAO2. They are displayed in Fig. 5.44. The blinded and non-blinded data are consistent with each other. Unfortunately, the expected monopole at the considered effective redshift and with the expected bias from Laurent et al. (2017) is higher, especially for high-redshift slices.

To avoid any problem with the window function or other systematic effects that can impact

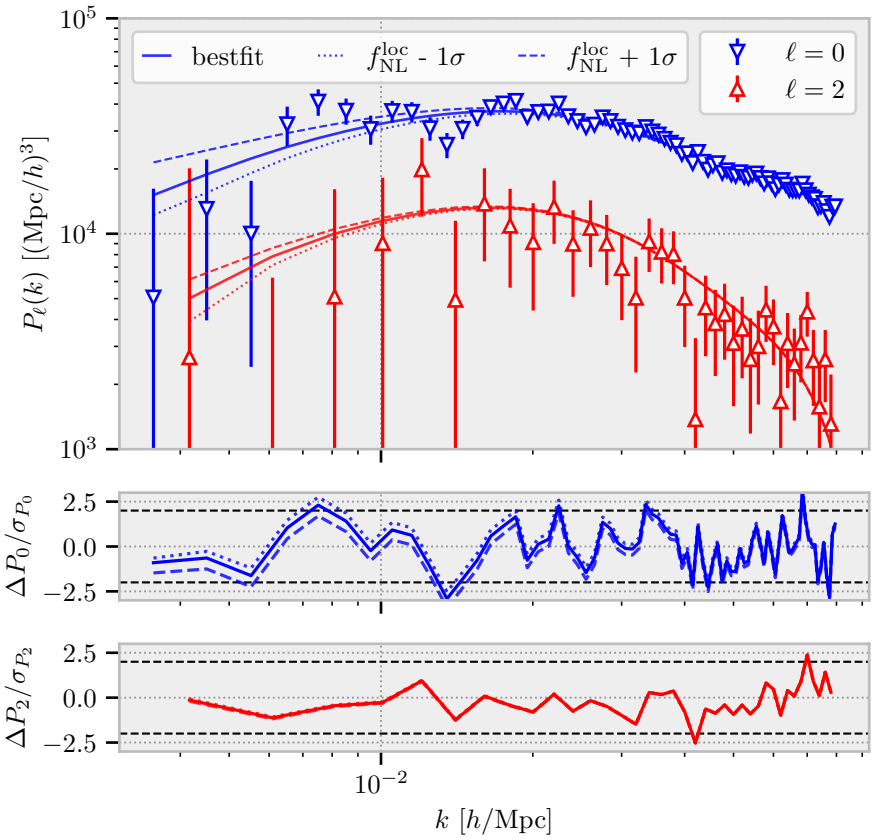


Figure 5.42: Top: Power spectrum, estimated with the OQE weights (p=1.6), of the DESI QSO Y1 sample and the associated model at the best fit (full line) values and with +/- 1σ (dashed/dotted) for $f_{\rm NL}^{\rm loc}$, given in Table 5.7. Bottom: Residuals of the monopole and quadrupole compare to the model at the best-fit values and with +/- 1σ for $f_{\rm NL}^{\rm loc}$.

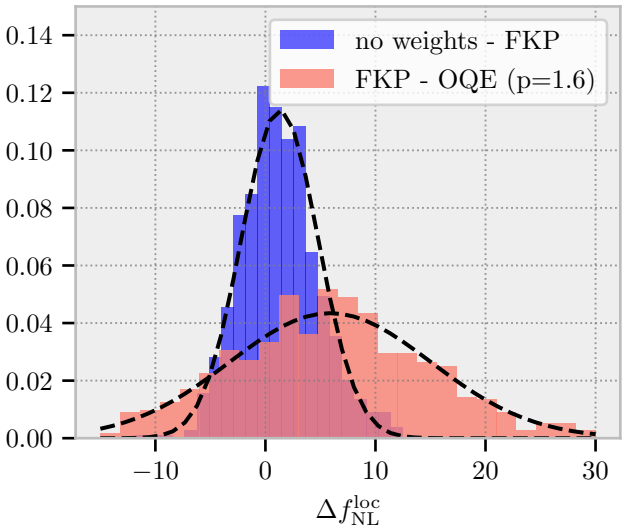


Figure 5.43: Normalized distribution of the difference between the best-fit value obtained with different sets of weights for each EZmock. Black dashed lines are the Gaussian fit of the two histograms. For 'no weights - FKP', the fit gives $(\mu, \sigma) = (1.3, 3.4)$, and for 'FKP - OQE', $(\mu, \sigma) = (5.9, 9.1)$.

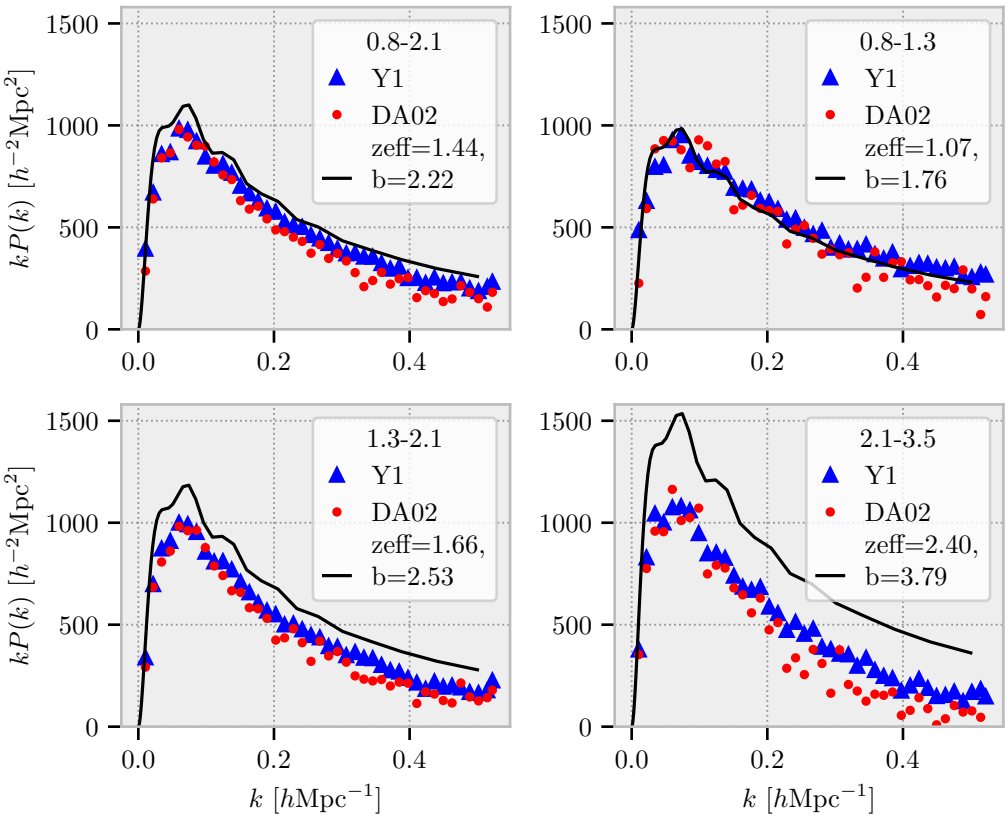


Figure 5.44: Power spectrum computed without FKP or OQE weights in different redshift slices for the blinded data of the Y1 sample (blue) and the non-blinded data of the first two months of observation (red). For comparison, the expected monopole at the considered effective redshift and with the expected bias from Laurent et al. (2017) is plotted in black.

the amplitude of the power spectrum, we decide to fit the bias for the DESI QSO sample by comparing the amplitude of the monopole of the correlation function ξ_0 between 30 and 80 h^{-1}Mpc to the simple Kaiser formula (1.122). We split the QSO sample into ten redshift bins.

This measurement is displayed in Fig.5.45 where the bias measured with ξ_0 are plotted in blue. For comparison, we also plot the bias found in Table 5.7 which is compatible with the measurement from the correlation function, albeit slightly lower. As in Laurent et al. (2017), we fit the QSO bias with the phenomenological relation

$$b_{\rm QSO}(z) = \alpha\left[(1+z)^2 - 6.565\right] + \beta, \tag{5.66}$$

with

$$\alpha = 0.214 \pm 0.011, \quad \beta = 2.206 \pm 0.029. \tag{5.67}$$

The biases measured in our QSO sample are actually lower than the measurement from (Laurent et al. 2017): $\alpha = 0.278 \pm 0.018, \beta = 2.393 \pm 0.042$.

Remark: In the final analysis, we will therefore need to use the measure of bias evolution given for our data to calculate the OQE weights.

► **Systematic budget:** To assess our measurement, we need to quantify the amount of systematics in the analysis. They can be preliminarily classified as follows:

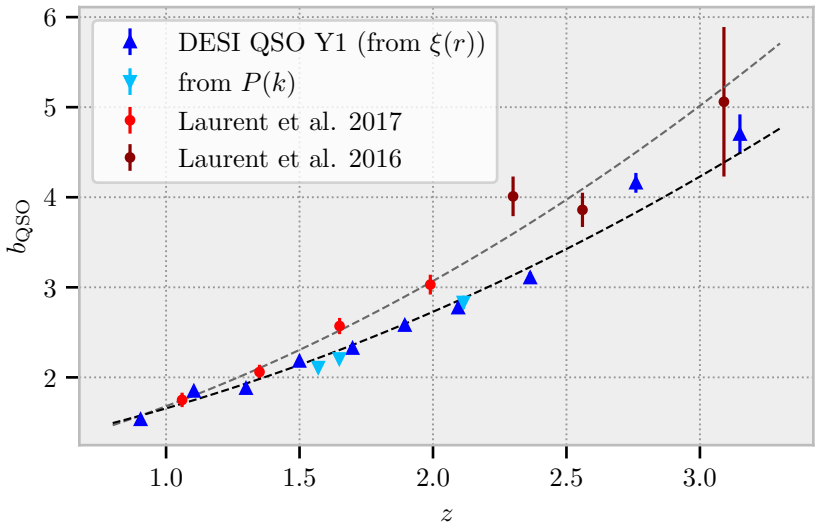


Figure 5.45: Linear bias for the QSO measured from the monopole of the correlation function between 30 and 80 h^{-1}Mpc (blue) in several redshift bins. The errors use a realistic covariance matrix. The best fit of (5.66) is plotted in black. The gray line is the best fit from the previous measurement done with BOSS (dark red, Laurent et al. 2016) and with (red, Laurent et al. 2017). For comparison, the biases found with the power spectrum, see Table 5.7, are displayed in sky blue, and these are not independent measures from those derived from the correlation function.

- Observational effect mitigation: From Section 5.2.2.2, we saw that `regressis` correction leads to a shift on $f_{\rm NL}^{\rm loc}$ about 10. Preliminary analysis based on a few FastPM simulations, see Fig. 5.15, indicates an additive shift. Once this shift is fully calibrated, we can rigorously calculate the associated systematic error. For the moment, and to be conservative, we can simply consider the half of the shift as a systematic error: $\sigma_{sys} = 10/2 = 5$.

- Fitting procedure: As discussed above, the use of OQE weights instead of FKP weights leads to a shift on $f_{\rm NL}^{\rm loc}$ about 6. Similarly to the previous shift and to be conservative, we consider a systematic error about $\sigma_{sys} = 6/2 = 3$.

- Error on the model: It is not discussed here since we do not have enough FastPM mocks to validate it. In particular, we do not study the impact of using the spherical halo collapse model in (2.120). For recent model validation, see, for instance, Biagetti et al. (2017). Similarly, the universal mass relation (2.120) could also be treated as theoretical systematics; however, we will give the measurement and the constraints with $p = 1$ in the next paragraph.

These systematic errors will be summed up in quadrature with the final statistical errors. We do not use any correlation coefficient between these two systematic errors to be as conservative as possible. In the OQE case, we found a statistical error about $\sigma_{stat} = 13.5$, and including the systematic errors increase the final errors to $\sigma_{tot} = \sqrt{\sigma_{stat}^2 + \sum_i \sigma_{sys,i}^2} \simeq 14.2$.

For this analysis, the systematic shifts observed are sufficiently small compared to the statistical error. However, the systematic budget will dominate over the statistical errors with either the upcoming Y5 data, the combination with the other tracers, or the use of more advanced methods. It will be necessary to better understand these shifts, whether they are additive or multiplicative, model them to correct them and estimate a systematic error in the correction.

▶ **Comparison with $p = 1$:** The previous measurement assumes $p = 1.6$ *i.e.* quasars come from recent mergers. This increases, unfortunately, the constraints on $f_{\rm NL}^{\rm loc}$. Here, we compare the evolution of the constraints by assuming either $p = 1.0$ or $p = 1.6$. Note that the OQE weights are recomputed with $p = 1$ and then the redshift effective is lower than the one with $p = 1.6$.

The best-fit values are given in Table 5.8 and the posteriors are displayed in Fig. 5.46 above the posteriors with $p = 1.6$. Using $p = 1.0$ halves the error for FKP weights ($\sigma = 20.5 \to 10$), while the gain is smaller for OQE weights (35 %, $\sigma = 13.5 \to 8.9$), mainly due to a smaller effective redshift.

Table 5.8: Best fit values for the DESI QSO Y1 data with $p = 1$ instead of $p = 1.6$. Errors are the 1σ (68% CL) from the posteriors shown in Fig. 5.38.

	$f_{NL}^{\rm loc}$	b	$s_{n,0}$	Σ_s	$\chi^2/(115-4)$
no weights	4^{+12}_{-11}	$2.110^{+0.044}_{-0.046}$	1050^{+410}_{-510}	$7.95^{+0.42}_{-0.48}$	1.53
FKP weights	$3^{+10.}_{-10.}$	$2.206^{+0.045}_{-0.046}$	810^{+480}_{-450}	$7.97^{+0.45}_{-0.43}$	1.38
OQE weights	$-10.9^{+9.3}_{-8.5}$	$2.676^{+0.057}_{-0.063}$	-730^{+560}_{-590}	$8.10^{+0.45}_{-0.44}$	1.51

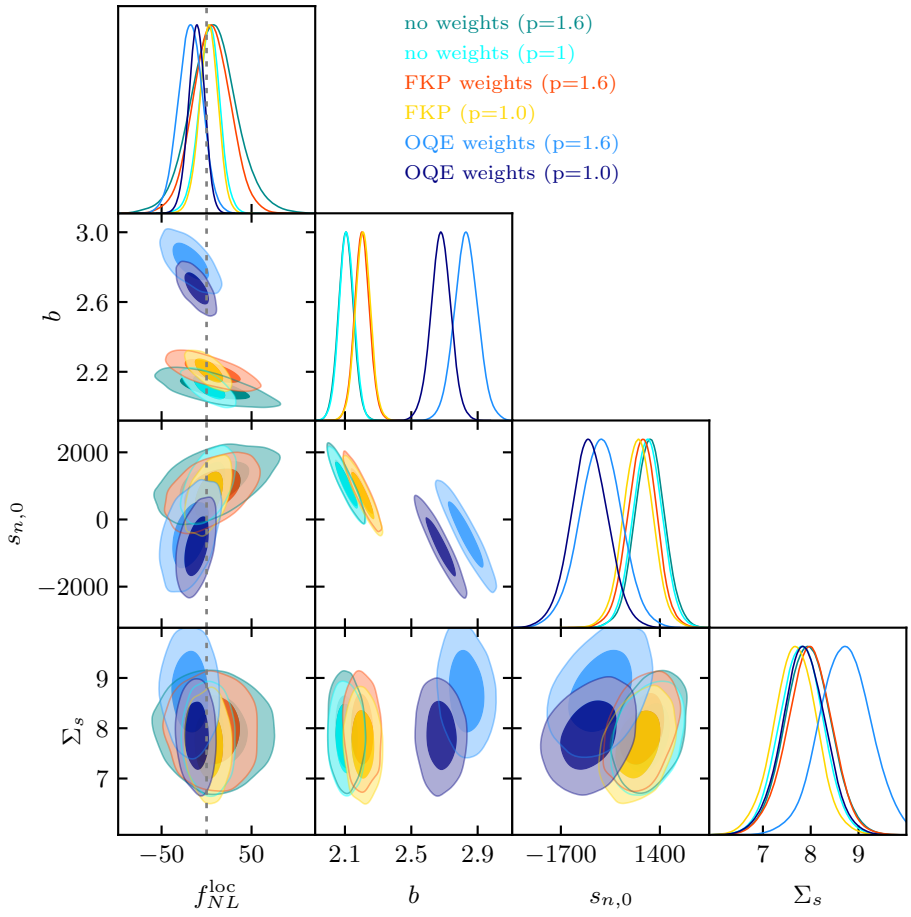


Figure 5.46: Similar to Fig. 5.46, where we have added the posteriors for $p = 1$.

5.4.3 Forecast for the final Y5 data release

Similarly to the description done in Section 5.1.2.4, we can generate EZmocks mimicking the expected Y5 (final) footprint of DESI. For this forecast, we assume an isotropic density of 188.7 deg^{-2} between $0.8 < z < 3.1$, an isotropic redshift distribution, and full completeness across all the footprint.

Here, we compute the power spectrum with the FKP weights, keeping in mind that the use of OQE weights can bring an additional 35% gain on the errors on $f_{\rm NL}^{\rm loc}$, and we recalibrate the power spectrum of the EZmocks to recover the bias measured in the Y1 data, see Table 5.7.

The posteriors, using FKP weights, for the mean of 1000 EZmocks mimicking either the Y1 dataset or the expected final Y5 dataset of DESI are shown in Fig 5.47 for $p = 1.6$ and $p = 1.0$. The best-fit values and the errors from the posteriors are given in Table 5.9.

Unfortunately, the lower bias found in the DESI QSO increases the expected errors and even with the OQE weights, one can only expect to have $\sigma \sim 10$ for the final Y5 dataset with $p = 1.6$. In order to have competitive constraints compared to the Planck ones, we will need to explore additional methods to extract the PNG signal from the large-scale structure data collected by DESI. For instance, we can think to combine this measurement with the measurement from the other tracers like LRG or split the QSO sample into different redshift bins. As done in Krolewski et al. (2023), we can also imagine combining the QSO data with the CMB lensing.

Table 5.9: Best-fits and errors from the 1σ posteriors displayed in Fig. 5.47 for the mean of 1000 EZmocks computed with the FKP weights.

	Y1 ($p = 1.6$)	Y1 ($p = 1.0$)	Y5 ($p = 1.6$)	Y5 ($p = 1.0$)
$f_{NL}^{\rm loc}$	$4^{+22.}_{-22.}$	$4^{+11.8}_{-9.9}$	6^{+14}_{-14}	$2^{+7.2}_{-6.5}$

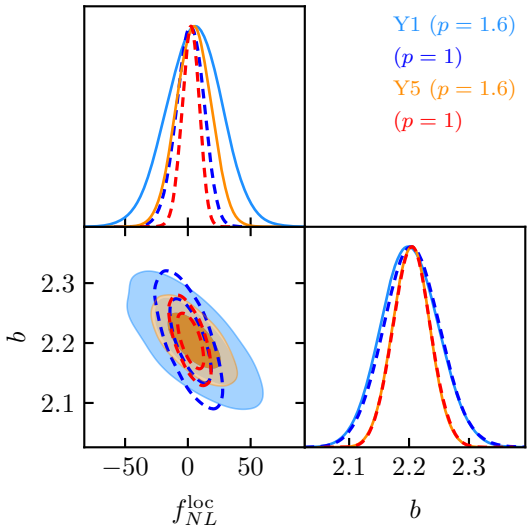


Figure 5.47: Posteriors for the mean of 1000 EZmocks, rescaled to the observed bias at $z_{eff} = 1.649$, mimicking either the current Y1 data or the expected Y5 final dataset. The power spectra are computed with FKP weights, and we do the forecast for $p = 1.6$ (full contours), the nominal value for quasars, and $p = 1.0$ (dashed contours).

Conclusion & Prospects

Pour chaque fin il y a toujours un nouveau départ.
Antoine de Saint-Exupéry, Le Petit Prince (1943)

In this dissertation, we have tried to observe the imprint that inflation, the theory describing the early Universe, may have left on the large-scale structures of the Universe. For this purpose, we have used the spectroscopic survey of quasars from the DESI instrument, in which we measure the 2-point correlation function, namely the power spectrum, to measure the presence of primordial non-gaussianity thanks to the scale-dependent bias relation.

This dissertation illustrates the fact that to measure the large scales of the power spectrum with controlled systematic errors, particular attention must be paid to every stage of the analysis: from target selection to power spectrum calculation.

First, I was in charge of the quasar target selection, one of the four main tracers for DESI (Chapter 4). Quasars have become a pillar of cosmological surveys over the past two decades. They are the source of choice to study large-scale structures at high redshift either as direct tracers of dark matter in the redshift range $0.8 < z < 2.1$ or as continuous tracers with the Lyman-α forest imprinted in their spectra at $z > 2.1$. The latter has received growing interest since the potential tension at high-redshift unveiled by eBOSS in the Lyman-α BAO measurement. Because DESI has a predefined number of targets for each tracer, the optimization of target selection is determinant for quasars that are shotnoise limited and whose selection is heavily contaminated by stars.

Extending the seminal color-cut classification from eBOSS, I developed a method based on Random Forest classification, in order to increase the efficiency of the selection by 20%. Additionally, I found that the baseline spectroscopic pipeline missed $\sim$ 10-15% of true quasars and could be optimized using an Mg II line finder and a neural network classifier to increase identification efficiency and avoid line confusion. The final combination presented reaches a redshift efficiency of $98.0 \pm 0.4\%$ and purity of $99.5 \pm 0.4\%$.

The selection was made difficult by unexpected stellar contamination from the Sagittarius stream. To validate the selection, I extensively studied the final sample to be sure that large-scale angular density fluctuations were explained by the known set of photometric features. In particular, I developed a template regression method based on Random Forest regression to mitigate the imaging systematics, `regressis`, by accounting for the non-linear relation between the imaging features and improving the mitigation. This method is implemented and currently used in the official DESI clustering pipeline.

This novel target selection and the optimized spectroscopic pipeline were extensively tested during the Survey Validation with a considerable effort of visual inspection conducted by the collaboration in which I was involved. This work enables DESI to collect more than 2.8 million spectra of quasars with correct redshift, including nearly 1 million with redshift $z > 2.1$, thereby

exceeding DESI scientific requirements by 20%. Because of the priority, DESI has already collected more than 1.4 million quasars during the first year of observation.

Finally, I used the DESI Y1 QSOs to constrain PNG with the scale-dependent bias method. For this purpose, I assessed the efficiency of the mitigation on the spectroscopic sample, I developed an end-to-end pipeline based on FastPM simulations mimicking one-third of the sample, including contamination from imaging systematics. This ongoing analysis has revealed an additive systematic shift due to the mitigation procedure and will need to be precisely studied shortly to ensure unbiased measurement of primordial non-gaussianity.

To avoid any confirmation bias during the imaging systematic mitigation and clustering catalog validation, I also developed a blinding scheme that mimics the behavior of a fake scale-dependent bias, emulating an unknown value of $f_{\rm NL}^{\rm loc}$ in the data. This blinding scheme was added to the global DESI blinding procedure. A complete procedure to validate the catalog needs to be clearly stated before unblinding the data, and unfortunately, I only worked with the blind data.

With the blind data, we explored the Y1 QSO and found an excellent agreement of the power spectrum at large scales between the different photometric regions when applying separately the imaging systematic mitigation in two redshift bins. After validating the fitting pipeline on realistic mocks, I fitted the blind data and found a lower linear bias than expected increasing the statistical errors. I found with FKP weights $f_{\rm NL}^{\rm loc} = 6^{+22}_{-19}$ and with OQE weights $f_{\rm NL}^{\rm loc} = -17^{+15}_{-12}$, with $p = 1.6$. The statistical gain between the two methods comes mainly from the increase of effective redshift.

This work led to several scientific publications: Chaussidon et al. (2022) for the imaging systematic mitigation, Chaussidon et al. (2023) for the quasar target selection and Chaussidon et al. (2023) for the blinding procedure. The measure of $f_{\rm NL}^{\rm loc}$ will be published with the Y1 data release over the next year. I was also involved in additional publications, see, in particular, Alexander et al. (2023) or Krolewski et al. (2023).

► **Prospects** Realistic forecasts for Y5 DESI data, using FKP weights, give $\sigma(f_{\rm NL}) = 22 \to 14$ for $p = 1.6$, and $\sigma(f_{\rm NL}) = 11 \to 7$ for $p = 1.0$. Although an additional gain of about 35% is expected using the OQE weights, the QSOs alone will not be sufficient to obtain a measurement competitive with Planck. However, this will be possible if all DESI data is considered.

Indeed, this measurement can be improved with the DESI data either by using several redshift bins for the QSO (low versus high redshift), the other dark time tracer as LRGs which seems to be quite competitive (already $\sigma(f_{\rm NL}) \sim 15$ with Y1 data). In addition, another promising approach is to use the CMB lensing to cross-correlate it with the DESI tracers. Below, we propose two detailed research avenues:

- **Improving PNG measurement with the bispectrum** The nominal survey of DESI is expected to constrain local PNG with scale-dependent bias up to $\sigma\left(f_{NL}^{loc}\right) \sim 5$ DESI Collaboration et al. (2016a). Adding the bispectrum information can improve the constraint on f_{NL}^{loc} compared to the power spectrum only measurement Tellarini et al. (2016), Karagiannis et al. (2018), Moradinezhad Dizgah et al. (2021). To test PNG at a level necessary to rule out the simplest inflation models, the use of bispectrum will be required.

The measurement of f_{NL}^{loc} via the scale-dependent bias is degenerated with an unknown bias term b_ϕ describing the response of the tracer density to local PNG. The seminal work Slosar

et al. (2008) provides an analytical description of b_ϕ, but more recent studies Barreira (2020; 2022a) based on hydrodynamical simulations show that it may not be correct for all tracers. The degeneracy is broken by the use of the bispectrum, which allows f_{NL}^{loc} to be measured without extending the prior on b_ϕ and improves the constraint on f_{NL}^{loc} by more than a factor of 5 compared to the power spectrum-only measurement Moradinezhad Dizgah et al. (2021). I would also like to extend my work to other forms of PNG as the equilateral and orthogonal forms, which can be probed with the bispectrum Coulton et al. (2023).

The first analyses using the bispectrum in the context of PNG appeared earlier in the year Cabass et al. (2022), D'Amico et al. (2022). However, there is still much work to do to reach the desired precision; in particular, observational systematic effects have been little studied so far. Particular attention must be paid to these since the squeeze configuration (triangle shape with two large sides and a small one: $k_1, k_2 \gg k_3$) can be contaminated by observational effects from both large (*e.g.* imaging systematics) and small (*e.g.* redshift determination) scales.

Additionally, geometrical effects such as the window function, are more complicated to model than in the power spectrum case Pardede et al. (2022), especially for the high multipoles. Cutting-edge analyses with bispectrum Ivanov et al. (2022), Philcox et al. (2022) used only the monopole, and special care would be needed for the higher-order multipoles. These geometrical effects can be modelled with a simulation-based approach to circumvent heavy analytical computation.

- **PNG with DESI successor** Once the mapping of the high-redshift Universe has been completed with DESI quasars, the successor of DESI could be a specific survey designed to target higher redshift objects as proposed for DESI-II (Schlegel et al. 2022). The previous analysis could then be performed with Lyman Break Galaxies (LBGs) in the redshift range ($2.2 < z < 3.6$), probing a larger volume and expected to yield better constraints on PNG.

For instance, to circumvent cosmic variance and reduce the errors drastically, it sounds enticing to cross-correlate several tracers with different biases (Seljak 2009, Hamaus et al. 2011). This method could be applied to DESI-II with two potential main tracers: Lyman-α emitters (LAEs) and LBGs. These two probes together open great opportunities to constrain f_{NL}^{loc} with the scale-dependent bias and other forms of non-gaussianity with bispectrum analyses (Yamauchi et al. 2017).

Exciting times lie just ahead.

Bibliography

Each new idea is dependent upon the ideas of the past.

Owen Chamberlain (1959)

Abazajian, K., A. Abdulghafour, G. E. Addison, P. Adshead, Z. Ahmed, et al. (2022). Snowmass 2021 CMB-S4 White Paper. arXiv e-prints 30, 32. arXiv:2203.08024.

Abazajian, K. N., P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, et al. (2016). CMB-S4 Science Book, First Edition. arXiv e-prints. arXiv:1610.02743.

Abbott, B. P., R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, et al. (2016). Observation of gravitational waves from a binary black hole merger. Physical Review Letters 116(6), 061102. arXiv:1602.03837.

Abbott, L. F., E. Farhi, and M. B. Wise (1982). Particle production in the new inflationary cosmology. Physics Letters B 117(1-2), 29–33.

Adamek, J., D. Daverio, R. Durrer, and M. Kunz (2016). General relativity and cosmic structure formation. Nature Physics 12(4), 346–349. arXiv:1509.01699.

Adams, F. C., J. R. Bond, K. Freese, J. A. Frieman, and A. V. Olinto (1993). Natural inflation: Particle physics models, power-law spectra for large-scale structure, and constraints from the Cosmic Background Explorer. Physical Review D 47(2), 426–455. arXiv:hep-ph/9207245.

Ahumada, R., C. A. Prieto, A. Almeida, F. Anders, S. F. Anderson, et al. (2020). The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. The Astrophysical Journal Supplement Series 249(1), 3. arXiv:1912.02905.

Alam, S., M. Ata, S. Bailey, F. Beutler, D. Bizyaev, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological analysis of the DR12 galaxy sample. Monthly Notices of the Royal Astronomical Society 470(3), 2617–2652. arXiv:1607.03155.

Alam, S., M. Aubert, S. Avila, C. Balland, J. E. Bautista, et al. (2021). Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory. Physical Review D 103(8). arXiv:2007.08991.

Albrecht, A. and P. J. Steinhardt (1982). Cosmology for grand unified theories with radiatively induced symmetry breaking. Physical Review Letters 48(17), 1220–1223.

Alcock, C. and B. Paczynski (1979). An evolution free test for non-zero cosmological constant. Nature 281.

Alexander, D. M., T. M. Davis, E. Chaussidon, V. A. Fawcett, A. X. Gonzalez-Morales, et al. (2023). The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra. The Astronomical Journal 165(3), 124. arXiv:2208.08517.

Alonso, D. (2012). CUTE solutions for two-point correlation functions from large cosmological datasets. arXiv e-prints. arXiv:1210.1833.

Anderson, L., É. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, et al. (2014). The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data releases 10 and 11 galaxy samples. Monthly Notices of the Royal Astronomical Society 441(1), 24–62. arXiv:1312.4877.

Angulo, R. E. and O. Hahn (2022). Large-scale dark matter simulations. Living Reviews in Computational Astrophysics 8(1). arXiv:2112.05165.

Angulo, R. E. and A. Pontzen (2016). Cosmological N-body simulations with suppressed variance. Monthly Notices of the Royal Astronomical Society: Letters 462(1), L1–L5. arXiv:1603.05253.

Antoja, T., P. Ramos, C. Mateu, A. Helmi, F. Anders, et al. (2020). An all-sky proper motion map of the Sagittarius stream using Gaia DR2. Astronomy and Astrophysics 635(2019), L3. arXiv:2001.10012.

Arnowitt, R., S. Deser, and C. W. Misner (2008). Republication of: The dynamics of general relativity. General Relativity and Gravitation 40(9), 1997–2027. arXiv:gr-qc/0405109.

ATLAS Collaboration, G. Aad, T. Abajyan, B. Abbott, J. Abdallah, et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Physics Letters B 716(1), 1–29. arXiv:1207.7214.

Aubourg, É., S. Bailey, J. E. Bautista, F. Beutler, V. Bhardwaj, et al. (2015). Cosmological implications of baryon acoustic oscillation measurements. Physical Review D - Particles, Fields, Gravitation and Cosmology 92(12), 9. arXiv:1411.1074.

Aver, E., K. A. Olive, and E. D. Skillman (2015). The effects of He i λ10830 on helium abundance determinations. Journal of Cosmology and Astroparticle Physics 2015(7). arXiv:1503.08146.

Avila, S. and A. G. Adame (2023). Validating galaxy clustering models with fixed and paired and matched-ICs simulations: application to primordial non-Gaussianities. Monthly Notices of the Royal Astronomical Society 519(3), 3706–3722. arXiv:2204.11103.

Bailey, S. J. (2023). Redrock in prep. (in prep).

Bardeen, J. M. (1980). Gauge-invariant cosmological perturbations. Physical Review D 22(8), 1882–1905.

Bardeen, J. M., P. J. Steinhardt, and M. S. Turner (1983). Spontaneous creation of almost scale-free density perturbations in an inflationary universe. Physical Review D 28(4), 679–693.

Barreira, A. (2020). On the impact of galaxy bias uncertainties on primordial non-gaussianity constraints. Journal of Cosmology and Astroparticle Physics 2020(12). arXiv:2009.06622.

Barreira, A. (2022a). Can we actually constrain $f_{\rm NL}$ using the scale-dependent bias effect? An illustration of the impact of galaxy bias uncertainties using the BOSS DR12 galaxy power spectrum. Journal of Cosmology and Astroparticle Physics 2022(11). arXiv:2205.05673.

Barreira, A. (2022b). Predictions for local PNG bias in the galaxy power spectrum and bispectrum and the consequences for $f_{\rm NL}$ constraints. Journal of Cosmology and Astroparticle Physics 2022(1). arXiv:2107.06887.

Baumann, D. (2022). Cosmology. Cambridge University Press.

Bautista, J. E., R. Paviot, M. Vargas Magaña, S. De La Torre, S. Fromenteau, et al. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. Monthly Notices of the Royal Astronomical Society 500(1), 736–762. arXiv:2007.08993.

Behroozi, P. S., R. H. Wechsler, and H. Y. Wu (2013). The ROCKSTAR phase-space temporal halo finder and the velocity offsets of cluster cores. The Astrophysical Journal 762(2). arXiv:1110.4372.

Bennett, C. L., M. Bay, M. Halpern, G. Hinshaw, C. Jackson, et al. (2003). The Microwave Anisotropy Probe Mission. The Astrophysical Journal 583(1), 1–23. arXiv:astro-ph/0301158.

Bennett, C. L., D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Final maps and results. The Astrophysical Journal, Supplement Series 208(2). arXiv:1212.5225.

Bernardeau, F. (2007). Cosmologie: des fondements théoriques aux observations.

Bernardeau, F., S. Colombi, E. Gaztanaga, and R. Scoccimarro (2002). Large-scale structure of the Universe and cosmological pertubation theory. Physics Report 367(1-3), 1–248. arXiv:astro-ph/0112551.

Beutler, F., E. Castorina, and P. Zhang (2019). Interpreting measurements of the anisotropic galaxy power spectrum. Journal of Cosmology and Astroparticle Physics 2019(3). arXiv:1810.05051.

Beutler, F., H. J. Seo, S. Saito, C. H. Chuang, A. J. Cuesta, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Anisotropic galaxy clustering in Fourier space. Monthly Notices of the Royal Astronomical Society 466(2), 2242–2260. arXiv:1607.03150.

Biagetti, M., T. Lazeyras, T. Baldauf, V. Desjacques, and F. Schmidt (2017). Verifying the consistency relation for the scale-dependent bias from local primordial non-Gaussianity. Monthly Notices of the Royal Astronomical Society 468(3), 3277–3288. arXiv:1611.04901.

Bianchi, D., A. Burden, W. J. Percival, D. Brooks, R. N. Cahn, et al. (2018). Unbiased clustering estimates with the DESI fibre assignment. Monthly Notices of the Royal Astronomical Society 481(2), 2338–2348. arXiv:1805.00951.

Block, D. L. (2012). Georges Lemaître and Stigler's Law of Eponymy. In arXiv e-prints, pp. 89–96.

Bolton, A. S. and D. J. Schlegel (2010). Spectro-Perfectionism: An Algorithmic Framework for Photon Noise-Limited Extraction of Optical Fiber Spectroscopy. Publications of the Astronomical Society of the Pacific, 100119133735095–000. arXiv:0911.2689.

Bolton, A. S., D. J. Schlegel, É. Aubourg, S. Bailey, V. Bhardwaj, et al. (2012). Spectral classification and Redshift measurement for the SDSS-III baryon oscillation spectroscopic survey. The Astronomical Journal 144(5), 144. arXiv:1207.7326.

Brieden, S., H. Gil-Marín, L. Verde, and J. L. Bernal (2020). Blind observers of the sky. Journal of Cosmology and Astroparticle Physics 2020(9), 052. arXiv:2006.10857.

Brodzeller, A., K. Dawson, S. Bailey, J. Yu, A. J. Ross, et al. (2023). Performance of the Quasar Spectral Templates for the Dark Energy Spectroscopic Instrument. The Astronomical Journal 166(2), 66. arXiv:2305.10426.

Bromm, V. and N. Yoshida (2011). The first galaxies. Annual Review of Astronomy and Astrophysics 49, 373–407. arXiv:1102.4638.

Brout, D., D. Scolnic, B. Popovic, A. G. Riess, A. Carr, et al. (2022). The Pantheon+ Analysis: Cosmological Constraints. The Astrophysical Journal 938(2), 110. arXiv:2202.04077.

Burden, A., W. J. Percival, and C. Howlett (2015). Reconstruction in Fourier space. Monthly Notices of the Royal Astronomical Society 453(1), 456–468. arXiv:1504.02591.

Burleigh, K. J. (2018). No Title. Ph. D. thesis, UC Berkeley.

Busca, N. and C. Balland (2018). QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks. Monthly Notices of the Royal Astronomical Society 000, 1–8. arXiv:1808.09955.

Cabass, G., M. M. Ivanov, O. H. Philcox, M. Simonović, and M. Zaldarriaga (2022). Constraints on Single-Field Inflation from the BOSS Galaxy Survey. Physical Review Letters 129(2). arXiv:2201.07238.

Carlson, J. and M. White (2010). Embedding realistic surveys in simulations through volume remapping. The Astrophysical Journal, Supplement Series 190(2), 311–314. arXiv:1003.3178.

Carrasco, J. J. M., M. P. Hertzberg, and L. Senatore (2012). The effective field theory of cosmological large scale structures. Journal of High Energy Physics 2012(9), 82. arXiv:1206.2926.

Castorina, E., N. Hand, U. Seljak, F. Beutler, C. H. Chuang, et al. (2019). Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space. Journal of Cosmology and Astroparticle Physics 2019. arXiv:1904.08859.

Castorina, E. and M. White (2018). Beyond the plane-parallel approximation for redshift surveys. Monthly Notices of the Royal Astronomical Society 476(4), 4403–4417. arXiv:1709.09730.

Chambers, K., E. Magnier, N. Metcalfe, H. Flewelling, M. Huber, et al. (2016). The Pan-STARRS1 Surveys. arXiv e-prints, arXiv:1612.05560. arXiv:astro-ph.IM/1612.05560.

Chaussidon, E., A. de Mattia, and C. Yeche (2023). Blinding method for scale-dependent bias relation. (in prep).

Chaussidon, E., C. Yèche, N. Palanque-Delabrouille, D. M. Alexander, J. Yang, et al. (2023). Target Selection and Validation of DESI Quasars. The Astrophysical Journal 944(1), 107. arXiv:2208.08511.

Chaussidon, E., C. Yèche, N. Palanque-Delabrouille, A. De Mattia, A. D. Myers, et al. (2022). Angular clustering properties of the DESI QSO target selection using DR9 Legacy Imaging Surveys. Monthly Notices of the Royal Astronomical Society 509(3), 3904–3923. arXiv:2108.03640.

Chen, X. (2005). Running non-Gaussianities in Dirac-Born-Infeld inflation. Physical Review D - Particles, Fields, Gravitation and Cosmology 72(12). arXiv:astro-ph/0507053.

Choi, K. Y. and C. T. Byrnes (2010). Review of local non-gaussianity from multifield inflation. Advances in Astronomy 2010. arXiv:1002.3110.

Chuang, C. H., F. S. Kitaura, F. Prada, C. Zhao, and G. Yepes (2015). EZmocks: Extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics. Monthly Notices of the Royal Astronomical Society 446(3), 2621–2628. arXiv:1409.1124.

CMS Collaboration, S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Physics Letters B 716(1), 30–61. arXiv:1207.7235.

Cole, S. and N. Kaiser (1989). Biased clustering in the cold dark matter cosmogony. Monthly Notices of the Royal Astronomical Society 237(4), 1127–1146.

Colless, M., G. Dalton, S. Maddox, W. Sutherland, P. Norberg, et al. (2001). The 2dF Galaxy Redshift Survey: Spectra and redshifts. Monthly Notices of the Royal Astronomical Society 328(4), 1039–1063. arXiv:astro-ph/0106498.

Consiglio, R., P. F. de Salas, G. Mangano, G. Miele, S. Pastor, et al. (2018). PArthENoPE reloaded. Computer Physics Communications 233, 237–242. arXiv:1712.04378.

Cooke, R. J., M. Pettini, and C. C. Steidel (2018). One Percent Determination of the Primordial Deuterium Abundance. The Astrophysical Journal 855(2), 102. arXiv:1710.11129.

Cooper, A. P., S. E. Koposov, C. Allende Prieto, C. J. Manser, N. Kizhuprakkat, et al. (2023). Overview of the DESI Milky Way Survey. The Astrophysical Journal 947(1), 37. arXiv:2208.08514.

Coulton, W. R., F. Villaescusa-Navarro, D. Jamieson, M. Baldi, G. Jung, et al. (2023). Quijote-PNG: Simulations of Primordial Non-Gaussianity and the Information Content of the Matter Field Power Spectrum and Bispectrum. The Astrophysical Journal 943(1), 64. arXiv:2206.01619.

Creminelli, P., G. D'Amico, M. Musso, and J. Norema (2011). The (not so) squeezed limit of the primordial 3-point function. Journal of Cosmology and Astroparticle Physics 2011(11). arXiv:1106.1462.

Creminelli, P. and M. Zaldarriaga (2004). A single-field consistency relation for the three-point function. Journal of Cosmology and Astroparticle Physics (10), 101–107. arXiv:astro-ph/0407059.

Croom, S. M., B. J. Boyle, T. Shanks, R. J. Smith, L. Miller, et al. (2005). The 2dF QSO redshift survey - XIV. Structure and evolution from the two-point correlation function. Monthly Notices of the Royal Astronomical Society 356(2), 415–438. arXiv:astro-ph/0409314.

Croom, S. M., R. J. Smith, B. J. Boyle, T. Shanks, N. S. Loaring, et al. (2001). The 2dF QSO Redshift Survey - V. The 10k catalogue. Monthly Notices of the Royal Astronomical Society 322(4), L29–L36. arXiv:astro-ph/0104095.

Dalal, N., O. Doré, D. Huterer, and A. Shirokov (2008). Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects. Physical Review D - Particles, Fields, Gravitation and Cosmology 77(12), 1–17. arXiv:0710.4560.

D'Amico, G., M. Lewandowski, L. Senatore, and P. Zhang (2022). Limits on primordial non-Gaussianities from BOSS galaxy-clustering data. arXiv:2201.11518.

Davis, M., G. Efstathiou, C. S. Frenk, and S. D. M. White (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. The Astrophysical Journal 292, 371.

Davis, M., J. Huchra, D. W. Latham, and J. Tonry (1982). A survey of galaxy redshifts. II - The large scale space distribution. The Astrophysical Journal 253, 423.

Dawson, K. S., J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data. The Astronomical Journal 151(2), 44. arXiv:1508.04473.

Dawson, K. S., D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, et al. (2013). The baryon oscillation spectroscopic survey of SDSS-III. The Astronomical Journal 145(1). arXiv:1208.0022.

De Boer, W. (1994). Grand unified theories and supersymmetry in particle physics and cosmology. Progress in Particle and Nuclear Physics 33(C), 201–301. arXiv:hep-ph/9402266.

de Jong, R., O. Bellido-Tirado, J. G. Brynnel, A. Ezzati Amini, S. Frey, et al. (2022). 4MOST: the 4-metre multi-object spectroscopic telescope project in the assembly, integration, and test phase. In SPIE, Volume 12184, pp. 40. SPIE-Intl Soc Optical Eng.

de Lapparent, V., M. J. Geller, and J. P. Huchra (1986). A slice of the universe. The Astrophysical Journal 302, L1.

De Mattia, A. and V. Ruhlmann-Kleider (2019). Integral constraints in spectroscopic surveys. Journal of Cosmology and Astroparticle Physics 2019(8). arXiv:1904.08851.

De Mattia, A., V. Ruhlmann-Kleider, A. Raichoor, A. J. Ross, A. Tamone, et al. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1. Monthly Notices of the Royal Astronomical Society 501(4), 5616–5645. arXiv:2007.09008.

Dembinski, H., P. Ongmongkolkul, and E. Al. (2020). scikit-hep/iminuit.

DES Collaboration, T. M. C. Abbott, M. Adamow, M. Aguena, S. Allam, et al. (2021). The Dark Energy Survey Data Release 2. The Astrophysical Journal Supplement Series 255(29pp), 20. arXiv:2101.05765.

DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, et al. (2022). Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument. The Astronomical Journal 164(62pp), 207. arXiv:2205.10939.

DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, et al. (2023a). Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument. arXiv:2306.06307v1.

DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, et al. (2023b). The Early Data Release of the Dark Energy Spectroscopic Instrument. arXiv:2306.06308v1.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016a). The DESI Experiment Part I: Science,Targeting, and Survey Design. arXiv e-prints. arXiv:1611.00036.

DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, et al. (2016b). The DESI Experiment Part II: Instrument Design. arXiv e-prints. arXiv:1611.00037.

Desjacques, V., D. Jeong, and F. Schmidt (2018). Large-scale galaxy bias. Physics Reports 733, 1–193. arXiv:1611.09787.

Desjacques, V., U. Seljak, and I. T. Iliev (2009). Scale-dependent bias induced by local non-Gaussianity: A comparison to N-body simulations. Monthly Notices of the Royal Astronomical Society 396(1), 85–96. arXiv:0811.2748.

Dey, A., D. J. Schlegel, D. Lang, R. Blum, K. Burleigh, et al. (2019). Overview of the DESI Legacy Imaging Surveys. The Astronomical Journal 157(5), 168. arXiv:1804.08657.

Di Valentino, E., O. Mena, S. Pan, L. Visinelli, W. Yang, et al. (2021). In the realm of the Hubble tension - A review of solutions. Classical and Quantum Gravity 38(15). arXiv:2103.01183.

Dimopoulos, K. (2020). An analytic treatment of quartic hilltop inflation. Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics 809. arXiv:2006.06029.

Ding, Z., C.-H. Chuang, Y. Yu, L. H. Garrison, A. E. Bayer, et al. (2022). The DESI N-body Simulation Project II: Suppressing Sample Variance with Fast Simulations. MNRAS 000, 1–19. arXiv:2202.06074.

Dodelson, S. and F. Schmidt (2020). Modern Cosmology, Second Edition. Modern Cosmology, Second Edition, 1–494.

Donley, J. L., A. M. Koekemoer, M. Brusa, P. Capak, C. N. Cardamone, et al. (2012). Identifying luminous active galactic nuclei in deep surveys: Revised IRAC selection criteria. The Astrophysical Journal 748(2). arXiv:1201.3899.

Drinkwater, M. J., R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, et al. (2010). The WiggleZ Dark Energy Survey: Survey design and first data release. Monthly Notices of the Royal Astronomical Society 401(3), 1429–1452. arXiv:0911.4246.

du Mas des Bourboux, H., J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, et al. (2020). The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests. The Astrophysical Journal 901(2), 153. arXiv:2007.08995.

Dyson, F. W., A. S. Eddington, and C. Davidson (1920). IX. A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 220(571-581), 291–333.

Eddington, A. S. (1919). The total eclipse of 1919 May 29 and the influence of gravitation on light. The Observatory 42, 119–122.

Edelstein, J., P. N. Jelinsky, M. Levi, G. Tarle, and D. Brooks (2018). The DESI spectrograph system and production. In SPIE, Volume 10702, pp. 272. SPIE-Intl Soc Optical Eng.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. Annalen der Physik 354(7), 769–822.

Einstein, A. (1917). Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie.

Eisenstein, D. J., D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, et al. (2011). SDSS-III: Massive spectroscopic surveys of the distant Universe, the MILKY WAY, and extra-solar planetary systems. The Astronomical Journal 142(3), 72. arXiv:1101.1529.

Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. The Astrophysical Journal 633(2), 560–574. arXiv:astro-ph/0501171.

Elsner, F., B. Leistedt, and H. V. Peiris (2016). Unbiased methods for removing systematics from galaxy clustering measurements. Monthly Notices of the Royal Astronomical Society 456(2), 2095–2104. arXiv:1509.08933.

Elvin-Poole, J., M. Crocce, A. J. Ross, T. Giannantonio, E. Rozo, et al. (2018). Dark Energy Survey year 1 results: Galaxy clustering for combined probes. Physical Review D 98(4), 61. arXiv:1708.01536.

Englert, F. and R. Brout (1964). Broken symmetry and the mass of gauge vector mesons. Physical Review Letters 13(9), 321–323.

Farr, J., A. Font-Ribera, and A. Pontzen (2020). Optimal strategies for identifying quasars in DESI. Journal of Cosmology and Astroparticle Physics 2020(11), 15. arXiv:2007.10348.

Feldman, H. A., N. Kaiser, and J. A. Peacock (1994). Power-spectrum analysis of three-dimensional redshift surveys. The Astrophysical Journal 426, 23. arXiv:astro-ph/9304022.

Feng, Y., M. Y. Chu, U. Seljak, and P. McDonald (2016). FastPM: A new scheme for fast simulations of dark matter and haloes. Monthly Notices of the Royal Astronomical Society 463(3), 2273–2286. arXiv:1603.00476.

Feng, Y. and C. Modi (2017). A fast algorithm for identifying friends-of-friends halos. Astronomy and Computing 20, 44–51. arXiv:1607.03224.

Flaugher, B., H. T. Diehl, K. Honscheid, T. M. Abbott, O. Alvarez, et al. (2015). the Dark Energy Camera. The Astronomical Journal 150(5). arXiv:1504.02900.

Foreman-Mackey, D., D. W. Hogg, D. Lang, and J. Goodman (2013). emcee : The MCMC Hammer. Publications of the Astronomical Society of the Pacific 125(925), 306–312. arXiv:1202.3665.

Freese, K., J. A. Frieman, and A. V. Olinto (1990). Natural inflation with pseudo Nambu-Goldstone bosons. Physical Review Letters 65(26), 3233.

Gaia Collaboration, A. G. Brown, A. Vallenari, T. Prusti, J. H. De Bruijne, et al. (2018). Gaia Data Release 2: Summary of the contents and survey properties. Astronomy and Astrophysics 616, A1. arXiv:1804.09365.

Gangui, A., F. Lucchin, S. Matarrese, and S. Mollerach (1994). The three-point correlation function of the cosmic microwave background in inflationary models. The Astrophysical Journal 430, 447. arXiv:astro-ph/9312033.

Gariazzo, S., P. F. de Salas, O. Pisanti, and R. Consiglio (2022). PArthENoPE revolutions. Computer Physics Communications 271. arXiv:2103.05027.

Garrison, L. H., D. J. Eisenstein, D. Ferrer, N. A. Maksimova, and P. A. Pinto (2021). The abacus cosmological N-body code. Monthly Notices of the Royal Astronomical Society 508(1), 575–596. arXiv:2110.11392.

Giannantonio, T., A. J. Ross, W. J. Percival, R. Crittenden, D. Bacher, et al. (2014). Improved primordial non-Gaussianity constraints from measurements of galaxy clustering and the integrated Sachs-Wolfe effect. Physical Review D - Particles, Fields, Gravitation and Cosmology 89(2), 023511. arXiv:1303.1349.

Gil-Maŕin, H., J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. De La Torre, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. Monthly Notices of the Royal Astronomical Society 498(2), 2492–2531. arXiv:2007.08994.

Gong, J. O. (2017). Multi-field inflation and cosmological perturbations. International Journal of Modern Physics D 26(1). arXiv:1606.06971.

Gorski, K. M., E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, et al. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. The Astrophysical Journal 622(2), 759–771. arXiv:astro-ph/0409513.

Grossi, M., L. Verde, C. Carbone, K. Dolag, E. Branchini, et al. (2009). Large-scale non-Gaussian mass function and halo bias: Tests on N-body simulations. Monthly Notices of the Royal Astronomical Society 398(1), 321–332. arXiv:0902.2013.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. Physical Review D 23(2), 347–356.

Guth, A. H. and S. Y. Pi (1982). Fluctuations in the new inflationary universe. Physical Review Letters 49(15), 1110–1113.

Guth, A. H. and S. H. Tye (1980). Phase transitions and magnetic monopole production in the very early universe. Physical Review Letters 44(10), 631–635.

Guy, J., S. Bailey, A. Kremin, S. Alam, D. M. Alexander, et al. (2023). The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument. The Astronomical Journal 165(4), 144. arXiv:2209.14482.

Hahn, C., M. J. Wilson, O. Ruiz-Macias, S. Cole, D. H. Weinberg, et al. (2022). DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation. The Astronomical Journal 165(6), 253. arXiv:2208.08512.

Hamaus, N., U. Seljak, and V. Desjacques (2011). Optimal constraints on local primordial non-Gaussianity from the two-point statistics of large-scale structure. Physical Review D - Particles, Fields, Gravitation and Cosmology 84(8). arXiv:1104.2321.

Hand, N., Y. Li, Z. Slepian, and U. Seljak (2017). An optimal FFT-based anisotropic power spectrum estimator. Journal of Cosmology and Astroparticle Physics 2017(7). arXiv:1704.02357.

Hartlap, J., P. Simon, and P. Schneider (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. Astronomy and Astrophysics 464(1), 399–404. arXiv:astro-ph/0608064.

Hawking, S. W. (1982). The development of irregularities in a single bubble inflationary universe. Physics Letters B 115(4), 295–297.

Hazumi, M., P. A. Ade, A. Adler, E. Allys, K. Arnold, et al. (2020). LiteBIRD satellite: JAXA's new strategic L-class mission for all-sky surveys of cosmic microwave background polarization. arXiv:2101.12449.

Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. Physical Review Letters 13(16), 508–509.

Hinshaw, G., D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, et al. (2013). Nine-year wilkinson microwave anisotropy probe (WMAP) observations: Cosmological parameter results. The Astrophysical Journal, Supplement Series 208(2). arXiv:1212.5226.

Ho, S., A. Cuesta, H. J. Seo, R. De Putter, A. J. Ross, et al. (2012). Clustering of sloan digital sky survey III photometric luminous galaxies: The measurement, systematics, and cosmological implications. The Astrophysical Journal 761(1), 14. arXiv:1201.2137.

Hoffmann, K., J. Bel, and E. Gaztañaga (2015). Comparing halo bias from abundance and clustering. Monthly Notices of the Royal Astronomical Society 450(2), 1674–1692. arXiv:1503.00313.

Hogg, D. W. (1999). Distance measures in cosmology. arXiv e-prints. arXiv:astro-ph/9905116.

Holman, R. and A. J. Tolley (2008). Enhanced non-Gaussianity from excited initial states. Journal of Cosmology and Astroparticle Physics 2008(5). arXiv:0710.1302.

Hou, J., A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, et al. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the quasar sample in configuration space between redshift 0.8 and 2.2. Monthly Notices of the Royal Astronomical Society 500(1), 1201–1221. arXiv:2007.08998.

Howlett, C., A. J. Ross, L. Samushia, W. J. Percival, and M. Manera (2015). The clustering of the SDSS main galaxy sample - II: Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z = 0.15. Monthly Notices of the Royal Astronomical Society 449(1), 848–866. arXiv:1409.3238.

Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the National Academy of Sciences 15(3), 168–173.

Huchra, J., M. Davis, D. Latham, and J. Tonry (1983). A survey of galaxy redshifts. IV - The data. The Astrophysical Journal Supplement Series 52, 89.

Huchra, J. P. and M. J. Geller (1982). Groups of galaxies. I - Nearby groups. The Astrophysical Journal 257, 423.

Huchra, J. P., M. S. Vogeley, and M. J. Geller (1999). The CfA Redshift Survey: Data for the South Galactic Cap. The Astrophysical Journal Supplement Series 121(2), 287–368.

Isi, M. (2022). Parametrizing gravitational-wave polarizations. arXiv:2208.03372.

Ivanov, M. M. (2023). Effective Field Theory for Large Scale Structure. arXiv:2212.08488v1.

Ivanov, M. M., O. H. Philcox, T. Nishimichi, M. Simonovic, M. Takada, et al. (2022). Precision analysis of the redshift-space galaxy bispectrum. Physical Review D 105(6). arXiv:2110.10161.

Ivezić, Ž., S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, et al. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. The Astrophysical Journal 873(2), 111. arXiv:0805.2366.

Izumi, K. and S. Mukohyama (2010). Trispectrum from ghost inflation. Journal of Cosmology and Astroparticle Physics 2010(6). arXiv:1004.1776.

James, F. and M. Roos (1975). Minuit - a system for function minimization and analysis of the parameter errors and correlations. Computer Physics Communications 10(6), 343–367.

Jeong, D. and E. Komatsu (2006). Perturbation Theory Reloaded: Analytical Calculation of Nonlinearity in Baryonic Oscillations in the Real-Space Matter Power Spectrum. The Astrophysical Journal 651(2), 619–626. arXiv:astro-ph/0604075.

Jones, D. H., M. A. Read, W. Saunders, M. Colless, T. Jarrett, et al. (2009). The 6dF Galaxy Survey: Final redshift release (DR3) and southern large-scale structures. Monthly Notices of the Royal Astronomical Society 399(2), 683–698. arXiv:0903.5451.

Kaiser, N. (1984). On the spatial correlations of Abell clusters. The Astrophysical Journal 284, L9.

Kaiser, N. (1987). Clustering in real space and in redshift space. Monthly Notices of the Royal Astronomical Society 227(1), 1–21.

Kallosh, R. and A. Linde (2021). BICEP/Keck and cosmological attractors. Journal of Cosmology and Astroparticle Physics 2021(12). arXiv:2110.10902.

Kalus, B., W. J. Percival, D. J. Bacon, E. M. Mueller, L. Samushia, et al. (2019). A map-based method for eliminating systematic modes from galaxy clustering power spectra with application to BOSS. Monthly Notices of the Royal Astronomical Society 482(1), 453–470. arXiv:1806.02789.

Karagiannis, D., A. Lazanu, M. Liguori, A. Raccanelli, N. Bartolo, et al. (2018). Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys. Monthly Notices of the Royal Astronomical Society 478(1), 1341–1376. arXiv:1801.09280.

Keck Array and BICEP2 Collaborations, P. A. Ade, Z. Ahmed, R. W. Aikin, K. D. Alexander, et al. (2018). Constraints on Primordial Gravitational Waves Using Planck, WMAP, and New BICEP2/ Keck Observations through the 2015 Season. Physical Review Letters 121(22). arXiv:1810.05216.

Keck Array and BICEP2 Collaborations, P. A. Ade, Z. Ahmed, M. Amiri, D. Barkats, et al. (2021). Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season. Physical Review Letters 127(15). arXiv:2110.00483v1.

Kellermann, K. (2014). The discovery of quasars and its aftermath. Journal of Astronomical History and Heritage 17(03), 267–282. arXiv:1304.3627.

Kingma, D. P. and J. L. Ba (2015). Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. arXiv:1412.6980.

Kitanidis, E., M. White, Y. Feng, D. Schlegel, J. Guy, et al. (2020). Imaging systematics and clustering of DESI main targets. Monthly Notices of the Royal Astronomical Society 496(2), 2262–2291. arXiv:1911.05714.

Kodama, H. and M. Sasaki (1984). Cosmological Perturbation Theory. Progress of Theoretical Physics Supplement 78, 1–166. arXiv:1303.2509.

Komatsu, E. (2010). Hunting for primordial non-Gaussianity in the cosmic microwave background. Classical and Quantum Gravity 27(12), 124010. arXiv:1003.6097.

Komatsu, E., J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, et al. (2009). Five-year wilkinson microwave anisotropy probe observations: Cosmological interpretation. The Astrophysical Journal, Supplement Series 180(2), 330–376. arXiv:0803.0547.

Komatsu, E., A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, et al. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Tests of Gaussianity. The Astrophysical Journal Supplement Series 148(1), 119–134. arXiv:astro-ph/0302223.

Komatsu, E. and D. N. Spergel (2001). Acoustic signatures in the primary microwave background bispectrum. Physical Review D - Particles, Fields, Gravitation and Cosmology 63(6), 13. arXiv:astro-ph/0005036.

Kong, H., K. J. Burleigh, A. Ross, J. Moustakas, C. H. Chuang, et al. (2020). Removing imaging systematics from galaxy clustering measurements with Obiwan: Application to the SDSS-IV extended Baryon Oscillation Spectroscopic Survey emission-line galaxy sample. Monthly Notices of the Royal Astronomical Society 499(3), 3943–3960. arXiv:2007.08992.

Krolewski, A., W. J. Percival, S. Ferraro, E. Chaussidon, M. Rezaie, et al. (2023). Constraining primordial non-Gaussianity from DESI quasar targets and Planck CMB lensing. arXiv:2305.07650.

Kurki-Suonio, H. (2019). Galaxy Survey Cosmology, Part 1.

Lan, T.-W., R. Tojeiro, E. Armengaud, J. X. Prochaska, T. M. Davis, et al. (2023). The DESI Survey Validation: Results from Visual Inspection of Bright Galaxies, Luminous Red Galaxies, and Emission-line Galaxies. The Astrophysical Journal 943(1), 68. arXiv:2208.08516.

Landau, L. and E. Lifshitz (1975). The Classical Theory of Fields (Fourth ed.). Course of theoretical physics.

Landau, L. and E. Lifshitz (1976). Mechanics (Third ed.).

Landy, S. D. and A. S. Szalay (1993). Bias and variance of angular correlation functions. The Astrophysical Journal 412, 64.

Laureijs, R., J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, et al. (2011). Euclid Definition Study Report. arXiv e-prints. arXiv:1110.3193.

Laurent, P., S. Eftekharzadeh, J. M. L. Goff, A. Myers, E. Burtin, et al. (2017). Clustering of quasars in SDSS-IV eBOSS: Study of potential systematics and bias determination. Journal of Cosmology and Astroparticle Physics 2017(7), 017. arXiv:1705.04718.

Laurent, P., J.-M. L. Goff, E. Burtin, J.-C. Hamilton, D. W. Hogg, et al. (2016). A 14 h^{-3} Gpc 3 study of cosmic homogeneity using BOSS DR12 quasar sample. Journal of Cosmology and Astroparticle Physics 2016(11), 060–060. arXiv:1602.09010.

Leistedt, B., H. V. Peiris, F. Elsner, A. Benoit-Lévy, A. Amara, et al. (2016). Mapping and simulating systematics due to spatially-varying observing conditions in DES Science Verification data. The Astrophysical Journal Supplement Series 226(2), 24. arXiv:1507.05647.

Leistedt, B., H. V. Peiris, D. J. Mortlock, A. Benoit-Lévy, and A. Pontzen (2013). Estimating the large-scale angular power spectrum in the presence of systematics: A case study of sloan digital sky survey quasars. Monthly Notices of the Royal Astronomical Society 435(3), 1857–1873. arXiv:1306.0005.

Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. Annales de la Société Scientifique de Bruxelles 47, 49–59.

Lemaître, G. (1931). A Homogeneous Universe of Constant Mass and Increasing Radius accounting for the Radial Velocity of Extra-galactic Nebulae. Monthly Notices of the Royal Astronomical Society 91(5), 483–490.

Limber, D. N. (1954). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. II. The Astrophysical Journal 119, 655.

Linde, A. D. (1982a). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. Physics Letters B 108(6), 389–393.

Linde, A. D. (1982b). Scalar field fluctuations in the expanding universe and the new inflationary universe scenario. Physics Letters B 116(5), 335–339.

Linder, E. V. (2005). Cosmic growth history and expansion history. Physical Review D - Particles, Fields, Gravitation and Cosmology 72(4), 1–8. arXiv:astro-ph/0507263.

Lovelock, D. (1972). The four-dimensionality of space and the Einstein tensor. Journal of Mathematical Physics 13(6), 874–876.

Lucchin, F. and S. Matarrese (1985). Power-law inflation. Physical Review D 32(6), 1316–1322.

Lyke, B. W., A. N. Higley, J. N. McLane, D. P. Schurhammer, A. D. Myers, et al. (2020). The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release. The Astrophysical Journal Supplement Series 250(1), 8. arXiv:2007.09001.

Ma, C.-P. and E. Bertschinger (1995). Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. The Astrophysical Journal 455, 7. arXiv:astro-ph/9506072.

Majewski, S. R., M. F. Skrutskie, M. D. Weinberg, and J. C. Ostheimer (2003). A Two Micron All Sky Survey View of the Sagittarius Dwarf Galaxy. I. Morphology of the Sagittarius Core and Tidal Arms. The Astrophysical Journal 599(2), 1082–1115. arXiv:astro-ph/0304198.

Maldacena, J. (2003). Non-gaussian features of primordial fluctuations in single field inflationary models. Journal of High Energy Physics 7(5), 233–264. arXiv:astro-ph/0210603.

Manera, M., R. K. Sheth, and R. Scoccimarro (2010). Large-scale bias and the inaccuracy of the peak-background split. Monthly Notices of the Royal Astronomical Society 402(1), 589–602. arXiv:0906.1314.

Martin, J., C. Ringeval, and V. Vennin (2014). Encyclopædia Inflationaris. Physics of the Dark Universe 5-6, 75–235. arXiv:1303.3787.

Mather, J. C., E. S. Cheng, J. Eplee, R. E., R. B. Isaacman, S. S. Meyer, et al. (1990). A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite. The Astrophysical Journal 354, L37.

Matsubara, T. (2008). Resumming cosmological perturbations via the Lagrangian picture: One-loop results in real space and in redshift space. Physical Review D - Particles, Fields, Gravitation and Cosmology 77(6). arXiv:0711.2521.

McCarthy, F., M. S. Madhavacheril, and A. S. Maniyar (2022). Constraints on primordial non-Gaussianity from halo bias measured through CMB lensing cross-correlations. arXiv:2210.01049.

McGreer, I. D., L. Jiang, X. Fan, G. T. Richards, M. A. Strauss, et al. (2013). The $z = 5$ quasar luminosity function from SDSS stripe 82. The Astrophysical Journal 768(2), 105. arXiv:1212.4493.

Meisner, A. M., D. Lang, and D. J. Schlegel (2017). Deep Full-sky Coadds from Three Years of WISE and NEOWISE Observations. The Astronomical Journal 154(4), 161. arXiv:1705.06746.

Meszaros, P. (1974). The behaviour of point masses in an expanding cosmological substratum. Astronomy and Astrophysics 37(2), 225–228.

Miller, T. N., P. Doel, G. Gutierrez, R. Besuner, D. Brooks, et al. (2023). The Optical Corrector for the Dark Energy Spectroscopic Instrument. arXiv:2306.06310.

Mo, H. J. and S. D. White (1996). An analytic model for the spatial clustering of dark matter haloes. Monthly Notices of the Royal Astronomical Society 282(2), 347–361. arXiv:astro-ph/9512127.

Mohammad, F. G., W. J. Percival, H. J. Seo, M. J. Chapman, D. Bianchi, et al. (2020). The completed SDSS-IV extended baryon oscillation spectroscopic survey: Pairwise-inverse probability and angular correction for fibre collisions in clustering measurements. Monthly Notices of the Royal Astronomical Society 498(1), 128–143. arXiv:2007.09005.

Moradinezhad Dizgah, A., M. Biagetti, E. Sefusatti, V. Desjacques, and J. Norea (2021). Primordial non-Gaussianity from biased tracers: Likelihood analysis of real-space power spectrum and bispectrum. Journal of Cosmology and Astroparticle Physics 2021(5). arXiv:2010.14523.

More, S., A. V. Kravtsov, N. Dalal, and S. Gottlöber (2011). The overdensity and masses of the friends-of-friends halos and universality of halo mass function. The Astrophysical Journal, Supplement Series 195(1), 4. arXiv:1103.0005.

Mueller, E. M., W. J. Percival, and R. Ruggeri (2019). Optimizing primordial non-Gaussianity measurements from galaxy surveys. Monthly Notices of the Royal Astronomical Society 485(3), 4160–4166. arXiv:1702.05088.

Mueller, E.-M., M. Rezaie, W. J. Percival, A. J. Ross, R. Ruggeri, et al. (2022). Primordial non-Gaussianity from the completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey II: measurements in Fourier space with optimal weights. Monthly Notices of the Royal Astronomical Society 514(3), 3396–3409. arXiv:2106.13725.

Mukhanov, V. F., H. A. Feldman, and R. H. Brandenberger (1992). Theory of cosmological perturbations. Physics Reports 215(5-6), 203–333.

Myers, A. D., R. J. Brunner, G. T. Richards, R. C. Nichol, D. P. Schneider, et al. (2006). First Measurement of the Clustering Evolution of Photometrically Classified Quasars. The Astrophysical Journal 638(2), 622–634. arXiv:astro-ph/0510371.

Myers, A. D., J. Moustakas, S. Bailey, B. A. Weaver, A. P. Cooper, et al. (2023). The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument. The Astronomical Journal 165(2), 50. arXiv:2208.08518.

Myers, A. D., N. Palanque-Delabrouille, A. Prakash, I. Pâris, C. Yeche, et al. (2015). The SDSS-IV extended Baryon oscillation spectroscopic survey: Quasar target selection. The Astrophysical Journal, Supplement Series 221(2), 27. arXiv:1508.04472.

Myers, A. D., M. White, and N. M. Ball (2009). Incorporating photometric redshift probability density information into real-space clustering measurements. Monthly Notices of the Royal Astronomical Society 399(4), 2279–2287. arXiv:0903.3121.

Nascimento, C. (2023). An accurate fluid approximation for massive neutrinos in cosmology. Physical Review D 108(2), 023505. arXiv:2303.09580.

Neveux, R., E. Burtin, A. de Mattia, A. Smith, A. J. Ross, et al. (2020). The completed SDSS-IV extended baryon oscillation spectroscopic survey: BAO and RSD measurements from the anisotropic power spectrum of the quasar sample between redshift 0.8 and 2.2. Monthly Notices of the Royal Astronomical Society 499(1), 210–229. arXiv:2007.08999.

Newberg, H. J., B. Yanny, C. Rockosi, E. K. Grebel, H.-W. Rix, et al. (2002). The Ghost of Sagittarius and Lumps in the Halo of the Milky Way. The Astrophysical Journal 569(1), 245–274. arXiv:astro-ph/0111095.

Palanque-Delabrouille, N., C. Magneville, C. Yeche, I. Pâris, P. Petitjean, et al. (2016). The extended Baryon Oscillation Spectroscopic Survey: Variability selection and quasar luminosity function. Astronomy and Astrophysics 587, A41. arXiv:1509.05607.

Palanque-Delabrouille, N., C. Yèche, J. Baur, C. Magneville, G. Rossi, et al. (2015). Neutrino masses and cosmology with Lyman-alpha forest power spectrum. Journal of Cosmology and Astroparticle Physics 2015(11). arXiv:1506.05976.

Palanque-Delabrouille, N., C. Yeche, A. D. Myers, P. Petitjean, N. P. Ross, et al. (2011). Variability selected high-redshift quasars on SDSS Stripe 82. Astronomy and Astrophysics 530, A122. arXiv:1012.2391.

Palanque-Delabrouille, N., C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, et al. (2020). Hints, neutrino bounds, and WDM constraints from SDSS DR14 Lyman-α and Planck full-survey data. Journal of Cosmology and Astroparticle Physics 2020(4). arXiv:1911.09073.

Pardede, K., F. Rizzo, M. Biagetti, E. Castorina, E. Sefusatti, et al. (2022). Bispectrum-window convolution via Hankel transform. Journal of Cosmology and Astroparticle Physics 2022(10). arXiv:2203.04174.

Peacock, J. A. (1998). Cosmological Physics. Cambridge University Press.

Pedregosa, F., V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, et al. (2011). Scikit-learn: Machine Learning in Python.

Percival, W. J., S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, et al. (2007). Measuring the baryon acoustic oscillation scale using the sloan digital sky survey and 2dF galaxy redshift survey. Monthly Notices of the Royal Astronomical Society 381(3), 1053–1066. arXiv:0705.3323.

Percival, W. J., A. J. Ross, A. G. Sánchez, L. Samushia, A. Burden, et al. (2014). The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Including covariance matrix errors. Monthly Notices of the Royal Astronomical Society 439(3), 2531–2541. arXiv:1312.4841.

Percival, W. J. and M. White (2009). Testing cosmological structure formation using redshift-space distortions. Monthly Notices of the Royal Astronomical Society 393(1), 297–308. arXiv:0808.0003.

Perlmutter, S., G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. The Astrophysical Journal 517(2), 565–586. arXiv:astro-ph/9812133.

Peskin, M. E. and D. V. Schroeder (1995). An Introduction to quantum field theory.

Peter, P. and J.-P. Uzan (2005). Cosmologie primordiale (2nd editio ed.).

Peter, P. and J.-P. Uzan (2013). Primordial Cosmology. Oxford Graduate Texts.

Peterson, B. M. (1997). An introduction to active galactic nuclei.

Philcox, O. H., M. M. Ivanov, G. Cabass, M. Simonović, M. Zaldarriaga, et al. (2022). Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order. Physical Review D 106(4). arXiv:2206.02800.

Pillepich, A., C. Porciani, and O. Hahn (2010). Halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions. Monthly Notices of the Royal Astronomical Society 402(1), 191–206. arXiv:0811.4176.

Pisanti, O., A. Cirillo, S. Esposito, F. Iocco, G. Mangano, et al. (2008). PArthENoPE: Public algorithm evaluating the nucleosynthesis of primordial elements. Computer Physics Communications 178(12), 956–971. arXiv:0705.0290.

Pitrou, C., A. Coc, J. P. Uzan, and E. Vangioni (2018). Precision big bang nucleosynthesis with improved Helium-4 predictions. Physics Reports 754, 1–66. arXiv:1801.08023.

Pitrou, C., A. Coc, J.-P. Uzan, and E. Vangioni (2020). Deuterium: a new bone of contention for cosmology? Monthly Notices of the Royal Astronomical Society 000, 1–10. arXiv:2011.11320.

Planck Collaboration, P. A. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, et al. (2014). Planck 2013 results. XXIV. Constraints on primordial non-gaussianity. Astronomy and Astrophysics 571, A24. arXiv:1303.5084.

Planck Collaboration, P. A. Ade, N. Aghanim, M. Arnaud, F. Arroja, et al. (2016). Planck 2015 results: XVII. Constraints on primordial non-Gaussianity. Astronomy and Astrophysics 594, A17. arXiv:1502.01592.

Planck Collaboration, P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, et al. (2016). Planck 2015 results: XIII. Cosmological parameters. Astronomy and Astrophysics 594. arXiv:1502.01589.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, et al. (2020). Planck 2018 results: VI. Cosmological parameters. Astronomy and Astrophysics 641. arXiv:1807.06209.

Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, et al. (2020). Planck 2018 results: X. Constraints on inflation. Astronomy and Astrophysics 641, 61. arXiv:1807.06211.

Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, et al. (2020). Planck 2018 results: IX. Constraints on primordial non-Gaussianity. Astronomy and Astrophysics 641, 24. arXiv:1905.05697.

Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies 2(1), 37–63. arXiv:2010.16061.

Prakash, A., T. C. Licquia, J. A. Newman, A. J. Ross, A. D. Myers, et al. (2016). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous red galaxy target selection. The Astrophysical Journal Supplement Series 224(2), 34. arXiv:1508.04478.

Preskill, J. P. (1979). Cosmological production of superheavy magnetic monopoles. Physical Review Letters 43(19), 1365–1368.

Press, W. H. and M. Davis (1982). How to identify and weigh virialized clusters of galaxies in a complete redshift catalog. The Astrophysical Journal 259, 449.

Prieto, C. A., A. P. Cooper, A. Dey, B. T. Gänsicke, S. E. Koposov, et al. (2020). Preliminary Target Selection for the DESI Milky Way Survey (MWS). Research Notes of the AAS 4(10), 188. arXiv:2010.11284.

Prusti, T., J. H. De Bruijne, A. G. Brown, A. Vallenari, C. Babusiaux, et al. (2016). The Gaia mission. Astronomy and Astrophysics 595, 14–108. arXiv:1609.04153.

Raichoor, A. (2023). Merge target list in prep. (in prep).

Raichoor, A., A. De Mattia, A. J. Ross, C. Zhao, S. Alam, et al. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues and measurement of the isotropic BAO between redshift 0.6 and 1.1 for the Emission Line Galaxy Sample. Monthly Notices of the Royal Astronomical Society 500(3), 3254–3274. arXiv:2007.09007.

Raichoor, A., D. J. Eisenstein, T. Karim, J. A. Newman, J. Moustakas, et al. (2020). Preliminary Target Selection for the DESI Emission Line Galaxy (ELG) Sample. Research Notes of the AAS 4(10), 180. arXiv:2010.11281.

Raichoor, A., J. Moustakas, J. A. Newman, T. Karim, S. Ahlen, et al. (2023). Target Selection and Validation of DESI Emission Line Galaxies. The Astronomical Journal 165(3), 126. arXiv:2208.08513.

Ravoux, C., M. L. A. Karim, E. Armengaud, M. Walther, N. G. Karaçaylı, et al. (2023). The Dark Energy Spectroscopic Instrument: One-dimensional power spectrum from first Lyman-α forest samples with Fast Fourier Transform. arXiv:2306.06311.

Reid, B. A., L. Verde, K. Dolag, S. Matarresed, and L. Moscardinif (2010). Non-Gaussian halo assembly bias. Journal of Cosmology and Astroparticle Physics 2010(7), 013. arXiv:1004.1637.

Rezaie, M., A. J. Ross, H.-J. Seo, H. Kong, A. Porredon, et al. (2023). Local primordial non-Gaussianity from the large-scale clustering of photometric DESI luminous red galaxies. arXiv e-prints, arXiv:2307.01753. arXiv:2307.01753.

Rezaie, M., A. J. Ross, H.-J. Seo, E.-M. Mueller, W. J. Percival, et al. (2021). Primordial non-gaussianity from the completed SDSS-IV extended baryon oscillation spectroscopic survey I: Catalogue preparation and systematic mitigation. Monthly Notices of the Royal Astronomical Society 000, 1–17. arXiv:2106.13724.

Rezaie, M., H. J. Seo, A. J. Ross, and R. C. Bunescu (2020). Improving galaxy clustering measurements with deep learning: Analysis of the DECaLS DR7 data. Monthly Notices of the Royal Astronomical Society 495(2), 1613–1640. arXiv:1907.11355.

Richards, G. T., X. Fan, H. J. Newberg, M. A. Strauss, D. E. Vanden Berk, et al. (2002). Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Quasar Sample. The Astronomical Journal 123(6), 2945–2975. arXiv:astro-ph/0202251.

Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. The Astronomical Journal 116(3), 1009–1038. arXiv:astro-ph/9805201.

Riess, A. G., W. Yuan, L. M. Macri, D. Scolnic, D. Brout, et al. (2022). A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s1 Mpc1 Uncertainty from the Hubble Space Telescope and the SH0ES Team. The Astrophysical Journal Letters 934(1), L7. arXiv:2112.04510.

Rieutord, M. (2014). Fluid dynamics: an introduction.

Rocher, A., V. Ruhlmann-Kleider, E. Burtin, S. Yuan, A. de Mattia, et al. (2023). The DESI One-Percent survey: exploring the Halo Occupation Distribution of Emission Line Galaxies with AbacusSummit simulations. arXiv:2306.06319.

Ross, A. J., J. Bautista, R. Tojeiro, S. Alam, S. Bailey, et al. (2020). The completed SDSS-IV extended baryon oscillation spectroscopic survey: Large-scale structure catalogues for cosmological analysis. Monthly Notices of the Royal Astronomical Society 498(2), 2354–2371. arXiv:2007.09000.

Ross, A. J., F. Beutler, C. H. Chuang, M. Pellejero-Ibanez, H. J. Seo, et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function. Monthly Notices of the Royal Astronomical Society 464(1), 1168–1191. arXiv:1607.03145.

Ross, A. J., S. Ho, A. J. Cuesta, R. Tojeiro, W. J. Percival, et al. (2011). Ameliorating systematic uncertainties in the angular clustering of galaxies: A study using the SDSS-III. Monthly Notices of the Royal Astronomical Society 417(2), 1350–1373. arXiv:1105.2320.

Ross, A. J., W. J. Percival, A. Carnero, G. B. Zhao, M. Manera, et al. (2013). The clustering of galaxies in the sdss-III DR9 baryon oscillation spectroscopic survey: Constraints on primordial non-gaussianity. Monthly Notices of the Royal Astronomical Society 428(2), 1116–1127. arXiv:1208.1491.

Ross, A. J., L. Samushia, C. Howlett, W. J. Percival, A. Burden, et al. (2015). The clustering of the SDSS DR7 main Galaxy sample - I: A 4 per cent distance measure at z = 0.15. Monthly Notices of the Royal Astronomical Society 449(1), 835–847. arXiv:1409.3242.

Ross, N. P., A. D. Myers, E. S. Sheldon, C. Yèche, M. A. Strauss, et al. (2012). THE SDSS-III baryon oscillation spectroscopic survey: Quasar target selection for data release nine. The Astrophysical Journal, Supplement Series 199(1), 3. arXiv:1105.0606.

Ruiz-Macias, O., P. Zarrouk, S. Cole, P. Norberg, C. Baugh, et al. (2020). Preliminary Target Selection for the DESI Bright Galaxy Survey (BGS). Research Notes of the AAS 4(10), 187. arXiv:2010.11283.

Rybicki, G. B. and W. H. Press (1992). Interpolation, realization, and reconstruction of noisy, irregularly sampled data. The Astrophysical Journal 398, 169.

Salopek, D. S. and J. R. Bond (1990). Nonlinear evolution of long-wavelength metric fluctuations in inflationary models. Physical Review D 42(12), 3936–3962.

Sargent, W. L. W. and E. L. Turner (1977). A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. The Astrophysical Journal 212, L3.

Scaramella, R., J. Amiaux, Y. Mellier, C. Burigana, C. S. Carvalho, et al. (2022). Euclid preparation: I. the Euclid Wide Survey. Astronomy and Astrophysics 662. arXiv:2108.01201.

Schlafly, E. F. and D. P. Finkbeiner (2011). Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD. The Astrophysical Journal 737(2), 103. arXiv:1012.4804.

Schlafly, E. F., D. Kirkby, D. J. Schlegel, A. D. Myers, A. Raichoor, et al. (2023). Survey Operations for the Dark Energy Spectroscopic Instrument. arXiv:2306.06309.

Schlegel, D. J., S. Ferraro, G. Aldering, C. Baltay, S. BenZvi, et al. (2022). A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5. arXiv e-prints. arXiv:2209.03585.

Schlegel, D. J., D. P. Finkbeiner, and M. Davis (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. The Astrophysical Journal 500, 525–553. arXiv:astro-ph/9710327.

Schmidt, M. (1963). 3C 273: A star-like object with large red-shift. Nature 197(4872), 1040.

Schwinger, J. (1961). Brownian motion of a quantum oscillator. Journal of Mathematical Physics 2(3), 407–432.

Scoccimarro, R. (2004). Redshift-space distortions, pairwise velocities, and nonlinearities. Physical Review D - Particles, Fields, Gravitation and Cosmology 70(8). arXiv:astro-ph/0407214.

Sefusatti, E., M. Crocce, R. Scoccimarro, and H. M. Couchman (2016). Accurate estimators of correlation functions in Fourier space. Monthly Notices of the Royal Astronomical Society 460(4), 3624–3636. arXiv:1512.07295.

Seljak, U. (2009). Extracting primordial non-Gaussianity without cosmic variance. Physical Review Letters 102(2). arXiv:0807.1770v1.

Shen, Y., W. N. Brandt, G. T. Richards, K. D. Denney, J. E. Greene, et al. (2016). the Sloan Digital Sky Survey Reverberation Mapping Project: Velocity Shifts of Quasar Emission Lines. The Astrophysical Journal 831(1), 7. arXiv:1602.03894.

Silber, J. H., P. Fagrelius, K. Fanning, M. Schubnell, J. N. Aguilar, et al. (2023). The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI). The Astronomical Journal 165(1), 9. arXiv:2205.09014.

Silva, D. R., R. D. Blum, L. Allen, A. Dey, D. J. Schlegel, et al. (2016). The Mayall z-band Legacy Survey. In American Astronomical Society Meeting Abstracts #228, Volume 228, pp. 317.02.

Slosar, A., C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan (2008). Constraints on local primordial non-Gaussianity from large scale structure. Journal of Cosmology and Astroparticle Physics 2008(8). arXiv:0805.3580.

Smith, A., A. de Mattia, E. Burtin, C. H. Chuang, and C. Zhao (2021). Reducing the variance of redshift space distortion measurements from mock galaxy catalogues with different lines of sight. Monthly Notices of the Royal Astronomical Society 500(1), 259–271. arXiv:2007.11417.

Smoot, G., C. Bennett, R. Weber, J. Maruschak, R. Ratliff, et al. (1990). COBE Differential Microwave Radiometers - Instrument design and implementation. The Astrophysical Journal 360, 685.

Spergel, D. N., R. Bean, O. Dore, M. R. Nolta, C. L. Bennett, et al. (2007). Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. The Astrophysical Journal Supplement Series 170(2), 377–408. arXiv:astro-ph/0603449.

Springel, V. (2005). The cosmological simulation code GADGET-2. Monthly Notices of the Royal Astronomical Society 364(4), 1105–1134. arXiv:astro-ph/0505010.

Starobinsky, A. A. (1980). A new type of isotropic cosmological models without singularity. Physics Letters B 91(1), 99–102.

Starobinsky, A. A. (1982). Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. Physics Letters B 117(3-4), 175–178.

Stern, D., P. Eisenhardt, V. Gorjian, C. S. Kochanek, N. Caldwell, et al. (2005). Mid-Infrared Selection of Active Galaxies. The Astrophysical Journal 631(1), 163–168. arXiv:astro-ph/0410523.

Suchyta, E., E. M. Huff, J. Aleksić, P. Melchior, S. Jouvel, et al. (2016). No galaxy left behind: Accurate measurements with the faintest objects in the Dark Energy Survey. Monthly Notices of the Royal Astronomical Society 457(1), 786–808. arXiv:1507.08336.

Sugiyama, N. S., E. Komatsu, and T. Futamase (2013). δn formalism. Physical Review D - Particles, Fields, Gravitation and Cosmology 87(2). arXiv:1208.1073v3.

Tamone, A., A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, et al. (2020). The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the Emission Line Galaxy sample. Monthly Notices of the Royal Astronomical Society 499(4), 5527–5546. arXiv:2007.09009.

Taruya, A., T. Nishimichi, and S. Saito (2010). Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. Physical Review D - Particles, Fields, Gravitation and Cosmology 82(6). arXiv:1006.0699.

Tauber, J. A., N. Mandolesi, J. L. Puget, T. Banos, M. Bersanelli, et al. (2010). Planck pre-launch status: The Planck mission. Astronomy and Astrophysics 520(1), A1.

Tegmark, M., A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay (1998). Measuring the Galaxy Power Spectrum with Future Redshift Surveys. The Astrophysical Journal 499(2), 555–576. arXiv:astro-ph/9708020.

Tellarini, M., A. J. Ross, G. Tasinato, and D. Wands (2016). Galaxy bispectrum, primordial non-Gaussianity and redshift space distortions. Journal of Cosmology and Astroparticle Physics 2016(6). arXiv:1603.06814.

Thompson, L. A. and S. A. Gregory (2011). An Historical View: The Discovery of Voids in the Galaxy Distribution. arXiv e-prints. arXiv:1109.1268.

Thorne, J., A. Robotham, L. Davies, and S. Bellstedt (2022). AGN Unification Diagram.

Tifft, W. G. and S. A. Gregory (1976). Direct observations of th large-scale distribution of galaxies. The Astrophysical Journal 205, 696.

Wang, F., X.-B. Wu, X. Fan, J. Yang, W. Yi, et al. (2016). a Survey of Luminous High-Redshift Quasars With Sdss and Wise. I. Target Selection and Optical Spectroscopy . The Astrophysical Journal 819(1), 24. arXiv:1602.04659.

Wang, Y., G. B. Zhao, C. Zhao, O. H. Philcox, S. Alam, et al. (2020). The clustering of the SDSS-IV extended baryon oscillation spectroscopic survey DR16 luminous red galaxy and emission-line galaxy samples: Cosmic distance and structure growth measurements using multiple tracers in configuration space. Monthly Notices of the Royal Astronomical Society 498(3), 3470–3483. arXiv:2007.09010.

Warren, M. S., P. J. Quinn, J. K. Salmon, and W. H. Zurek (1992). Dark halos formed via dissipationless collapse. I - Shapes and alignment of angular momentum. The Astrophysical Journal 399, 405.

Weaverdyck, N. and D. Huterer (2021). Mitigating contamination in LSS surveys: a comparison of methods. Monthly Notices of the Royal Astronomical Society 503(4), 5061–5084. arXiv:2007.14499.

Weinberg, S. (2005). Quantum contributions to cosmological correlations. Physical Review D - Particles, Fields, Gravitation and Cosmology 72(4), 1–19. arXiv:hep-th/0506236.

White, M. (2014). The Zel'dovich approximation. Monthly Notices of the Royal Astronomical Society 439(4), 3630–3640. arXiv:1401.5466.

Wise, J. H., M. J. Turk, M. L. Norman, and T. Abel (2012). The birth of a galaxy: Primordial metal enrichment and stellar populations. The Astrophysical Journal 745(1), 50. arXiv:1011.2632.

Wright, E. L., P. R. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, et al. (2010). The Wide-field Infrared Survey Explorer (wise): Mission description and initial on-orbit performance. The Astronomical Journal 140(6), 1868–1881. arXiv:1008.0031.

Wu, Q. and Y. Shen (2022). A Catalog of Quasar Properties from Sloan Digital Sky Survey Data Release 16. The Astrophysical Journal Supplement Series 263(2), 42. arXiv:2209.03987.

Xia, J. Q., C. Baccigalupi, S. Matarrese, L. Verde, and M. Viel (2011). Constraints on primordial non-Gaussianity from large scale structure probes. Journal of Cosmology and Astroparticle Physics 2011(8). arXiv:1104.5015.

Yamamoto, K., M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka (2006). A measurement of the quadrupole power spectrum in the clustering of the 2dF QSO survevy. Publications of the Astronomical Society of Japan 58(1), 93–102. arXiv:astro-ph/0505115.

Yamauchi, D., S. Yokoyama, and K. Takahashi (2017). Multitracer technique for galaxy bispectrum: An application to constraints on nonlocal primordial non-Gaussianities. Physical Review D 95(6). arXiv:1611.03590.

Yang, J., X. Fan, X.-B. Wu, F. Wang, F. Bian, et al. (2017). Discovery of 16 New $z \sim 5.5$ Quasars: Filling in the Redshift Gap of Quasar Color Selection. The Astronomical Journal 153(4), 184. arXiv:1703.03526.

Yèche, C., N. Palanque-Delabrouille, C.-A. Claveau, D. D. Brooks, E. Chaussidon, et al. (2020). Preliminary Target Selection for the DESI Quasar (QSO) Sample. Research Notes of the AAS 4(10), 179. arXiv:2010.11280.

Yèche, C., P. Petitjean, J. Rich, E. Aubourg, N. Busca, et al. (2010). Artificial neural networks for quasar selection and photometric redshift determination. Astronomy and Astrophysics 523(1), A14. arXiv:0910.3770.

York, D. G., J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, et al. (2000). The Sloan Digital Sky Survey: Technical Summary. The Astronomical Journal 120(3), 1579–1587. arXiv:astro-ph/0006396.

Yuan, S., H. Zhang, A. J. Ross, J. Donald-McCann, B. Hadzhiyska, et al. (2023). The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit. arXiv:2306.06314.

Zarrouk, P., E. Burtin, H. Gil-Marin, A. J. Ross, R. Tojeiro, et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: Measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2. Monthly Notices of the Royal Astronomical Society 477(2), 1639–1663. arXiv:1801.03062.

Zeldovich, Y. B. (1970). Gravitational Instability: An Approximate Theory for Large Density Perturbations. Astronomy and Astrophysics 5, 84–89.

Zeldovich, Y. B. and M. Y. Khlopov (1978). On the concentration of relic magnetic monopoles in the universe. Physics Letters B 79(3), 239–241.

Zhang, T., C. H. Chuang, R. H. Wechsler, S. Alam, J. DeRose, et al. (2023). Covariance matrices for variance-suppressed simulations. Monthly Notices of the Royal Astronomical Society 518(3), 3737–3745. arXiv:2112.10845.

Zhao, C., C.-H. Chuang, J. Bautista, A. de Mattia, A. Raichoor, et al. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: 1000 multi-tracer mock catalogues with redshift evolution and systematics for galaxies and quasars of the final data release. Monthly Notices of the Royal Astronomical Society 503(1), 1149–1173. arXiv:2007.08997.

Zheng, Y. and Y. S. Song (2016). Study on the mapping of dark matter clustering from real space to redshift space. Journal of Cosmology and Astroparticle Physics 2016(8). arXiv:1603.00101.

Zhou, R., B. Dey, J. A. Newman, D. J. Eisenstein, K. Dawson, et al. (2023). Target Selection and Validation of DESI Luminous Red Galaxies. The Astronomical Journal 165(2), 58. arXiv:2208.08515.

Zhou, R., J. A. Newman, K. S. Dawson, D. J. Eisenstein, D. D. Brooks, et al. (2020). Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample. Research Notes of the AAS 4(10), 181. arXiv:2010.11282.

Zonca, A., L. Singer, D. Lenz, M. Reinecke, C. Rosset, et al. (2019). healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python. Journal of Open Source Software 4(35), 1298.

Zou, H., X. Zhou, X. Fan, T. Zhang, Z. Zhou, et al. (2019). The Third Data Release of the Beijing–Arizona Sky Survey. The Astrophysical Journal Supplement Series 245(1), 4. arXiv:1908.07099.

Appendix

Contents

1 Perturbed energy stress tensor

The energy stress tensor for a non-perfect fluid is

$$T^{\mu\nu} = (\rho + P)\, u^\mu u^\nu + P\,(g^{\mu\nu} + \Pi^{\mu\nu})\,, \tag{A.1}$$

where $\Pi^{\mu\nu}$ is the contribution from anisotropic stress[1], called also anisotropic pressure. It is symmetric, traceless, such that $\Pi^{\mu 0} = 0$ and unitless in this definition. The perturbed stress-energy tensor in linear order, defined as the linear part of $\delta T^{\mu\nu} = T^{\mu\nu} - \bar{T}^{\mu\nu}$ with $\bar{T}^{\mu\nu}$ following (1.10), is

$$\delta T^{\mu\nu} = (\delta\rho + \delta P)\,\bar{u}^\mu \bar{u}^\nu + \left(\bar{\rho} + \bar{P}\right)(\delta u^\mu \bar{u}^\nu + \bar{u}^\mu \delta u^\nu) + \bar{P}\left(h^{\mu\nu} + \frac{\delta P}{\bar{P}}\bar{g}^{\mu\nu} + \Pi^{\mu\nu}\right), \tag{A.2}$$

where $\bar{\rho}$, $\bar{P}$, $\bar{u}^\mu$ are the unperturbed quantities of the underlying perfect fluid. Let derive the four-velocity $u^\mu = \bar{u}^\mu + \delta u^\mu$ in the Newtonian gauge (1.58). By isotropy, the underlying fluid is at rest and since $\bar{g}_{\mu\nu}\bar{u}^\mu\bar{u}^\nu = -1$,

$$\bar{u}^\mu = a^{-1}\delta^0_\mu. \tag{A.3}$$

[1]For instance, neutrinos develop anisotropic stress which cannot be neglected up to the matter-dominated area.

By construction u^μ respects also $g_{\mu\nu}u^\mu u^\nu = -1$. Therefore, in linear order

$$2\bar{g}_{\mu\nu}\bar{u}^\mu \delta u^\nu + h_{\mu\nu}\bar{u}^\mu \bar{u}^\nu = 0, \tag{A.4}$$

and then, $\delta u^0 = -a^{-1}\Phi$. Writing $\delta u^i = a^{-1}v^i$, the four-velocity becomes

$$u^\mu = a^{-1}\left(1 - \Phi, v^i\right). \tag{A.5}$$

The perturbed energy tensor (A.2) is currently decomposed into 2 scalars ($\delta\rho$, δP), one 3-vector v^i and one symmetric traceless 3-tensor Π^{ij}. Following the SVT decomposition, v_i can be separated into 1 scalar and 1 3-vector with two degrees of freedom, such that,

$$v_i = v_i^S + v_i^V$$

with $v_i^S = \partial_i v$ and $\nabla \cdot \vec{v}^V = 0$. Similarly, the anisotropic pressure Π_{ij} can be separated into 1 scalar Π, 1 3-vector Π_i that is divergence-free ($\Pi_{i,i} = 0$), and one 3-tensor with two degrees of freedom Π_{ij}^T, such that

$$\Pi_{ij} = \Pi_{ij}^S + \Pi_{ij}^V + \Pi_{ij}^T$$

with $\Pi_{ij}^S = \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right)\Pi$, $\Pi_{ij}^V = (\partial_j\Pi_i + \partial_i\Pi_j)$ and $\delta^{ik}\Pi_{ij,k}^T = 0$.

Hence, in the Newtonian gauge (1.58) and considering only the scalar modes, the perturbed energy tensor (A.2) simplifies as follows

$$\begin{aligned}
\delta T_{00} &= \bar{\rho}a^2(\delta + 2\Phi) \\
\delta T_{0i} &= -\bar{\rho}a^2(1+w)\partial_i v \\
\delta T_{ij} &= \bar{P}\left(h_{ij} + \frac{\delta P}{\bar{P}}\bar{g}_{ij} + \Pi_{ij}^S\right),
\end{aligned} \tag{A.6}$$

where δ is the usual contrast density $\delta = \delta\rho/\bar{\rho}$ and $\bar{P} = w\bar{\rho}$.

2 A Universe with multiple species

The Universe is composed of different species whose contributions to the stress-energy tensor are added $T_{\mu\nu} = \sum_I T_{\mu\nu}^I$. Hence, from (1.61), the perturbed quantities are

$$\delta\rho = \sum_I \delta\rho_I, \quad \delta P = \sum_I \delta P_I, \quad (\bar{\rho} + \bar{P})v^i = \sum_I (\bar{\rho}_I + \bar{P}_I)v_I^i, \quad \Pi^{ij} = \sum_I \frac{\bar{P}_I}{\bar{P}}\Pi_I^{ij}. \tag{A.7}$$

The Einstein equations involve the total fluid quantities and express the gravity generated by all the contributions together. However, each component can be described separately by its stress-energy conservation equations

$$\boldsymbol{\nabla}_\mu T_I^{\mu\nu} = Q_I^{\mu\nu}, \tag{A.8}$$

where $Q_I^{\mu\nu}$ represents the interaction term for each component. If the component I interacts with the others only through gravity *i.e.* there is no energy transfer, then $Q_I^{\mu\nu} = 0$. To satisfy the overall energy conservation equation (1.12), $\sum_I Q_I^{\mu\nu} = 0$. These conservation equations take the form of (1.65) and (1.66), but with an additional contribution to the right-hand side if necessary.

3 Perturbation evolution: beyond the linear regime

In Section 1.2.3.5, we described the evolution of the perturbations in the linear regime. During the cosmic evolution, these perturbations become too large to be well described by the linear theory, and this happens for small scales: linearity holds up to $k_{NL} \sim 0.1\ h\ \mathrm{Mpc}^{-1}$ at $z = 0$.

When we derive the perturbed tensor at linear order (A.2), we have neglected all the second order terms, especially $\delta\rho \times \delta u$. For smaller scales ($k > k_{NL}$) need a particular description.

The general description of non-relativistic matter perturbations in matter area is given by (1.102), and the linear regime by (1.103). In this regime, the k modes are not coupled, and they evolve independently.

Without neglecting any terms, (1.102) can be written in Fourier space as

$$\begin{aligned}
\delta'(\mathbf{k},\eta) + \theta(\mathbf{k},\eta) &= -\int \frac{\mathrm{d}^3k_1\mathrm{d}^3k_2}{(2\pi)^3}\delta_D^{(3)}\left(\mathbf{k}-\mathbf{k}_{12}\right)\alpha\left(\mathbf{k}_1,\mathbf{k}_2\right)\theta\left(\mathbf{k}_1,\eta\right)\delta\left(\mathbf{k}_2,\eta\right) \\
\theta'(\mathbf{k},\eta) + \mathcal{H}\theta(\mathbf{k},\eta) + \frac{3}{2}\Omega_m\mathcal{H}^2\delta(\mathbf{k},\eta) &= -\int \frac{\mathrm{d}^3k_1\mathrm{d}^3k_2}{(2\pi)^3}\delta_D^{(3)}\left(\mathbf{k}-\mathbf{k}_{12}\right)\beta\left(\mathbf{k}_1,\mathbf{k}_2\right)\theta\left(\mathbf{k}_1,\eta\right)\theta\left(\mathbf{k}_2,\eta\right)
\end{aligned} \tag{A.9}$$

where the kernels α and β are, noting $\mathbf{k}_{12} = \mathbf{k}_1 + \mathbf{k}_2$,

$$\alpha\left(\mathbf{k}_1,\mathbf{k}_2\right) = \frac{\mathbf{k}_{12}\cdot\mathbf{k}_1}{k_1^2}, \quad \beta\left(\mathbf{k}_1,\mathbf{k}_2\right) = \frac{\left(\mathbf{k}_{12}\cdot\mathbf{k}_2\right)\left(\mathbf{k}_1\cdot\mathbf{k}_2\right)}{k_1^2k_2^2} = \frac{k_{12}^2\mathbf{k}_1\cdot\mathbf{k}_2}{2k_1^2k_2^2}. \tag{A.10}$$

Beyond the linear regime, k modes do not evolve independently, they are coupled by gravity, as shown in the RHS of (A.9).

To solve these equations, one can use the perturbation method, known as the standard perturbation theory (SPT):

$$\delta = \sum_{n=1}^{\infty}\delta^{(n)}, \quad \theta = \sum_{n=1}^{\infty}\theta^{(n)}, \tag{A.11}$$

where δ^n, θ^n are solved iteratively with (A.9). For instance, $(\delta^{(1)}, \theta^{(1)})$ are the solution of the linear regime described by (1.103). Hence, (δ^n, θ^n) can be fully described by $\delta^{(1)}$ (Bernardeau et al. 2002)

$$\begin{aligned}
\delta^{(n)}(\mathbf{k},\tau) &= \int_{\mathbf{k}_1}\int_{\mathbf{k}_n}(2\pi)^3\delta_D\left(\mathbf{k}-\mathbf{k}_{1\cdots n}\right)F_n\left(\mathbf{k}_1,\cdots,\mathbf{k}_n,\tau\right)\prod_i^n\delta^{(1)}\left(\mathbf{k}_i,\tau\right) \\
\theta^{(n)}(\mathbf{k},\tau) &= -\mathcal{H}(\tau)f(\tau)\int_{\mathbf{k}_1}\int_{\mathbf{k}_n}(2\pi)^3\delta_D\left(\mathbf{k}-\mathbf{k}_{1\cdots n}\right)G_n\left(\mathbf{k}_1,\cdots,\mathbf{k}_n,\tau\right)\prod_i^n\delta^{(1)}\left(\mathbf{k}_i,\tau\right)
\end{aligned}, \tag{A.12}$$

where $\int_{\mathbf{k}} = 1/(2\pi)^3\int d^3\mathbf{k}$. F_n (*resp.* G_n) is called the density (*resp.* divergence) kernels. From the linear order

$$F_1 = 1 \quad G_1 = 1. \tag{A.13}$$

In matter-dominated area and neglecting dark energy (Einstein-De Sitter Universe), these kernels do not depend on time and are easy to compute. For instance, the second-order kernels are

$$\begin{aligned}
F_2\left(\mathbf{k}_1,\mathbf{k}_2\right) &= \frac{5}{7} + \frac{2}{7}\frac{\left(\mathbf{k}_1\cdot\mathbf{k}_2\right)^2}{k_1^2k_2^2} + \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right) \\
G_2\left(\mathbf{k}_1,\mathbf{k}_2\right) &= \frac{3}{7} + \frac{4}{7}\frac{\left(\mathbf{k}_1\cdot\mathbf{k}_2\right)^2}{k_1^2k_2^2} + \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{2k_1k_2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)
\end{aligned}. \tag{A.14}$$

In this configuration, $(\delta^{(n)}, \theta^{(n)})$ will be completely described by the linear growth function $D(\tau)$ which then encodes the entire cosmology dependence of the non-linear structure formation. For simplicity, the SPT kernels are usually computed in the EdS Universe and used for all cosmologies.

Since $\langle\delta(\mathbf{k}, z)\delta(\mathbf{k}', z)\rangle = (2\pi)^3 P(\mathbf{k}, z)$, the matter power spectrum at first order is given by

$$P(\mathbf{k}, z) = P_{lin}(\mathbf{k}, z) + P_{22}(\mathbf{k}, z) + 2P_{13}(\mathbf{k}, z), \tag{A.15}$$

where

$$\begin{aligned} P_{22}(\mathbf{k}) &= \int \frac{\mathrm{d}^3 q}{(2\pi)^3} P_{lin}(\mathbf{q}) P_{lin}(\mathbf{k}-\mathbf{q}) \left|F_2(\mathbf{q}, \mathbf{k}-\mathbf{q})\right|^2 \\ P_{13}(\mathbf{k}) &= 3P_{lin}(\mathbf{k}) \int \frac{\mathrm{d}^3 q}{(2\pi)^3} P_{lin}(\mathbf{q}) F_3(\mathbf{k}, \mathbf{q}, -\mathbf{q}). \end{aligned} \tag{A.16}$$

Similar computation can be done for the high-order statistics, in particular, the mode coupling will create a non-zero bispectrum that is in the Gaussian case given by

$$B\left(\mathbf{k}_1, \mathbf{k}_2\right) = 2F_2\left(\mathbf{k}_1, \mathbf{k}_2\right) P\left(k_1\right) P\left(k_2\right) + \text{perm.} \tag{A.17}$$

This part of the bispectrum is usually called the gravitational contribution in order to separate it from the potential primordial non-gaussianities which are also encoded in the bispectrum.

Finally, the accuracy of SPT for the matter power spectrum and bispectrum has been extensively tested using N-body simulations that emulate the gravity, see for instance Jeong and Komatsu (2006) which shows that the accuracy of STP increases when the non-linearity is less prone, typically at high redshift ($z > 1$). Several alternative methods, also based on perturbative development, have been proposed in recent years to increase the accuracy of perturbative development without increasing the order used. See, for instance, the effective field theory of large-scale structures (EFFofLSS, Carrasco et al. 2012, Ivanov 2023).

4 Einstein equations in different gauges

As described in Section 1.2.2.1, at linear order, the scalar, vector and tensor perturbations evolve independently.

4.1 Scalar perturbations

The perturbed metric for the scalar modes reads

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Phi)\mathrm{d}\eta^2 + 2B_{,i}\mathrm{d}x^i\mathrm{d}\eta + \left[(1-2\Psi)\delta_{ij} + 2E_{,ij}\right]\mathrm{d}x^i\mathrm{d}x^j\right], \tag{A.18}$$

where $B_{,i} = \partial B/\partial x^i$ and $E_{,ij} = \partial^2 E/\partial x^i \partial x^j$.

They are two scalar gauge transformations

$$\begin{aligned} \eta &\to \eta + \alpha \\ x^i &\to x^i + \partial^i \beta' \end{aligned} \tag{A.19}$$

which lead to the following scalar metric perturbation transformations

$$\begin{aligned} \Phi &\to = \Phi - \alpha' - \frac{a'}{a}\alpha \\ B &\to B - \beta' + \alpha \\ \Psi &\to \Psi + \mathcal{H}\alpha \\ E &\to E - \beta \end{aligned}, \tag{A.20}$$

The physics is encoded in a gauge-invariant combination of the scalar metric perturbations. The two scalar gauge-invariant quantities, introduced by Bardeen (Bardeen 1980), are

$$\begin{aligned}\Phi_{\rm B} &\equiv \Phi + \mathcal{H}\left(B - E'\right) + \left(B - E'\right)' \\ \Psi_{\rm B} &\equiv \Psi + \mathcal{H}(E' - B)\end{aligned}. \tag{A.21}$$

Similarly, the perturbed fluid quantities are not invariant under the above gauge transformations. δ, δP and v are transformed as

$$\delta\rho \to \delta\rho - \bar{\rho}'\alpha, \quad \delta P \to \delta P - \bar{P}'\alpha, \quad v \to v + \beta'. \tag{A.22}$$

Remark: The perturbed fluid quantities depend on the choice of the gauge. They will be noted with an additional superscript. For instance, the Newtonian gauge quantities are written δ^N, v^N.

4.1.1 Einstein equations

The computation of the perturbed Einstein tensor and the perturbed stress-energy tensor leads to the perturbed Einstein equations in gauge-invariant quantities

$$\begin{aligned}&\Delta\Psi_{\rm B} = 4\pi G a^2 \bar{\rho}\delta^C \\ &\Psi_{\rm B} - \Phi_{\rm B} = 8\pi G a^2 \bar{P}\Pi \\ &\Psi_{\rm B}' + \mathcal{H}\Phi_{\rm B} = -4\pi G a^2 \bar{\rho}(1+w)v^N \\ &\mathcal{H}X' + \left(\mathcal{H}^2 + 2\mathcal{H}'\right)X = 4\pi G a^2 \bar{\rho}\left(w\Gamma + c_s^2\delta^F + \frac{2}{3}w\Delta\Pi\right)\end{aligned}, \tag{A.23}$$

where $X = \Psi_{\rm B} + \Phi_{\rm B} + (\Psi_{\rm B}/\mathcal{H})'$, δ^C is the comoving density perturbations (A.37), δ^F is the spatially-flat density perturbations (A.33) and v^N is the Newtonian velocity perturbations (A.27). The entropy perturbations Γ is defined as $\delta P = c_{\rm S}^2\delta\rho + P\Gamma$ such that for an adiabatic perturbation, $\delta P/\delta\rho = c_s^2$ and therefore $\Gamma = 0$. Note that Γ and Π are gauge-invariant[2].

Similarly, one can compute the continuity equation and the Euler equation for fluid in gauge-invariant quantities

$$\begin{aligned}&\delta^{C\prime} - 3\mathcal{H}w\delta^C = (1+w)\Delta^2 v^N + 2\mathcal{H}w\Delta^2\Pi \\ &v^{N\prime} + \mathcal{H}v^N = -\Phi - \frac{c_s^2}{1+w}\delta^C - \frac{2}{3}\frac{w}{1+w}\Delta^2\Pi\end{aligned}, \tag{A.24}$$

where Φ is in the newtonian gauge.

4.1.2 Newtonian gauge

The Newtonian gauge reduces to Newtonian gravity in the classical limit, and so it is very popular for analytic computation since the scalar modes can be understood as the classical gravity. The Newtonian gauge fixes

$$B = E = 0, \tag{A.25}$$

such that the perturbed metric becomes

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1+2\Phi)\mathrm{d}\eta^2 + (1-2\Psi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{A.26}$$

[2] Π is gauge-invariant thanks to the Stewart-Walker lemma and $\Gamma \to \Gamma + \left(\bar{P}' - c_s^2\bar{\rho}'\right)\alpha = \Gamma$

The Newtonian density perturbations and velocity perturbations

$$\delta^N = \delta + \frac{\rho'}{\rho}(B - E'), \quad v^N = v + E' \tag{A.27}$$

are gauge-invariant.

Remark: In this gauge, $(\Psi_\mathrm{B}, \Phi_\mathrm{B}) = (\Psi, \Phi)$ and that is why we have used these notations in the perturbation of the metric since in most cases we will work with the Newtonian gauge, although it can be confusing in general. In the following, (Ψ, Φ) will always refer to their expression in the Newtonian gauge unless otherwise stated.

4.1.3 Synchronous gauge

The synchronous gauge is used in the numerical implementation of the perturbation equations as in many Boltzmann solvers. `CLASS`[3] can solve the perturbed equations in both Newtonian and Synchronous gauges. The synchronous gauge fixes

$$\Phi^S = B^S = 0, \tag{A.28}$$

such that the perturbed metric becomes

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + \left[(1 - 2\Psi^S)\delta_{ij} + 2E^S_{,ij}\right]\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{A.29}$$

4.1.4 Spatially-flat gauge

The spatially-flat gauge is generally used to compute the inflationary perturbation, such that the scalar perturbations are described by $\delta\phi$. The spatially-flat gauge fixes

$$\Psi^F = E^F = 0, \tag{A.30}$$

such that the perturbed metric becomes

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1 + 2\Phi^F)\mathrm{d}\eta^2 + 2B^F_{,i}\mathrm{d}x^i\mathrm{d}\eta + \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]. \tag{A.31}$$

The two remaining scalars can be expressed as

$$\begin{aligned} \Phi^F &= \Phi + \Psi + (\Psi/\mathcal{H})' \\ B^F &= B - E' - \Psi/\mathcal{H} \end{aligned}, \tag{A.32}$$

where Φ, Ψ, B, E are from (A.18).

The spatially-flat density perturbations

$$\delta^F = \delta + \frac{\rho'}{\mathcal{H}}\Psi \tag{A.33}$$

is gauge-invariant.

[3]Note that `CLASS` follows the convention of Ma and Bertschinger (1995). In particular, the Newtonian gauge is defined as

$$\mathrm{d}s^2 = a^2(\eta)\left[-(1 + 2\Psi)\mathrm{d}\eta^2 + (1 - 2\Phi)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right],$$

swapping $\Psi \leftrightarrow \Phi$ in our notation (A.26).

4.1.5 Comoving gauge

Known also as the comoving orthogonal gauge, it is defined such that the coordinate system is comoving with the fluid flow *i.e.* requiring a zero 3-velocity and the orthogonality of constant time hypersurfaces to 4-velocity. This gauge is then,

$$v^C = B^C = 0. \tag{A.34}$$

This can be achieved with the following gauge transformation

$$\alpha = -(v + B), \quad \beta = -\int v d\tau + \mathcal{C}(\mathbf{x}), \tag{A.35}$$

where $\mathcal{C}(\mathbf{x})$ is the residual gauge freedom in shifts of the spatial coordinates. Hence, the scalar quantities are

$$\begin{aligned} \Phi^C &= \Phi + \mathcal{H}(B + v) + (B + v)' \\ \Psi^C &= \Psi - \mathcal{H}(B + v) \\ E^C &= E + \int v d\tau - \mathcal{C} \end{aligned} . \tag{A.36}$$

The comoving density perturbations

$$\delta^C = \delta + \frac{\rho'}{\rho}(v + B) = \delta - 3\mathcal{H}(1 + w)(v + B) \tag{A.37}$$

is gauge-invariant. One can introduce also another very useful gauge-invariant quantity called the *comoving curvature perturbation*,

$$\mathcal{R} \equiv -\Psi^C = -\Psi + \mathcal{H}(B + v), \tag{A.38}$$

which can be expressed only as a function of the metric in the Newtonian gauge ($B = 0$, $E = 0$)[4]

$$\mathcal{R} = -\Psi + \frac{\mathcal{H}(\Psi' + \mathcal{H}\Phi)}{\mathcal{H}' - \mathcal{H}^2} = -\Psi - \frac{2}{3\mathcal{H}}\frac{(\Psi' + \mathcal{H}\Phi)}{1 + w}, \tag{A.39}$$

where from (1.28), we have $3/2(1 + w) = 1 - \mathcal{H}'/\mathcal{H}^2$. Taking the conformal time derivative with some simplifications from the background evolution[5] leads to

$$\mathcal{R}' = -\frac{2\mathcal{H}^{-1}}{3(1 + w)}\Psi'' - \frac{4 + 6c_s^2}{3(1 + w)}\Psi' - \frac{2}{3(1 + w)}\Phi' + 2\mathcal{H}\frac{w - c_s^2}{1 + w}\Phi. \tag{A.40}$$

Then, the Einstein equations (1.62) give

$$\mathcal{R}' = -\frac{2}{3}\frac{\mathcal{H}}{1 + w}\left[-c_s^2\left(\frac{k}{\mathcal{H}}\right)^2\Psi + \frac{1}{3}\left(\frac{k}{\mathcal{H}}\right)^2(\Phi - \Psi) + \frac{3}{2\bar{\rho}}\left(\delta P - c_s^2\delta\rho\right)\right]. \tag{A.41}$$

Finally, for super-Hubble modes ($k \ll \mathcal{H}$) and adiabatic evolution ($\Gamma = 0 \to \delta P = c_s^2\delta\rho$),

$$\mathcal{R}' = 0. \tag{A.42}$$

Since the expansion of the Universe is an adiabatic process, the comoving curvature perturbation $\mathcal{R}$ **stays constant** outside the horizon!

[4] From Einstein equations, we have $\Psi' + \mathcal{H}\Phi = 4\pi G a^2\left(\bar{\rho} + \bar{P}\right)v^N$ and combining the Friedmann equations gives $4\pi G a^2\left(\bar{\rho} + \bar{P}\right) = \mathcal{H}^2 - \mathcal{H}'$.

[5] $w' = -3\mathcal{H}(1 + w)(c_s^2 - w)$ and $\mathcal{H}' = -1/2\mathcal{H}^2(1 + 3w)$

Neglecting the anisotropic pressure ($\Pi = 0$), $\Psi = \Phi$ and (A.39) becomes

$$\mathcal{R} = -\Phi - \frac{2}{3\mathcal{H}} \frac{(\Phi' + \mathcal{H}\Phi)}{1+w}. \tag{A.43}$$

For adiabatic perturbations at super-Hubble modes ($\mathcal{R} = const.$), (A.43) is a differential equation for Φ during the period where $w = const.$:

$$\frac{2}{3}\mathcal{H}^{-1}\Phi' + \frac{5+3w}{3}\Phi = -(1+w)\mathcal{R}. \tag{A.44}$$

The solution is the sum of the particular and the homogeneous solution

$$\Phi = -\frac{3+3w}{5+3w}\mathcal{R} + Ca^{-\frac{5+3w}{2}}. \tag{A.45}$$

If $w = const.$ for a long time enough *i.e.* a sufficiently high value of a, one can neglect the second term and

$$\Phi = -\frac{3+3w}{5+3w}\mathcal{R} = const. \tag{A.46}$$

In particular, this relation holds for a radiation-dominated Universe and a matter-dominated Universe. Hence, for adiabatic perturbations,

$$\begin{array}{ll} \Phi_{\mathbf{k}} = -\dfrac{2}{3}\mathcal{R}_{\mathbf{k}} & \text{(radiation-dominated)} \\ \Phi_{\mathbf{k}} = -\dfrac{3}{5}\mathcal{R}_{\mathbf{k}} & \text{(matter-dominated)} \end{array}, \quad (k \ll \mathcal{H}). \tag{A.47}$$

Even if this question does hold during the radiation-matter transition, outside the horizon $\mathcal{R}_{\mathbf{k}}$ remains constant and since $\Phi_{\mathbf{k}}$ changes from $-\frac{2}{3}\mathcal{R}_{\mathbf{k}}$ to $-\frac{3}{5}\mathcal{R}_{\mathbf{k}}$, $\Phi_{\mathbf{k}}$ changes by a factor 9/10.

Remark: Several conventions exist to define the comoving curvature perturbation, some (see, for instance, Peter and Uzan 2013) prefer to use the name ζ defined as

$$\zeta = -\mathcal{R}. \tag{A.48}$$

It can be useful to define another gauge-invariant quantity called the *curvature perturbation in the flat-slicing gauge* (known also as the spatially-flat gauge) and defined as

$$\zeta_{\mathrm{BST}} = \Psi + \frac{\delta\rho}{3(\rho+P)} = \Psi + \frac{\delta}{3(1+w)}. \tag{A.49}$$

This quantity can be related to the comoving curvature perturbation via

$$\zeta = \zeta_{\mathrm{BST}} - \frac{\Delta\Phi}{3\left(\mathcal{H}' - \mathcal{H}^2\right)}. \tag{A.50}$$

Hence, for super-Hubble modes ($\Delta \propto k^2$), these two quantities are equal: $\zeta \underset{k\ll\mathcal{H}}{=} \zeta_{\mathrm{BST}}$.

4.2 Vector perturbations

The perturbed metric for the vector modes reads

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + 2B_i\mathrm{d}x^i\mathrm{d}\eta + (\delta_{ij} + 2E_{i,j} + 2E_{j,i})\,\mathrm{d}x^i\mathrm{d}x^j\right], \tag{A.51}$$

where $B_{i,i} = E_{i,i} = 0$.

There is one vector gauge transformation

$$x^i \to x^i + \beta^i, \quad \beta_{i,i} = 0, \tag{A.52}$$

which lead to the following vector metric perturbation transformations

$$\begin{aligned} B_i &\to B_i + a\dot{\beta}_i \\ E_i &\to E_i - \beta_i \end{aligned}, \tag{A.53}$$

The gauge-invariant quantity is $\dot{E}_i + B_i/a$ and is known as the *gauge-invariant vector shear perturbation.*

The associated perturbed Einstein equations are

$$\begin{aligned} \delta\dot{q}_i + 3H\delta q_i &= k^2\Pi_i \\ k^2\left(\dot{E}_i + B_i/a\right) &= 16\pi G\delta q_i \end{aligned}, \tag{A.54}$$

where $\delta q^i = (\bar{\rho} + \bar{p})v^i$ is the 3-momentum density, see Appendix 2, and Π_i is the vector part of the anisotropic pressure.

Remark: In the Newtonian gauge, B_i is set to 0.

Finally, in the absence of anisotropic stress ($\Pi_i = 0$), the first Einstein equation becomes $\delta\dot{q}_i + 3H\delta q_i = 0$ such that δq_i decays with the expansion of the Universe. Hence, the shear perturbation $\dot{E}_i + B_i/a$ vanishes with the expansion of the Universe. Vector modes will not play an important role in large-scale structure formation, especially as they are not created by inflation.

4.3 Tensor perturbations

The perturbed metric for the tensor modes reads

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + (\delta_{ij} + 2E_{ij})\,\mathrm{d}x^i\mathrm{d}x^j\right], \tag{A.55}$$

where E_{ij} is a traceless ($E^i_i = 0$), symmetric ($E_{ij} = E_{ji}$) and divergence-free ($k^iE_{ij} = 0$) tensor. This tensor is gauge invariant in linear order.

The Einstein equations for the tensor perturbations are

$$E''_{ij} + 2\mathcal{H}E'_{ij} - \Delta E_{ij} = 8\pi G a^2 \bar{P}\Pi^T_{ij}, \tag{A.56}$$

where Π^T_{ij} is the tensor part of the anisotropic pressure. For a perfect fluid or scalar field, $\Pi^T_{ij} = 0$.

E_{ij} is fully depicted by only two independent components corresponding to the polarization modes of the gravitational waves. Several settings exist for the polarization (see, Isi 2022). For instance, the *linear polarization* sets the propagation of the wave along the $\hat{z}$ axis, such as

$$E_+ \equiv E_{11} = -E_{22}, \quad E_\times \equiv E_{12} = E_{21}, \tag{A.57}$$

where the spatial basis is form to be right-handed with $\hat{x}$ and $\hat{y}$ two arbitrary orthonormal vectors that, with $\hat{z}$. The gravitational waves E_{ij} can be therefore expanded as

$$E_{ij}(\mathbf{k}) = E_+ e^+_{ij} + E_\times e^\times_{ij} \tag{A.58}$$

where $e_{ij}^{+,\times}$ are the *polarization tensors*, defined by

$$e_{ij}^{+} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad e_{ij}^{\times} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{A.59}$$

This frame is constructed to have $\mathbf{k}$ aligned with $\hat{z}$ and therefore the polarization tensors depend implicitly on $\mathbf{k}$.

5 Validation of imaging systematic mitigation method

5.1 Validation with mocks

To validate the systematic mitigation method and avoid overfitting, we use a set of 100 QSO EZ-mocks from eBOSS (Zhao et al. 2021). They have a smaller density than the nominal target selection density and a smaller area. However, it will not impact the analysis since the surface and the density are enough for our test. Besides, since our analysis for the legacy survey is different for the three photometric footprints, we will mimic one footprint only, similar to a sub-region of the North footprint.

The mocks contain a cosmological signal but no systematic effects. They are contaminated using the inverse weight estimated previously and this contamination will be mitigated with our method using the same observational features.

We present here contamination estimated with the RF method (it is the inverse of the map shown in Fig. 4.32). We test the method in two different ranges of contamination: one describing a weakly contaminated case (extracted from the R.A., Dec. $\in [100°, 270°] \times [32°, 60°]$ box) where all the methods work well; a second describing a strongly contaminated case (extracted from the R.A., Dec. $\in [120°, 290°] \times [55°, 90°]$ box) where the linear method is inefficient as shown in the systematics plots of Fig. 4.28. To compare the method efficiency and avoid biasing the results, we also applied an NN-based estimation for the strongly contaminated case.

The validation pipeline is shown in Fig. A5.1. There are two different tests (dashed lines). The first one is to check whether our method overfits the data. We apply the mitigation method to the initial (uncontaminated) mocks such that no correction is expected. The angular correlations of the uncontaminated mocks and of the corrected uncontaminated mocks should be identical. The second test is to validate our mitigation. We apply the mitigation method to the contaminated mocks, expecting to recover the same angular correlation as for the initial mocks.

Some subtleties have to be considered. First, the density of these mocks (70 deg^{-2}) is lower than that of the quasar targets (300 deg^{-2}). Thus, the size of the correction cannot be the same since the number of objects inside a pixel of $N_{side} = 256$ is too small. We, therefore, downgrade the contamination and the observational feature maps to $N_{side} = 128$.

Secondly, the footprint of these mocks is different from the North footprint. The size of the training sample is here twice smaller as the data. We, therefore, use only 3 folds instead of 6 for the DR9 North training. It is worth noting that we need to change the position of the folds to cover the footprint correctly as explained in 4.3.3.3, especially for the strongly contaminated case.

Nevertheless, the training conditions are slightly different and we expect some differences for the second test.

The angular correlation functions for the weakly contaminated case are shown in Fig. A5.2a. Each correlation function is the average of the angular correlations of the 100 mocks and the error

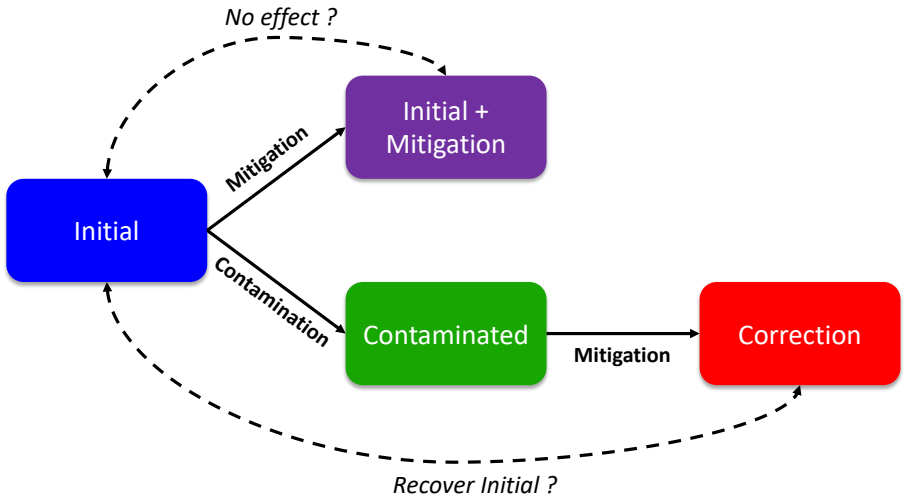


Figure A5.1: Pipeline validation with mocks. The initial mocks (blue) are uncontaminated and contain a cosmological signal. They are contaminated with systematic weights (green). Then, they are corrected with the mitigation presented above (red). We compare the corrected mock to the initial mock to verify if we recover the correct initial state. As a sanity check, we also applied the mitigation to the initial mock (purple) to verify whether our mitigation technique overfits target density variations. These tests are represented by the two black dashed lines.

bars are the standard deviation between all the realizations. We show the correlation function up to $\theta \sim 3.5$ deg, which corresponds to a comoving angular distance of ~ 250 Mpc h^{-1} at $z = 1.7$, the maximum size of the eBOSS analysis. The result for each regression method is shown in the three panels.

The purple lines are the angular correlations of the initial mocks mitigated with our methods. In the three cases, they lie on top of the blue lines that show the angular correlations of the initial mocks. This indicates no sign of overfitting compared to the significance of the correction. The green lines are the angular correlations for the mocks after contamination. The red lines are the angular correlations for the contaminated mocks after mitigation. The red line exactly recovers the blue line in the case of the RF method. In the two other cases, the red lines recover more or less the blue lines and the slight difference stems from the contamination was done from an estimation of the systematics by the RF method which benefits the RF case.

The angular correlations for the strongly contaminated case are shown in Fig. A5.2b for the RF-based contamination and in Fig. A5.2c for the NN-based one. The two figures are similar, indicating that the contaminated method does not benefit the RF or NN mitigation. The result from the same contamination method can be compared together.

Here again, there is no indication of overfitting for the NN and Linear method. For the RF method, the purple line is slightly below the blue one above 1 deg indicating tiny overfitting. Note that the mocks that we used here cover a smaller area than the DESI QSO sample, thereby making our method more prone to overfit for this test case. Optimization of the hyper-parameters to reduce overfitting on the DESI footprint will be pursued with DESI mocks. However, this tiny overfitting is much less significant than the level of systematic corrections validating that most of the impact of RF weights on the angular correlation function of the DESI QSO sample is real systematics mitigation and not overfitting.

The angular correlation of contaminated mocks is higher as a consequence of the stronger contamination compared to the previous case. The red lines do not perfectly recover the blue lines even in the case of the RF regression. Indeed, the smaller size of the training sample and the

modification of the fold form prevent the NN and RF methods to learn all the information needed to fully correct this contamination. In addition, the NN method is less efficient to correct small contaminated regions without any additional hyper-parametrization like regularization terms. The RF approach is more robust than the NN one to variations in the training sample and less dependent on hyper-parametrization. For the linear method, the correction is less efficient: in the region of extreme observational features, the systematic effects are less corrected, as shown in Section 4.3.4.1.

This set of mocks and the second test are used to optimize the hyper-parameters of the NN and perform the grid search method (4.3.3.3). Given the subtleties presented above, however, it is expected that the hyper-parameters used here are not the best ones for the DR9 training. More optimized parameter tuning will eventually be achieved with mocks matching the DESI QSO samples.

5.2 South footprint and Sagittarius Stream contamination

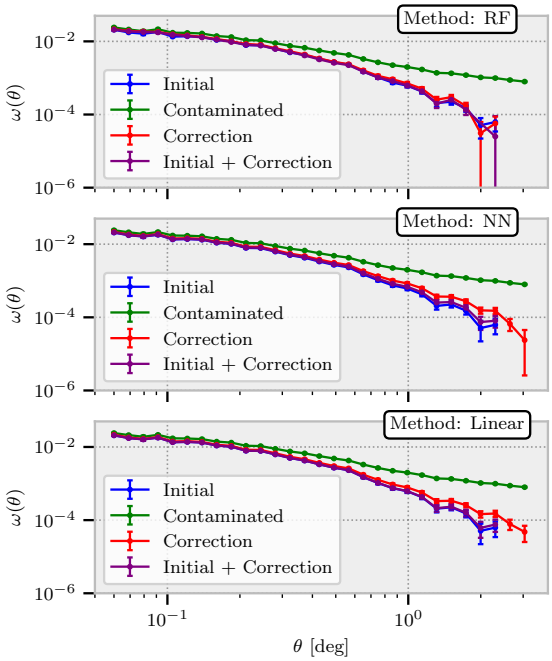
As shown in Fig. 4.35, even after the systematic mitigation, we do not recover the same level of correlation in the South footprint as in the two other regions. To analyze the excess of correlation in the South footprint, we divide it into 4 zones, represented in Fig. A5.3. These zones are:

- Zone 1 near the anti-galactic pole shows less over-density than the other regions near the Galactic plane. Its lower density is due to the lower value of the PSF Depth $W2$ in this zone.
- Zone 2 contains the Sagittarius Stream and it is strongly contaminated by stars.
- Zone 3 and zone 4 show strong over-density due to stars from the galactic plane. In addition, zone 4 describes the SGC part of the South footprint and we considered also zone 13 combining zone 1 with zone 3 which describes the NGC without the Sagittarius Stream part of the South.

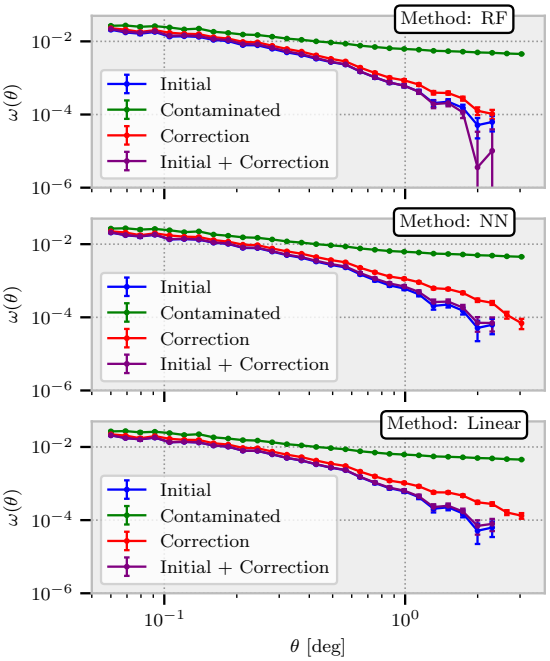
The angular correlation functions for these different zones are shown in Fig. A5.4. Here each mitigation is performed with the RF method. The top panel shows the correction done with training on all the South footprints. The bottom panel shows the correction where the training was done individually on each considered zone. We add also the angular correlation for the corrected targets in South (in green) and in DES (in black) for reference. The correlations of the corrected targets on each zone are lower than on all the South footprint but it is not the average on zones due to the missing cross-terms.

The excess of correlation at small scales cannot be removed by our method. It is due to either stars or non-considered features. The part caused by a non-considered feature is left for a future study with the spectroscopic sample. The excess caused by the stellar contamination will be removed with the spectroscopic data as explained in Sec. 4.3.5.2. Hence, we will discuss the excess of correlation at large scales compared to the correlation in DES.

The correction estimated with the training on all South footprint is not sufficient to recover the same level of correlation as in DES after mitigation for all zones. The correction in zone 2 is more efficient than in the other zones since the Sagittarius Stream feature separates the Sagittarius Stream from the rest during the training. This zone does not recover the same angular correlation at large scales as in DES since the Sagittarius Stream feature is built as the spatial average of candidate stars suppressing the angular correlation information contained in this feature.

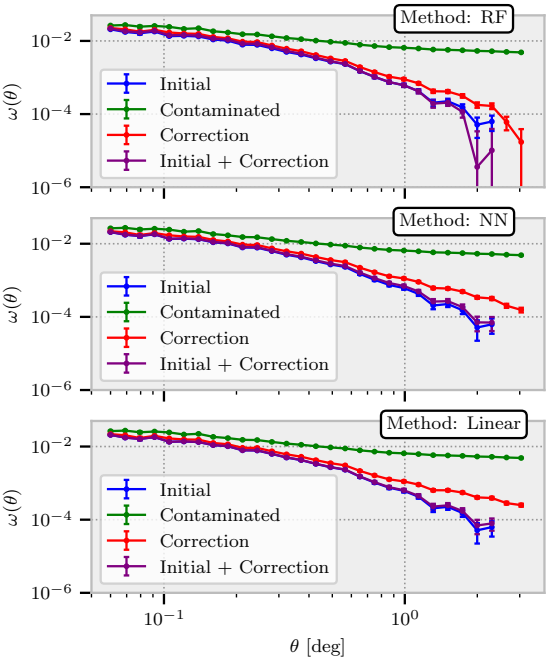


(a) Weakly contaminated case



(b) Strongly contaminated case

Figure A5.2: (a) Mean on 100 eBOSS EZ-mocks of the angular correlation functions. The error bars are the standard deviation between all realizations. Colors follow the same scheme as in Fig. A5.1. The systematic contamination is done with an RF estimation of a weakly contaminated area (R.A., Dec. $\in [100, 270] \times [32, 60]$) of the DR9 legacy survey. The method does not overfit the data and recovers the initial correlation. Note that the contamination was estimated with the RF method which explains the small residual systematics with the NN and Linear correction. (b) The systematic contamination depicts a strongly contaminated area (R.A., Dec. $\in [120, 290] \times [55, 90]$). The correction is not perfect since the training sample is smaller than the Legacy Surveys case and cannot contain the same information in each fold. Besides, the correction is done with a higher N_{side} which can explain why we do not fully recover the initial state. (c) Same as (b) but the contamination is estimated with the NN method.



(c) Strongly contaminated case

The angular correlation in zone 1 without systematic mitigation is lower than those in the three other zones since this region does not show strong over-density. However, the systematic mitigation is inefficient in this region when the training is done on all the South footprint. This is not the case when we apply the mitigation from the training only in zone 1. Our method is unable to extract the correct information for this zone in all of the South. Zone 4 and zone 13 are better when the training is performed only on each zone, recovering a level of correlation

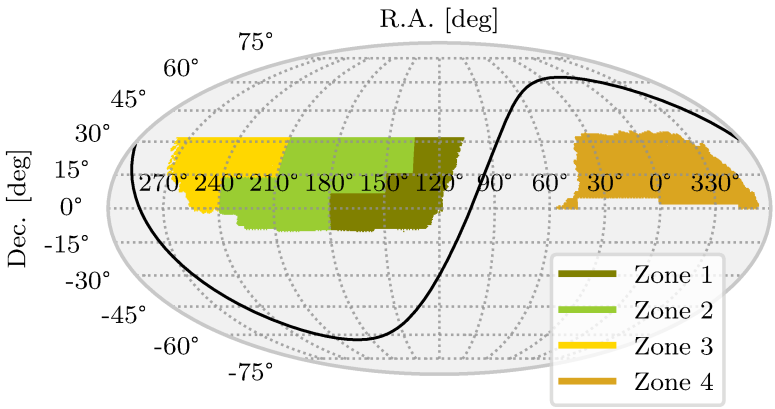


Figure A5.3: The South region is split into four distinct zones to analyze the stellar contamination. Zone 1 near the anti-galactic pole shows less overdensity than the three others. Zone 2 contains the Sagittarius Stream. Zone 3 and Zone 4 show strong overdensity due to stars from the galactic plane.

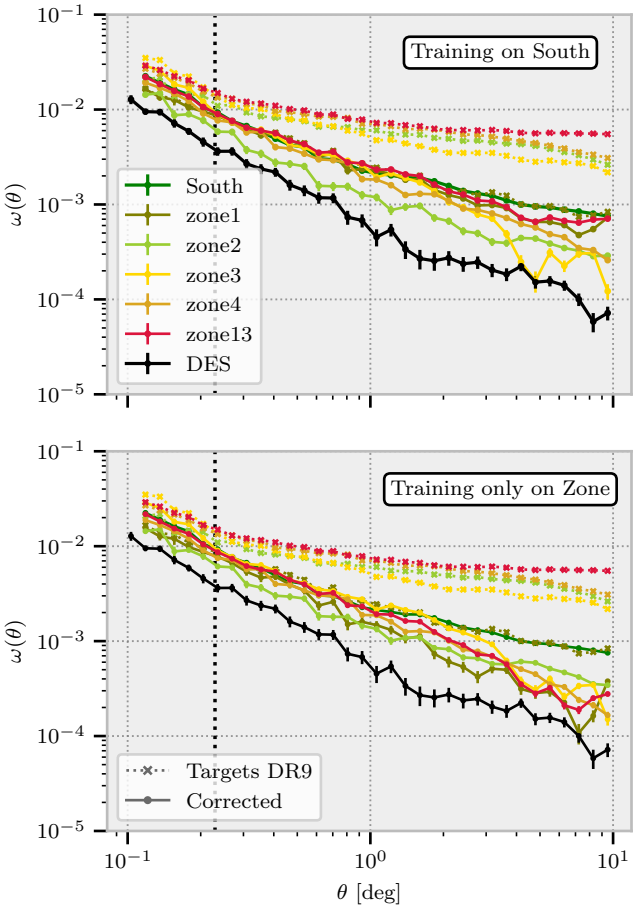


Figure A5.4: The angular correlation for the four zones which are represented in Fig. A5.3. The dotted lines are the correlations of DR9 targets, the solid ones are for the corrected targets. For comparison, we also plot the correlation in the South region (green line) and in the DES region (black line) which is known as the least contaminated region. On the top panel, corrections are calculated on all the South footprints. On the bottom panel, corrections are calculated in each zone independently.

that is in more agreement with that of DES at large scales.

The inefficiency of the regression in the South can be explained by the numerous stars in this region which bias the information about the observational systematic effects preventing the regression to learn correctly the true observational features dependency of the relative target density. For instance, the PSF Depth $W2$ is not characterized as an important feature (cf. Sec. 4.3.4.2) while it explains the lower density observed around the anti-galactic pole. Fig. A5.5 shows the feature importance estimated with the training on the considered zone only. The feature importances for all the zone used are plotted. PSF Depth $W2$ is almost uniform (cf. Fig. 4.25 for its distribution) in zone 1 and zone 3 and so does not appear as an important feature when the training is done in each region. However, it is found as the most important feature in the training on zone 13 illustrating that the stars in zone 2 and zone 4 bias the training, *e.g.* the fluctuation of the relative density in the function of the PSF Depth $W2$ in zone 2 is masked by the presence of numerous stars from the Sagittarius Stream.

Even if the training in an individual zone is more effective, especially at large scales than the training in all the South, we do not recover the same level of correlation as in DES at intermediate scales. This can be explained still by the presence of numerous stars in each zone. The stellar contamination is mixed with the imaging systematics and the two effects are not easily separable. Then, the excess of correlation can be caused either by the stellar contaminant or by an additional unconsidered imaging feature.

This analysis could be performed once the spectroscopic survey is done since the stellar contamination will vanish. The impact of imaging features can be then studied without any significant bias. If the set of features introduced during this analysis contains all the information, our systematic mitigation method will be able to correctly learn the true dependence on the observational features as in DES or the North and recover the correct angular correlation.

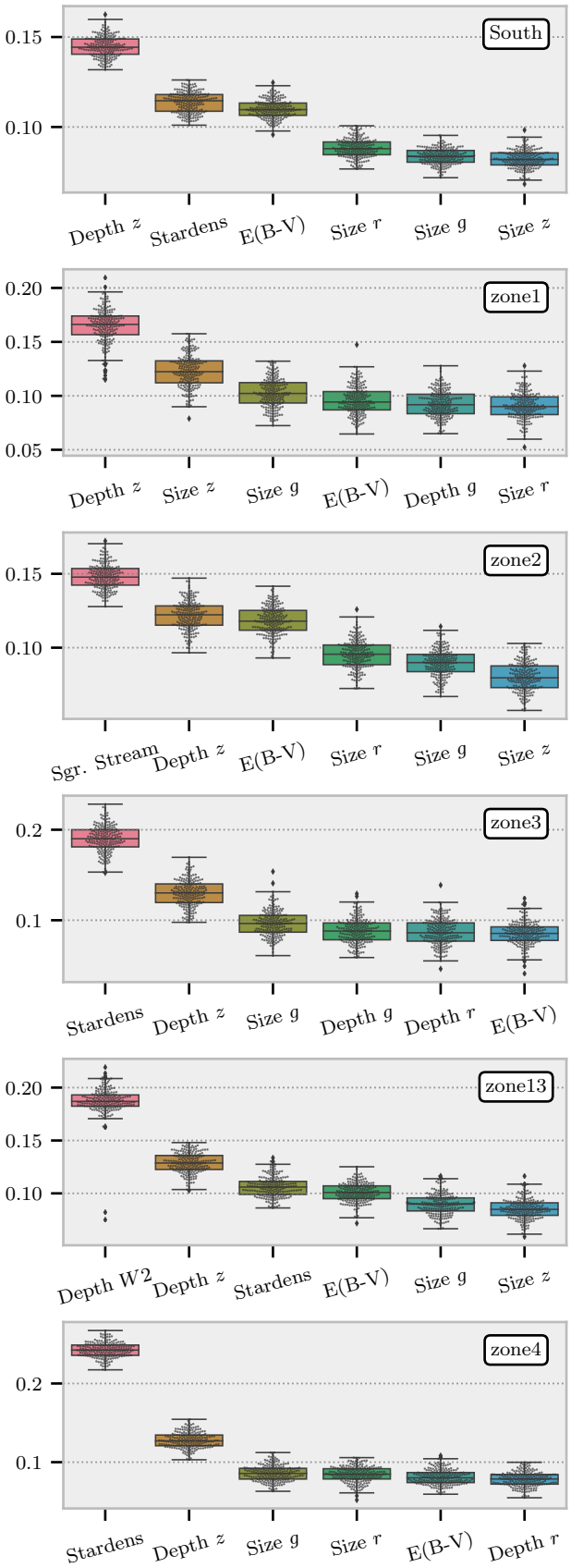


Figure A5.5: Same as Fig. 4.31 but the feature importances plotted are estimated with the training only on the considered zone. The different behaviors in each zone are highlighted by the different features which are qualified as important. In particular, the density of stars from the galactic plane prevents the regression to learn correctly the role of the PSF Depth $W2$ as expected.

Résumé

Contents

Rappelez-vous de regarder vers les étoiles et pas vers vos pieds. Essayez de donner un sens à ce que vous voyez et demandez-vous ce qui fait que l'Univers existe. Soyez curieux.

Stephen Hawking

La cosmologie, étude de l'Univers à ses plus grandes échelles, cherche à percer les mystères de son origine, de son évolution et de son destin. Elle englobe un large éventail de sujets, notamment la théorie du Big Bang, la structure et le comportement des structures à grande échelle dans l'Univers, et la nature de la matière noire et de l'énergie noire.

La naissance de la cosmologie "moderne" remonte aux travaux novateurs d'Albert Einstein et à sa théorie de la relativité générale en 1915, qui a révolutionné notre compréhension de la gravité en proposant un espace-temps dynamiquement influencé par la matière et l'énergie. L'expansion de notre Univers a été prédite quelques années plus tard par Georges Lemaître et Alexander Friedmann, qui ont résolu indépendamment les équations d'Einstein. La confirmation observationnelle, connue sous le nom de loi de Hubble-Lemaitre, a été l'un des tournants les plus importants de la physique moderne.

Longtemps remise en question, la théorie du Big Bang, qui décrit un Univers en expansion, a été largement adoptée après la découverte en 1964 par Arno Penzias et Robert Wilson du rayonnement de fond cosmologique. Cette découverte a confirmé de manière irréfutable que notre Univers était plus chaud et plus dense à ses débuts. Le fond diffus cosmologique a été un objet d'étude important dans le cadre des missions COBE, WMAP et Planck, et il est toujours au cœur des principaux programmes d'observation.

En 1998, des observations de supernovae de type Ia, une classe particulière d'étoiles naines blanches qui explosent, ont montré que l'expansion de l'Univers s'accélérait plutôt que de décélérer sous l'influence seule de la gravité. Cela a conduit à la découverte de l'énergie noire, dont l'existence et le comportement constituent actuellement la plus grande question en cosmologie.

Pour répondre à cette question, un autre ensemble de données est crucial : les observations des structures à grande échelle de l'Univers. Les galaxies et les vides cosmiques forment une

vaste toile cosmique, révélant la distribution de la matière dans l'Univers. L'observation des décalages vers le rouge des galaxies, d'abord réalisée par le CfA Redshift Survey à la fin des années 1970, puis étendue par le Sloan Digital Sky Survey dans les années 2000, et actuellement réalisée par le Dark Energy Spectroscopic Instrument, nous permet de cartographier l'Univers avec une grande précision, et de mesurer le taux d'expansion de l'Univers à différentes époques afin de caractériser l'énergie noire.

Toutes ces observations cherchent également à mettre en lumière les premiers instants de notre Univers. Depuis plusieurs décennies, l'inflation est considérée comme le paradigme principal pour décrire les premiers instants de l'Univers. L'inflation est un ensemble de nombreux modèles qui peuvent être testés et améliorés à partir de la mesure de différentes propriétés telles que l'inclinaison du spectre de puissance scalaire primordial, les ondes gravitationnelles primordiales ou de non-gaussianité primordiale (PNG). La première est déjà bien contrainte par les dernières données CMB de Planck, et la seconde suscite un intérêt croissant avec les futures missions d'observation de la polarisation en mode B du fond diffus cosmologique. En ce qui concerne la variance, le PNG reste à ce jour peu contraint par les expériences actuelles, et c'est le sujet principal de cette dissertation.

Les modèles d'inflation les plus simples prédisent une distribution quasi gaussienne des fluctuations primordiales *i.e.* une quantité minimale de PNG. En particulier, la détection de non-gaussianité locale : $\mathcal{O}\left(f_{NL}^{loc}\right) \sim 1$ exclura l'inflation à champ unique et laissera entrevoir des modèles plus complexes *e.g.* l'inflation à champs multiples.

Actuellement, les meilleures contraintes sur les PNGs sont obtenues à partir des données de Planck : $f_{NL}^{loc} = -0.9 \pm 5.1$. Pour contourner la limite de variance cosmique des observations du CMB, une possibilité serait d'utiliser l'énorme puissance statistique des regroupements de galaxies en 3D, en sondant un grand volume de l'Univers. Une approche prometteuse consiste à utiliser l'empreinte minuscule laissée à grande échelle sur le spectre de puissance de la matière par la PNG locale, connue sous le nom de biais dépendant de l'échelle. Grâce à cette méthode, l'instrument spectroscopique de l'énergie noire (DESI) devrait contraindre la PNG locale avec une précision similaire à celle de Planck. Bien que la mesure dépende fortement de la valeur du biais linéaire du traceur et de sa réponse au PNG, l'échantillon de quasars semble être la cible de choix, puisqu'il sonde un très grand volume.

Ce résumé succinct donne un aperçu extrêmement bref de l'ensemble du manuscrit, tout en essayant d'apporter les éléments clefs.

1 Cosmologie avec les structures à grande échelle

L'avènement de la relativité restreinte, développée par Albert Einstein en 1905, a marqué une transformation significative de la physique classique en introduisant le concept d'espace-temps et en reliant l'énergie et la masse. Si la relativité restreinte a posé les bases, il a fallu attendre l'invention de la relativité générale par Einstein en 1915 pour qu'une théorie respectant la relativité restreinte soit disponible pour expliquer la force gravitationnelle. Bien qu'elle ait déjà expliqué la précession du périhélie de Mercure, la relativité générale a été validée pour la première fois en 1919 avec la mesure de la déviation de la lumière par le soleil qu'elle avait prédite (Eddington 1919, Dyson et al. 1920). Cela a ouvert la voie aux débuts de la cosmologie moderne.

Pendant une grande partie du 20e siècle, l'idée que l'Univers était en expansion et donc plus dense et plus chaud dans le passé était controversée. En fait, Fred Hoyle, qui croyait en un Univers immobile, l'a même qualifiée de façon moqueuse de "Big Bang". Cependant,

la découverte du rayonnement de fond cosmologique (CMB) par Penzias et Wilson en 1965 a confirmé une prédiction faite par Gamow, Alpher et Herman en 1948.

Au fur et à mesure de l'expansion et du refroidissement de l'Univers, les protons et les électrons se sont combinés pour former de l'hydrogène, empêchant les photons présents à l'époque d'interagir par diffusion Thomson. Les photons ont alors été libérés dans l'Univers, créant le rayonnement de fond cosmologique. Ce rayonnement ne peut exister que dans un Univers en expansion dans lequel la température diminue avec le temps.

La chronologie, voir Fig. B1.1 résume l'histoire de l'Univers selon le modèle du "Big Bang chaud". Elle peut être rapidement décrite comme suit :

- La phase initiale entre l'"instant zéro" et le début réel du modèle du "Big Bang chaud" correspond à l'Univers primitif et est actuellement expliquée par l'inflation cosmique. Notons qu'avant l'époque de Planck ($t_{\rm pl} \sim 10^{-43}$ s), la théorie actuelle de la gravitation ne peut être appliquée puisque les 4 forces fondamentales doivent être considérées dans un cadre unique.
- Au tout début ($t \sim 10^{-30}$ s), l'Univers était constitué d'un plasma extrêmement chaud et dense, connu sous le nom de "soupe primordiale", composé des particules les plus fondamentales, comme les quarks et les gluons. Au fur et à mesure de l'expansion et du refroidissement de l'Univers, la température a chuté jusqu'au point où les premiers noyaux ont pu se former, limitant ainsi l'existence des particules les plus énergétiques.
- Le plasma était alors presque exclusivement constitué des particules les moins énergétiques, comme les protons, les électrons, les neutrinos et les photons. Pendant une courte période ($t \sim 10/10^3$ s), le plasma était suffisamment chaud pour permettre la création de noyaux plus lourds. C'est la nucléosynthèse primordiale (BBN) qui est responsable de la formation de la plupart de l'hélium et du deutérium dans l'Univers. La quantité observée de ces deux composants dans les galaxies (Aver et al. 2015, Cooke et al. 2018) ne peut être expliquée que par la BBN et est également en faveur du modèle du "Big Bang chaud".
- Au fur et à mesure de l'expansion et du refroidissement de l'Univers, les protons et les électrons se sont combinés. C'est la recombinaison ($t \sim 380\ 000$ yr). Les photons se détachent de la matière et se propagent librement dans l'Univers.
- Après la recombinaison, l'Univers n'était pas assez dense pour émettre de la lumière. C'est ce qu'on appelle l'âge des ténèbres. Sous l'effet de la gravité, la matière est tombée dans les régions les plus denses, créant les premières étoiles et galaxies. Ces premières structures ont émis de nouvelles premières lumières qui sont responsables de la réionisation ($t \sim 250/900$ Myr) de l'Univers (Bromm and Yoshida 2011, Wise et al. 2012).
- Ensuite, les galaxies ont continué à se former sous l'effet de la compétition entre la gravité et l'expansion de l'Univers, créant la toile cosmique telle qu'elle est observée aujourd'hui avec une structure filamentaire et des amas.

Bien que le modèle du "Big Bang chaud" ait reçu un soutien important grâce à la découverte du rayonnement cosmique de fond et aux abondances observées des éléments légers produits lors de la nucléosynthèse primordiale, il reste confronté à des défis qui ne peuvent être résolus sans théories supplémentaires. Par exemple, les problèmes suivants doivent être expliqués :

► **Problème de la platitude** L'évolution de la courbure peut s'écrire à partir de (1.28) comme suit

$$\frac{\mathrm{d}\Omega_K}{\mathrm{d}\ln a} = (3w+1)\,(1-\Omega_K)\,\Omega_K, \tag{B.1}$$

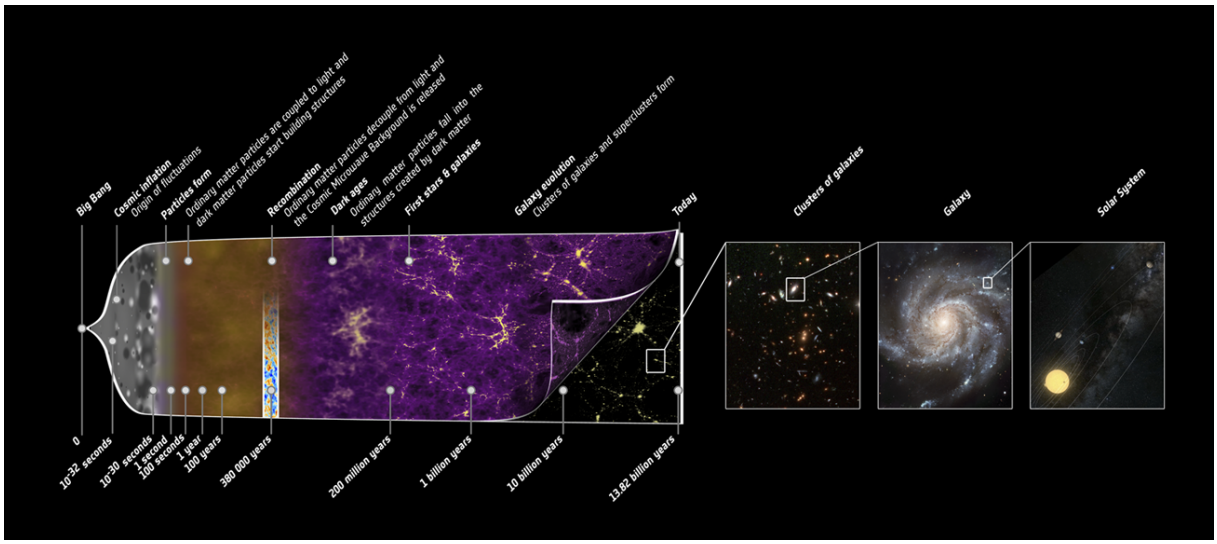


Figure B1.1: Chronologie cosmologique du modèle du Big Bang chaud qui résume l'histoire de notre Univers, longue de près de 14 milliards d'années. Depuis la phase initiale du cosmos, où les propriétés de l'Univers étaient presque uniformes et ponctuées seulement par de minuscules fluctuations, jusqu'à l'Univers bien formé et groupé que l'on observe aujourd'hui. Credits: ESA – C. Carreau

qui peut être intégré en fonction du décalage vers le rouge

$$\Omega_K = \frac{\Omega_{K0}}{(1-\Omega_{K0})\,(1+z)^{3w+1}+\Omega_{K0}} \tag{B.2}$$

Les observations actuelles sont en faveur d'un Univers plat, typiquement $|\Omega_{K0}| = |\Omega_0 - 1| < 0.1$. En négligeant la contribution de l'énergie noire, cela conduit à une égalité rayonnement-matière ($w = 0$) à

$$|\Omega\,(z_{\rm eq}\,) - 1| < 3 \times 10^{-5},$$

puis ($w = 1/3$) à un très faible aplatissement au temps de Planck

$$|\Omega\,(z_{pl}) - 1| < 10^{-60}.$$

Par conséquent, la théorie devrait pouvoir soit expliquer pourquoi $\Omega_K = 0$ avec une grande précision, soit identifier un mécanisme capable de produire une si petite courbure.

► **Problème de l'horizon** La température observée du CMB est prodigieusement isotrope, la fluctuation relative étant d'environ 10^{-5}. Cela nous amène à l'hypothèse fondamentale du principe cosmologique.

Pour expliquer simplement cette isotropie, les régions qui émettent des photons CMB à z_{rec} doivent être causalement liées, c'est-à-dire que la lumière a pu voyager entre elles avant z_{rec}. Calculons l'horizon des particules à z_{rec}, qui donne la distance maximale que les particules de lumière auraient pu parcourir jusqu'à un observateur depuis le début de l'Univers. Il s'agit exactement de la distance de rapprochement à z_{rec} :

$$\chi(z_{rec}) = \int_{z_{rec}}^{\infty} \frac{\mathrm{d}z}{H(z)} \approx \frac{2}{\sqrt{\Omega_m} H_0} \frac{1}{\sqrt{1+z_{rec}}}. \tag{B.3}$$

Comme la recombinaison se produit dans la zone dominée par la matière, $H^2(z)$ est approximé à $\Omega_m H_0^2 (1+z)^3$ à partir de (1.34). La distance angulaire correspondante est $d_A(z_{rec}) = a(z_{rec})\chi(z_{rec}) \approx 2^\circ$.

Enfin, seules des taches d'un diamètre d'environ 2°, *i.e.* avec une surface d'environ 0.0012 sr, sur la sphère céleste pourraient être causalement connectées au moment de la recombinaison. Par conséquent, le CMB devrait être constitué de $4\pi/0,0012 \approx 10^4$ taches causalement indépendantes qui devraient toutes avoir la même température. Une solution élégante consisterait à fournir un mécanisme qui augmente l'horizon des particules avant la recombinaison.

▶ **Problème des monopoles** Malheureusement, nombre de théories prédisent la création de "défauts topologiques" résultant de la transition de phase due à la brisure de symétrie lorsque l'Univers se refroidit au fur et à mesure de son expansion. En particulier, la plupart des théories GUT prédisent la création de monopôles magnétiques (voir Zeldovich and Khlopov 1978, Preskill 1979, Guth and Tye 1980) dont l'hypothèse avait déjà été émise par Maxwell ou Dirac. Dans le cas de la transition GUT, les monopôles magnétiques devraient être la matière dominante de l'Univers. Cependant, aucun monopôle n'a jamais été observé, directement ou indirectement, par l'homme. Les monopôles sont appelés *reliques indésirables* car ils sont produits par la théorie mais ne sont pas observés.

▶ **Inflation, solution miraculeuse** Ces deux premiers problèmes peuvent être résolus, avec une grande pertinence expérimentale, en ajoutant une période supplémentaire d'expansion accélérée au début de l'histoire de l'Univers : *inflation*. Le problème du monopôle a été l'une des premières motivations de la physique inflationniste, car l'inflation nous permet de contrôler les échelles d'énergie auxquelles les particules sont créées dans l'Univers, et le choix d'une échelle d'énergie inférieure à la transition GUT résout ce problème. Cette description de l'Univers primitif a été initiée par Guth (1981), Albrecht and Steinhardt (1982), Linde (1982a).

Les mécanismes qui peuvent créer l'inflation seront le sujet principal du chapitre 2. Notons que l'inflation fournira, naturellement, une explication de la taille et de la forme des fluctuations primordiales. Aujourd'hui, l'inflation et le modèle du Big Bang chaud constituent le modèle cosmologique standard bien que la notion de modèle standard puisse être très variable dans le temps.

Cette thèse vise à apporter quelques observations, malheureusement indirectes, à ce mécanisme via l'observation de la distribution des galaxies.

2 L'Univers primordial

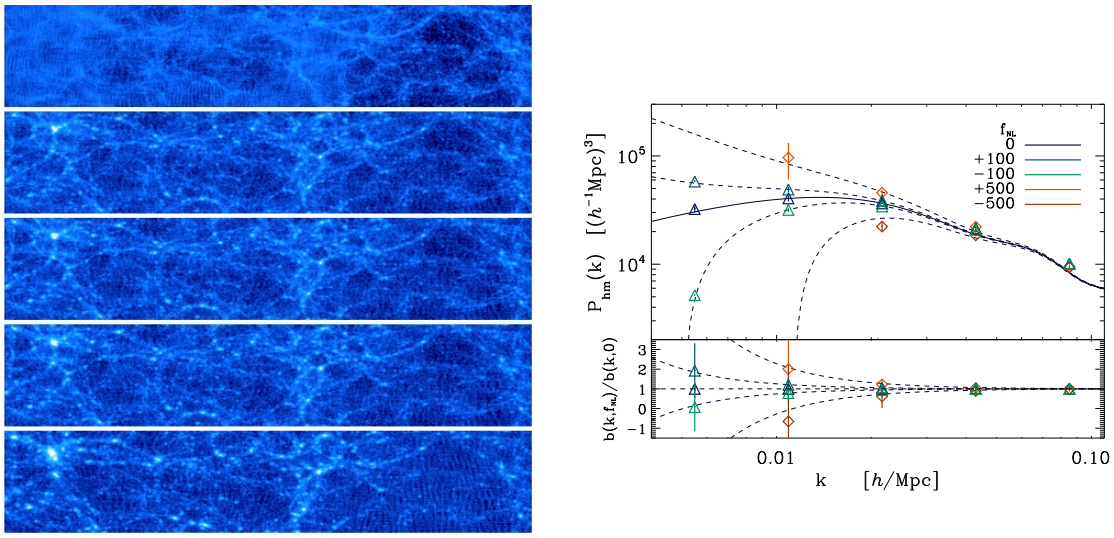
Bien que certains problèmes dans le modèle du Big Bang chaud puissent émerger comme les problèmes de planéité, d'horizon et de monopôle, il semble que, dans l'ensemble, il décrive correctement l'Univers. En particulier, la nucléosynthèse primordiale, la découverte du CMB et la description théorique de ses anisotropies sont des prédictions fortes du modèle ΛCDM.

Il est donc naturel d'essayer de compléter ce modèle plutôt que de le changer complètement. L'observation la plus lointaine que l'on puisse faire en collectant de la lumière provient du CMB, et il reste en amont toute une partie de l'histoire de l'Univers qui est encore inconnue. Une phase d'expansion accélérée au tout début de l'Univers, pourrait être le remède miracle à tous nos maux. Afin de classer les différents candidats à ce paradigme, nous étudierons les non-gaussianités primordiales et l'empreinte qu'elles ont laissée dans le clustering des galaxies.

Une approche prometteuse consiste à utiliser la petite empreinte inattendue laissée aux grandes échelles sur le spectre de puissance de la matière par les PNGs locale, connue sous le nom de biais dépendant de l'échelle. Cette relation, d'abord découverte par Dalal et al. (2008) puis confirmée par (Desjacques et al. 2009, Grossi et al. 2009, Pillepich et al. 2010),

a radicalement changé la recherche de la non-gaussianité primordiale au cours de la dernière décennie, puisqu'elle permet de mesurer les PNGs uniquement avec le spectre de puissance, et à des échelles où la théorie linéaire se vérifie.

Dans la Section 1.3.1.1, on a supposé que le biais des galaxies ne dépendait que du champ $\delta(\mathbf{x})$ à l'ordre linéaire. Cependant, la présence de PNG locaux peut avoir un impact sur la formation des galaxies. En effet, une valeur positive de $f_{\rm NL}^{\rm loc}$ dans (2.86) augmentera le champ gravitationnel primordial là où le champ de densité est le plus dense de sorte que la matière s'agrégera dans les puits de potentiel primordiaux formant des régions plus denses plus rapidement. Et le contraire pour une valeur négative $f_{\rm NL}^{\rm loc}$ qui ralentira l'agrégation de la matière dans les puits de potentiel primordiaux. La Fig. B2.2a montre des simulations avec des conditions initiales similaires mais avec différentes valeurs de $f_{\rm NL}^{\rm loc}$: -5000, -500, 0, 500, 5000 de haut en bas. L'impact de $f_{\rm NL}^{\rm loc}$ est clairement visible en faisant correspondre les structures à grande échelle.



(a) **(b)**

Figure B2.2: Ces figures sont tirées de Dalal et al. (2008), qui a été le premier à remarquer l'impact des PNGs locale sur le spectre de puissance des halos. (a) Simulation à $z = 0$ avec des conditions initiales similaires mais avec $f_{\rm NL}^{\rm loc} = -5000, -500, 0, +500, +5000$ de haut en bas. Chaque tranche a une largeur de 375 h^{-1} Mpc et une hauteur et une profondeur de 80 h^{-1} Mpc. L'impact d'une valeur positive ou négative de $f_{\rm NL}^{\rm loc}$ peut être visualisé en faisant correspondre les structures à grande échelle. Dans le cas d'une valeur positive, les régions primordiales sur-denses sont plus regroupées. (b) Le panneau supérieur montre les spectres de puissance croisée entre les halos et la matière noire. Les points sont des mesures issues de simulations avec plusieurs valeurs de $f_{\rm NL}^{\rm loc}$ à $z = 1$ pour des halos de masse $1.6 \times 10^{13} M_\odot < M < 3.2 \times 10^{13} M_\odot$. Les lignes noires et pointillées correspondent à la prédiction théorique donnée par (2.104). Le panneau inférieur montre le rapport $b(k, f_{\rm NL}^{\rm loc})/b(k, f_{\rm NL}^{\rm loc} = 0)$. Les mesures avec $f_{\rm NL}^{\rm loc} = -500, 500$ proviennent de simulations plus petites et ont donc des erreurs statistiques plus importantes. Le biais dépendant de l'échelle modifie fortement la valeur du spectre de puissance aux grandes échelles *i.e.* aux petits k. L'effet attendu est bien sûr beaucoup plus faible que celui montré ici.

Il faut donc compléter le biais des galaxies (1.111) (voir Desjacques et al. 2018) avec un terme supplémentaire prenant en compte les modifications locales impliquées par le champ gravitationnel primordial,

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + b_\Phi f_{\rm NL}^{\rm loc} \Phi_{prim}(\mathbf{x}) + \epsilon, \tag{B.4}$$

Le champ gravitationnel primordial Φ_{prim} peut être relié à la perturbation de la densité de matière δ par une fonction de transfert (1.77) telle que $\delta(k) = \alpha(k)\Phi_{prim}(k)$. Notons que pour correspondre à la définition (2.86), Φ_{prim} est normalisé dans la zone dominée par la matière de

telle sorte que $\Phi = 3/5\mathcal{R}$. La transformée de Fourier du champ δ_g est

$$\delta_g(\mathbf{k}) = \left(b_1 + \frac{b_\Phi}{\alpha(k)} f_{\rm NL}^{\rm loc}\right) \delta(\mathbf{k}), \tag{B.5}$$

où la contribution stochastique a été supprimée pour des raisons de commodité.

Enfin, en présence de PNG locale, le spectre de puissance des galaxies en ordre linéaire s'écrit

$$P(k,z) = \left(b_1(z) + \frac{b_\Phi(z)}{\alpha(k,z)} f_{\rm NL}^{\rm loc}\right)^2 \times P_{lin}(k,z), \tag{B.6}$$

où $\alpha(k,z)$ peut être calculer directement avec la fonction de transfert à partir de `CLASS`

$$\alpha(k,z) = T_{\Phi\to\delta}(k,z) = \sqrt{\frac{P_\delta(k,z)}{P_{\Phi_{prim}}(k)}}, \tag{B.7}$$

et puisque Φ_{prim} est normalisé dans la zone dominée par la matière de telle sorte que $\Phi = 3/5\mathcal{R}$, et avec (1.82), le spectre de puissance primordial peut être écrit comme suit

$$P_{\Phi_{prim}}(k) = \frac{9}{25}\frac{2\pi^2}{k^3} A_s \left(\frac{k}{k_{\rm pivot}}\right)^{n_s-1}. \tag{B.8}$$

Le biais dépendant de l'échelle laisse une empreinte unique sur les grandes échelles ($k < 0,001$ Mpc^{-1} h) dans le spectre de puissance, comme le montre la figure B2.2b. Bien que cette relation soit théoriquement très prometteuse, puisqu'elle implique des échelles où la théorie est linéaire, elle est malheureusement sujette à des effets systématiques très importants, connus sous le nom de *effets systématiques d'imagerie*, ce qui les rend délicats à exploiter, comme nous le verrons dans cette thèse.

3 Relevé de galaxies

La cartographie de l'Univers nécessite l'acquisition des coordonnées tridimensionnelles des galaxies qui nous entourent. Nous utilisons d'abord une étude photométrique pour détecter et localiser les galaxies dans le ciel. Ensuite, nous utilisons la spectroscopie pour déterminer leurs décalages vers le rouge (1.24), qui servent d'approximation pour leurs distances (1.26).

Un relevé photométrique capture des images du ciel avec ou sans filtres optiques. Des temps d'exposition plus longs permettent de détecter des objets moins lumineux. La position de chaque objet dans le ciel est généralement donnée en coordonnées équatoriales, représentées par (R.A., Dec.). La *Déclinaison* (Dec.) est analogue à la latitude terrestre, tandis que l'*Ascension droite* (R.A.) est comparable à la longitude terrestre.

Alors que les mesures photométriques sont relativement rapides puisqu'elles ne dépendent que du temps d'exposition, les mesures spectroscopiques prennent beaucoup plus de temps. En effet, la lumière de chaque objet répartie sur un CCD doit être collectée dans une fibre puis analysée individuellement à l'aide d'un spectrographe. Dans la suite nous utiliserons les données collectées par le nouvel instrument DESI (Dark Energy Spectroscopic Instrument) qui est montré schématiquement dans la Fig. B3.3.

4 Relevé de quasars de DESI

Au cours des deux dernières décennies, les quasars (ou objets quasi-stellaires, ou QSO) sont devenus un élément clé de notre compréhension de la cosmologie et de l'évolution des galaxies.

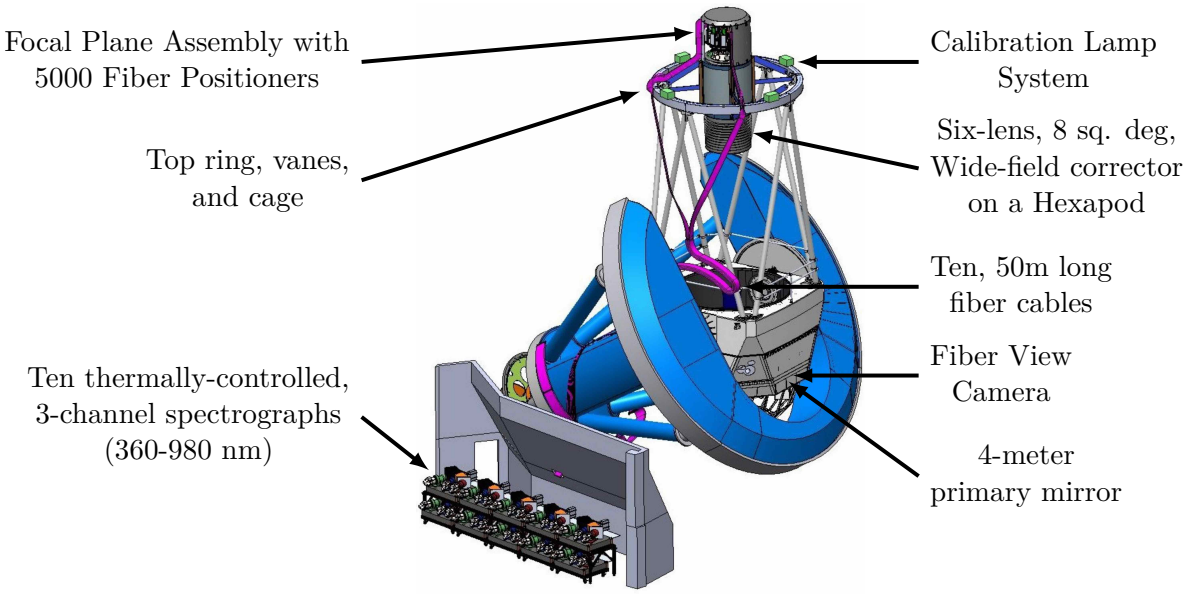


Figure B3.3: Modèle CAO 3D de DESI installé sur le télescope Mayall. Les parties les plus importantes de l'instrument sont annotées.

Faisant partie des sources extragalactiques les plus lumineuses, ils sont devenus un pilier des études cosmologiques telles que le 2dF Quasar Redshift Survey (2QZ ; Croom et al. 2001) et le Sloan Digital Sky Survey (SDSS ; York et al. 2000), où ils sont les cibles privilégiées pour étudier les structures à grande échelle à un décalage vers le rouge élevé.

Dans le cadre de la troisième génération du Sloan Digital Sky Survey (SDSS-III ; Eisenstein et al. 2011), le Baryon Acoustic Oscillation Survey (BOSS ; Dawson et al. 2013) a mesuré le spectre d'environ 300 000 quasars, dont 180 000 à $z > 2,15$, jusqu'à une magnitude limite de $g \sim 22$. Dans le cadre de SDSS-IV, l'étude spectroscopique étendue de l'oscillation des baryons (eBOSS ; Dawson et al. 2016) a observé 350 000 quasars avec des décalages vers le rouge de $0,8 < z < 2,2$ à $g \sim 22,5$, en plus de cibler 60 000 nouveaux quasars à $z > 2,2$ (Lyke et al. 2020). DESI vise à quadrupler le nombre de quasars connus et à obtenir les spectres de près de trois millions de quasars, atteignant des magnitudes limites $r \sim 23$.

DESI utilisera cet échantillon pour mesurer, voir Section 1.3.2.4, l'échelle des oscillations acoustiques du baryon (BAO) et la croissance de la structure à travers les distorsions de l'espace redshift (RSD) qui a été inaugurée pour les quasars par les études menées dans eBOSS (Zarrouk et al. 2018, Hou et al. 2021, Neveux et al. 2020).

En raison de leur morphologie ponctuelle et de leurs caractéristiques photométriques qui imitent les faibles étoiles bleues dans les longueurs d'onde optiques, en particulier pour les QSO de Ly-α la sélection des QSO est un défi. La sélection réussie d'un échantillon de QSO très complet et pur est généralement basée sur leur excès UV (Richards et al. 2002, Ross et al. 2012).

Les QSOs présentent généralement des spectres durs dans le régime des longueurs d'onde des rayons X, une émission brillante de Ly-α dans l'UV au repos, et un spectre en loi de puissance se comportant comme $F_\nu \propto \nu^\alpha$ avec $\alpha < 0$ dans les bandes de l'infrarouge moyen (Stern et al. 2005, Donley et al. 2012). Dans les couleurs optiques moyennes, les QSOs à la plupart des redshifts ne sont pas facilement distingués des étoiles beaucoup plus nombreuses. Une sélection réussie d'un échantillon de QSOs pur et très complet doit utiliser la photométrie UV ou infrarouge. Avec la mission étendue (WISE) qui a plus que quadruplé le temps d'exposition de l'étude originale (WISE), et en l'absence d'imagerie en bande 'u' sur l'ensemble de l'empreinte DESI, nous avons décidé de nous appuyer sur la photométrie optique et infrarouge pour la sélection des QSO.

Par conséquent, la sélection des cibles QSO de DESI est une combinaison de couleurs optiques uniquement et optiques+IR. Afin d'illustrer cette stratégie, nous utilisons deux couleurs, $grz-W$ vs. $g-z$ où grz est une moyenne pondérée des flux de la bande grz avec flux(grz) = [flux(g) + 0. 8×flux(r) + 0.5×flux(z)] / 2.3 et W une moyenne pondérée des flux $W1$ et $W2$ avec flux(W)=0.75×flux($W1$)+0.25×flux($W2$). Dans le Legacy Imaging Surveys (Dey et al. 2019), la conversion des flux linéaires en magnitudes est $m = 22.5 - 2.5\log_{10}(\text{flux})$. La Fig. B4.4 montre l'essentiel des cibles QSO qui sont identifiées dans une sélection optique+IR où l'excès d'émission infrarouge des QSOs résulte en une ségrégation claire des étoiles avec des flux optiques similaires. Les SEDs stellaires échantillonnent en effet la queue rapidement décroissante du spectre du corps noir à ces longueurs d'onde, où les QSOs ont un SED beaucoup plus plat que les étoiles. Cette méthode a été démontrée précédemment dans eBOSS et la Fig. 5 de Myers et al. (2015) montre la même séparation entre les étoiles et les QSOs grâce à l'imagerie WISE.

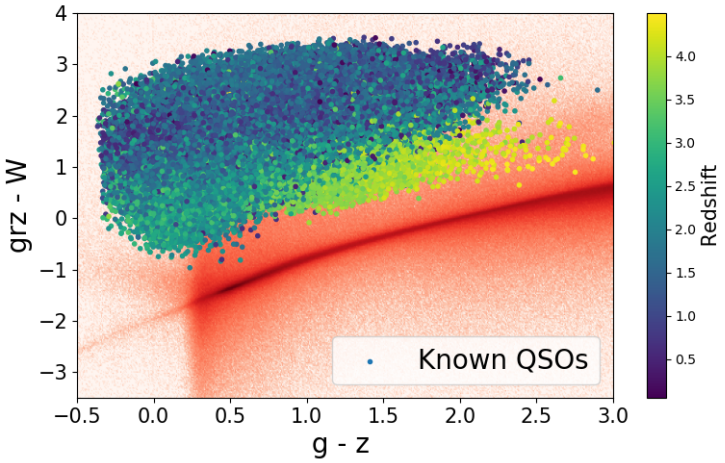


Figure B4.4: Couleurs dans l'optique ou le proche infrarouge des objets classés photométriquement comme étoiles (rouge) ou spectroscopiquement comme QSOs (points bleus à jaunes, en fonction de leur redshift). La couleur $grz - W$ nous permet de rejeter les étoiles en fonction de l'"excès infrarouge" des QSOs.

Les algorithmes basés sur les réseaux neuronnes mis en œuvre dans BOSS (Yèche et al. 2010) ont permis d'augmenter l'efficacité de la sélection des QSO d'environ 20% par rapport aux coupes colorées. De même, pour améliorer le taux de réussite de DESI, nous avons utilisé un algorithme d'apprentissage automatique basé sur les forêts aléatoires.

Comme décrit dans la Section 2.2.3, une approche prometteuse pour sonder l'inflation est l'empreinte minuscule laissée sur le spectre de puissance de la matière par la non-Gaussianité primordiale induite par l'inflation. Cette mesure Ross et al. (2013), Castorina et al. (2019), Mueller et al. (2022) est connue pour être limitée par des effets systématiques à grande échelle, dont la plupart sont dus aux systématiques d'imagerie imprimées sur la densité des cibles spectroscopiques lors de la sélection des cibles.

Différentes stratégies ont été développées pour traiter les effets systématiques de l'imagerie et améliorer la fiabilité des études de regroupement. Dans ce travail, nous suivons l'approche qui a été utilisée pour les études SDSS (Myers et al. 2006, Ross et al. 2011, Ho et al. 2012, Ross et al. 2017; 2020, Raichoor et al. 2021) et pour les études Dark Energy Survey (DES Collaboration et al. 2021) (Leistedt et al. 2016, Elvin-Poole et al. 2018). Cette méthode modélise la variation de la densité de la cible comme une fonction linéaire des caractéristiques d'imagerie (voir Myers et al. 2015, Prakash et al. 2016) afin d'éliminer les fluctuations causées par les systématiques d'imagerie. Un poids de correction est ensuite calculé et appliqué aux données. Comme cette méthode lisse les fluctuations de densité, il faut vérifier dans quelle mesure la

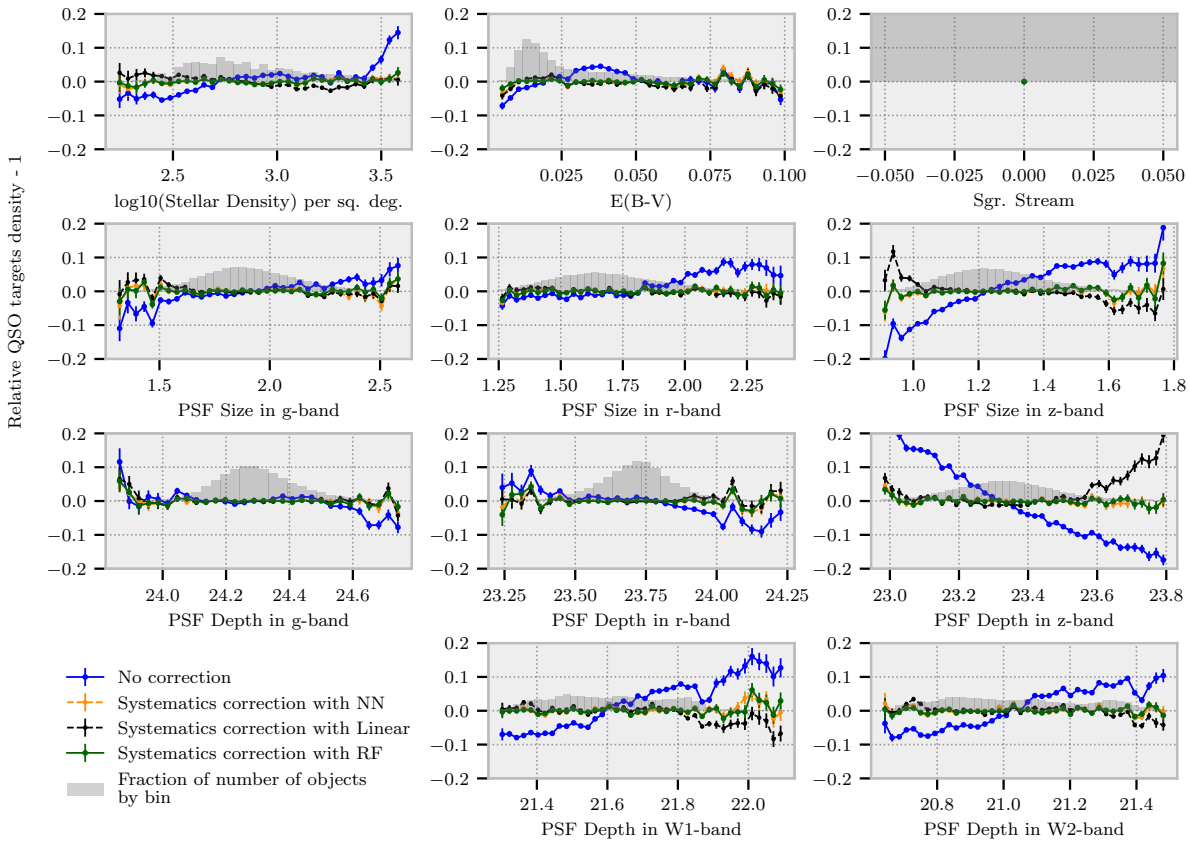


Figure B4.5: Densité relative des cibles QSO dans le nord en fonction de chaque caractéristique d'observation. La densité relative des cibles QSO est une valeur moyenne après avoir rejeté les valeurs aberrantes, c'est-à-dire les pixels dont la couverture est inférieure à 90%. Les lignes bleues représentent la sélection brute des cibles QSO de DESI. Les lignes vertes (*resp.* jaune / noir) représentent la sélection des cibles QSO après correction des effets systématiques à l'aide de la régression RF (*resp.* NN / linéaire). L'histogramme représente la fraction d'objets dans chaque case pour chaque caractéristique observationnelle et les barres d'erreur sont l'écart-type estimé de la densité de cibles normalisée dans chaque case. Les trois méthodes réussissent à aplanir la densité relative des cibles QSO en fonction de chaque caractéristique d'observation. Cependant, la méthode linéaire est moins efficace que les deux autres méthodes.

procédure d'atténuation affecte le signal cosmologique. Une autre approche moins courante est basée sur la projection de mode (Rybicki and Press 1992, Tegmark et al. 1998, Leistedt et al. 2013, Elsner et al. 2016, Kalus et al. 2019) : Les modes (dans l'espace de Fourier) ou les pixels (dans l'espace de configuration) se voient attribuer une variance accrue lorsque la carte de systématicité présente des valeurs importantes, de sorte que la matrice de covariance présente des valeurs plus importantes en présence de systématicité. Il s'agit d'une méthode robuste qui, toutefois, n'atténue la systématicité qu'en utilisant également une combinaison linéaire des cartes d'imagerie. Elle ne peut pas modéliser les effets non linéaires qui sont maintenant observés, comme illustré dans Ho et al. (2012). Les stratégies de pondération des corrections et de projection des modes peuvent être combinées dans un cadre commun, comme l'explique Weaverdyck and Huterer (2021).

Adoptant une approche similaire à Rezaie et al. (2020), j'ai développé ma propre méthode de correction utilisant des régressions avec des réseaux de neurones ou des forets aléatoires. Le résultat est montré sur la Fig. B4.5.

5 Mesure des non-gaussianités primordiales avec les quasars de DESI

Maintenant que DESI produit la plus grande carte 3D de l'Univers jamais réalisée, il est temps de passer à l'étape suivante, en analysant toutes ces nouvelles données. DESI aura plusieurs publications de données (DR) tout au long de son étude. Dans ce qui suit, nous travaillerons avec DR1, les données de la première année d'observation, voir Section 3.2.2.3.

Cette première livraison de données (DR1) est particulièrement intéressante pour l'échantillon des quasars, car ils ont la plus haute priorité pendant l'observation et sont donc collectés en premier. Nous nous attendons donc déjà à des contraintes compétitives sur la mesure de la non-gaussianité primordiale.

Le signal dans les données a été modifié suivant une valeur aléatoire de $f_{\rm NL}^{\rm loc} \in [-15, 15]$. Le meilleur ajustement des paramètres sont indiqués dans le tableau 7.1 et les postérieurs pour l'estimation du spectre de puissance sans pondération, avec les pondérations FKP et les pondérations OQE sont indiqués dans la figure B5.6.

Table 7.1: Meilleures valeurs d'ajustement pour les données DESI QSO Y1. Les erreurs sont les 1σ (68% CL) des postérités montrées dans la Fig. B5.6.

	$f_{NL}^{\rm loc}$	b	$s_{n,0}$	Σ_s	$\chi^2/(115-4)$
no weights	8^{+25}_{-26}	$2.110^{+0.043}_{-0.045}$	1050^{+460}_{-470}	$7.95^{+0.47}_{-0.47}$	1.53
FKP weights	6^{+22}_{-19}	$2.206^{+0.041}_{-0.048}$	810^{+440}_{-480}	$7.97^{+0.52}_{-0.41}$	1.38
OQE weights	-17^{+15}_{-12}	$2.833^{+0.074}_{-0.069}$	-480^{+630}_{-670}	$8.71^{+0.52}_{-0.53}$	1.54

Comme prévu, le biais linéaire est plus élevé à un redshift effectif plus élevé, mais la valeur trouvée ici est inférieure à la prédiction de Laurent et al. (2017) qui a travaillé avec le QSO eBOSS. De plus, la différence entre les poids FKP et OQE est plus élevée qu'avec les EZmocks. Ces deux effets seront discutés plus loin.

Les erreurs sur $f_{\rm NL}^{\rm loc}$ sont plus importantes que dans le cas des EZmocks car nous mesurons un biais plus faible dans les données. Pour éviter cette discussion, nous devrions plutôt mesurer $b_\phi f_{\rm NL}^{\rm loc}$ directement, et nous devrions alors trouver des erreurs similaires. Notons toutefois que le taux d'erreur est à peu près similaire entre EZmocks et les données utilisant la même matrice de covariance.

La figure B5.7 montre le modèle le mieux ajusté par rapport aux données pour le cas des poids OQE, ainsi que les résidus correspondants. Nous représentons également le modèle à la meilleure valeur lorsque nous modifions la valeur de $f_{\rm NL}^{\rm loc}$ pour obtenir la meilleure valeur d'ajustement $\pm 1\sigma$. Notons que le quadripôle n'apporte pas trop de contraintes sur $f_{\rm NL}^{\rm loc}$, mais qu'il rompt les dégénérescences entre les différents paramètres du modèle.

Remarque: Par rapport à Castorina et al. (2019), nous trouvons une postériorité gaussienne pour $f_{\rm NL}^{\rm loc}$, comme dans Mueller et al. (2022). Cela s'explique par le fait que nous ne sondons pas des valeurs suffisamment négatives de $f_{\rm NL}^{\rm loc}$ pour que la forme du biais dépendant de l'échelle soit différente.

Pour évaluer notre mesure, nous devons quantifier la quantité de systématiques dans l'analyse. Ils peuvent être classés de manière préliminaire comme suit :

- Atténuation de l'effet d'observation : Dans la section 5.2.2.2, nous avons vu que la correc-

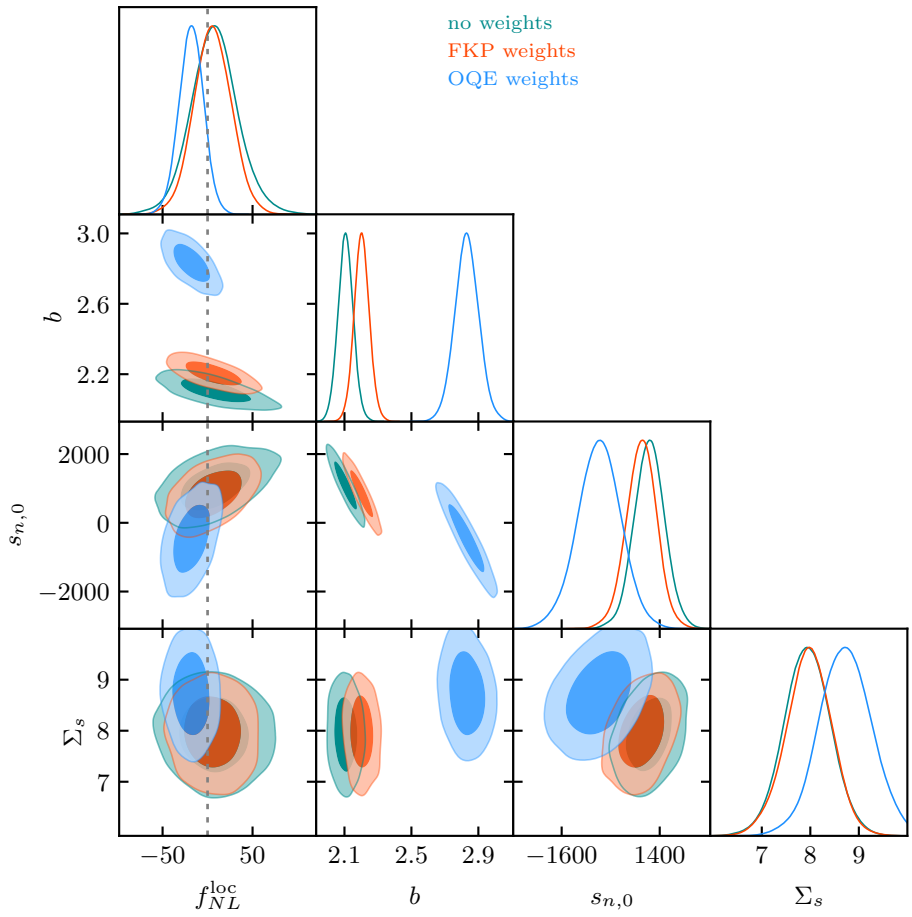


Figure B5.6: Postérieurs pour les données sans poids (vert), avec les poids FKP (orange) et les poids OQE (bleu) aux différents redshifts effectifs donnés dans le Tableau 5.1. Notons que le biais linéaire b_1 est inférieur à la prédiction au redshift effectif faite par Laurent et al. (2017).

tion `regressis` conduit à un décalage de $f_{\rm NL}^{\rm loc}$ d'environ 10. Une analyse préliminaire basée sur quelques simulations FastPM, voir la Fig. 5.15, indique un décalage additif. Une fois ce décalage entièrement calibré, nous pourrons calculer rigoureusement l'erreur systématique associée. Pour l'instant, et par prudence, nous pouvons simplement considérer la moitié du décalage comme une erreur systématique : $\sigma_{sys} = 10/2 = 5$.

- Procédure d'ajustement des éléments : Comme indiqué ci-dessus, l'utilisation de poids OQE au lieu de poids FKP entraîne un décalage de $f_{\rm NL}^{\rm loc}$ d'environ 6. Comme pour le décalage précédent et pour être prudent, nous considérons une erreur systématique d'environ $\sigma_{sys} = 6/2 = 3$.

- Erreur sur le modèle : Elle n'est pas discutée ici car nous n'avons pas assez de mocks FastPM pour la valider. En particulier, nous n'étudions pas l'impact de l'utilisation du modèle d'effondrement du halo sphérique dans (2.120). Pour une validation récente du modèle, voir, par exemple, Biagetti et al. (2017). De même, la relation de masse universelle (2.120) pourrait également être traitée comme une systématique théorique.

Ces erreurs systématiques seront additionnées en quadrature avec les erreurs statistiques finales. Nous n'utilisons pas de coefficient de corrélation entre ces deux erreurs systématiques afin d'être le plus conservateur possible. Dans le cas de l'OQE, nous avons trouvé une erreur statistique

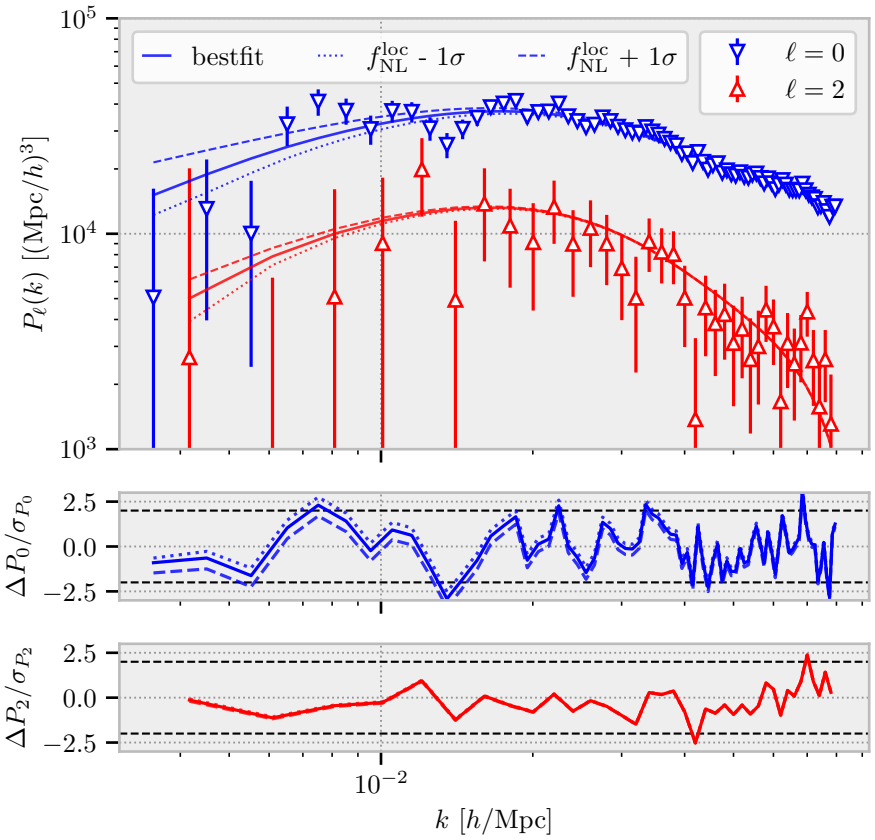


Figure B5.7: En haut : Spectre de puissance, estimé avec les pondérations OQE (p=1.6), de l'échantillon DESI QSO Y1 et du modèle associé aux meilleures valeurs d'ajustement (ligne pleine) et avec +/- 1σ (en pointillés) pour $f_{\rm NL}^{\rm loc}$, données dans le Tableau 5.7. En bas : Les résidus du monopôle et du quadripôle sont comparés au modèle aux valeurs les mieux ajustées et avec +/- 1σ pour $f_{\rm NL}^{\rm loc}$.

d'environ $\sigma_{stat} = 13,5$, et l'inclusion des erreurs systématiques augmente les erreurs finales à $\sigma_{tot} = \sqrt{\sigma_{stat}^2 + \sum_i \sigma_{sys,i}^2} \simeq 14.2$.

Pour cette analyse, les décalages systématiques observés sont suffisamment faibles par rapport à l'erreur statistique. Cependant, le budget systématique dominera les erreurs statistiques avec les données Y5 à venir, la combinaison avec les autres traceurs, ou l'utilisation de méthodes plus avancées. Il sera nécessaire de mieux comprendre ces décalages, qu'ils soient additifs ou multiplicatifs, de les modéliser pour les corriger et d'estimer une erreur systématique dans la correction.

6 Conclusion & perspective

Dans cette thèse, nous avons tenté d'observer l'empreinte que l'inflation, la théorie décrivant l'Univers primitif, a pu laisser sur les structures à grande échelle de l'Univers. Pour ce faire, nous avons utilisé le relevé spectroscopique des quasars de l'instrument DESI, dans lequel nous mesurons la fonction de corrélation en 2 points, à savoir le spectre de puissance, pour mesurer la présence d'une non-gaussianité primordiale grâce à la relation de biais dépendant de l'échelle.

Cette dissertation illustre le fait que pour mesurer les grandes échelles du spectre de puissance avec des erreurs systématiques contrôlées, une attention particulière doit être portée à chaque

étape de l'analyse : de la sélection de la cible au calcul du spectre de puissance.

Tout d'abord, j'ai été chargé de la sélection des cibles des quasars, l'un des quatre traceurs principaux de DESI (Chapitre 4). Les quasars sont devenus un pilier des études cosmologiques au cours des deux dernières décennies. Ils sont la source de choix pour étudier les structures à grande échelle à un redshift élevé, soit comme traceurs directs de la matière noire dans la gamme de redshifts $0.8 < z < 2.1$, soit comme traceurs continus avec la forêt Lyman-α imprimée dans leurs spectres à $z > 2.1$. Ce dernier cas a suscité un intérêt croissant depuis la tension potentielle à grande distance dévoilée par eBOSS dans la mesure du Lyman-α BAO. DESI ayant un nombre prédéfini de cibles pour chaque traceur, l'optimisation de la sélection des cibles est déterminante pour les quasars dont le bruit de fond est limité et dont la sélection est fortement contaminée par les étoiles.

En élargissant la classification color-cut d'eBOSS, j'ai développé une méthode basée sur la classification Random Forest, afin d'augmenter l'efficacité de la sélection de 20%. De plus, j'ai découvert que le pipeline spectroscopique de base manquait $\sim$ 10-15% de vrais quasars et pouvait être optimisé en utilisant un détecteur de raies Mg II et un classificateur de réseau neuronal pour augmenter l'efficacité de l'identification et éviter la confusion des raies. La combinaison finale présentée atteint une efficacité de redshift de $98.0 \pm 0.4\%$ et une pureté de $99.5 \pm 0.4\%$.

La sélection a été rendue difficile par une contamination stellaire inattendue provenant du courant du Sagittaire. Pour valider la sélection, j'ai étudié en profondeur l'échantillon final afin de m'assurer que les fluctuations de densité angulaire à grande échelle étaient expliquées par l'ensemble des caractéristiques photométriques connues. En particulier, j'ai développé une méthode de régression basée sur la régression Random Forest pour atténuer les systématiques d'imagerie, `regressis`, en tenant compte de la relation non linéaire entre les caractéristiques d'imagerie et en améliorant l'atténuation. Cette méthode est mise en œuvre et actuellement utilisée dans le pipeline officiel de regroupement DESI.

Cette nouvelle sélection de cibles et le pipeline spectroscopique optimisé ont été largement testés lors de la validation de l'enquête, avec un effort considérable d'inspection visuelle menée par la collaboration à laquelle j'ai participé. Ce travail permet à DESI de collecter plus de 2,8 millions de spectres de quasars avec un redshift correct, dont près d'un million avec un redshift $z > 2.1$, dépassant ainsi de 20% les exigences scientifiques de DESI. Grâce à cette priorité, DESI a déjà collecté plus de 1,4 million de quasars au cours de la première année d'observation.

Enfin, j'ai utilisé les QSOs Y1 de DESI pour contraindre PNG avec la méthode du biais dépendant de l'échelle. Pour cela, j'ai évalué l'efficacité de l'atténuation sur l'échantillon spectroscopique, j'ai développé un pipeline de bout en bout basé sur des simulations FastPM imitant un tiers de l'échantillon, y compris la contamination due aux systématiques d'imagerie. Cette analyse en cours a révélé un décalage systématique additif dû à la procédure d'atténuation et devra être étudiée avec précision sous peu pour garantir une mesure non biaisée de la non-gaussianité primordiale.

Afin d'éviter tout biais de confirmation lors de l'atténuation systématique de l'imagerie et de la validation du catalogue de regroupement, j'ai également développé un système d'aveuglement qui imite le comportement d'un faux biais dépendant de l'échelle, émulant une valeur inconnue de $f_{\rm NL}^{\rm loc}$ dans les données. Ce système d'aveuglement a été ajouté à la procédure globale d'aveuglement de DESI. Une procédure complète de validation du catalogue doit être clairement énoncée avant de lever l'aveuglement des données, et malheureusement, je n'ai travaillé qu'avec les données en aveugle.

Avec les données en aveugle, nous avons exploré le QSO Y1 et trouvé un excellent accord du spectre de puissance à grande échelle entre les différentes régions photométriques en appliquant séparément l'atténuation systématique de l'imagerie dans deux bins de redshift. Après avoir

validé le pipeline d'ajustement sur des mocks réalistes, j'ai ajusté les données en aveugle et j'ai trouvé un biais linéaire plus faible que prévu augmentant les erreurs statistiques. J'ai trouvé avec les poids FKP $f_{\rm NL}^{\rm loc} = 6^{+22}_{-19}$ et avec les poids OQE $f_{\rm NL}^{\rm loc} = -17^{+15}_{-12}$, avec $p = 1.6$. Le gain statistique entre les deux méthodes provient principalement de l'augmentation du décalage vers le rouge effectif.

Ces travaux ont donné lieu à plusieurs publications scientifiques : Chaussidon et al. (2022) pour l'atténuation systématique de l'imagerie, Chaussidon et al. (2023) pour la sélection des cibles quasars et Chaussidon et al. (2023) pour la procédure d'aveuglement. La mesure de $f_{\rm NL}^{\rm loc}$ sera publiée avec la publication des données Y1 au cours de l'année prochaine. J'ai également participé à d'autres publications, voir en particulier Alexander et al. (2023) ou Krolewski et al. (2023).

▶ **Prospects** Des prévisions réalistes pour les données DESI de l'année 5, en utilisant les pondérations FKP, donnent $\sigma(f_{\rm NL}) = 22 \to 14$ pour $p = 1,6$, et $\sigma(f_{\rm NL}) = 11 \to 7$ pour $p = 1,0$. Bien qu'un gain supplémentaire d'environ 35% soit attendu en utilisant les poids OQE, les QSOs seuls ne seront pas suffisants pour obtenir une mesure compétitive avec Planck. Cependant, cela sera possible si toutes les données DESI sont prises en compte.

En effet, cette mesure peut être améliorée avec les données DESI soit en utilisant plusieurs bins de redshift pour le QSO (redshift faible contre redshift élevé), l'autre traceur du temps sombre comme les LRG qui semble être assez compétitif (déjà $\sigma(f_{\rm NL}) \sim 15$ avec les données Y1). En outre, une autre approche prometteuse consiste à utiliser l'effet de lentille du CMB pour le mettre en corrélation croisée avec les traceurs DESI. Nous proposons ci-dessous deux pistes de recherche détaillées :

- **Améliorer la mesure du PNG avec le bispectre** L'étude nominale de DESI devrait contraindre le PNG local avec un biais dépendant de l'échelle jusqu'à $\sigma\left(f_{NL}^{loc}\right) \sim 5$ DESI Collaboration et al. (2016a). L'ajout d'informations sur le bispectre peut améliorer la contrainte sur f_{NL}^{loc} par rapport à la mesure du spectre de puissance uniquement Tellarini et al. (2016), Karagiannis et al. (2018), Moradinezhad Dizgah et al. (2021). Pour tester les PNGs à un niveau permettant d'exclure les modèles d'inflation les plus simples, l'utilisation du bispectre sera nécessaire.

La mesure de f_{NL}^{loc} via le biais dépendant de l'échelle est dégénérée avec un terme de biais inconnu b_ϕ décrivant la réponse de la densité du traceur au PNG locale. Le travail fondateur Slosar et al. (2008) fournit une description analytique de b_ϕ, mais des études plus récentes Barreira (2020; 2022a) basées sur des simulations hydrodynamiques montrent qu'elle peut ne pas être correcte pour tous les traceurs. La dégénérescence est brisée par l'utilisation du bispectre, qui permet de mesurer f_{NL}^{loc} sans étendre l'a priori sur b_ϕ et améliore la contrainte sur f_{NL}^{loc} de plus d'un facteur 5 par rapport à la mesure du spectre de puissance seul Moradinezhad Dizgah et al. (2021). J'aimerais également étendre mon travail à d'autres formes de PNG, comme les formes équilatérales et orthogonales, qui peuvent être sondées avec le bispectre Coulton et al. (2023).

Les premières analyses utilisant le bispectre dans le contexte des PNGs sont apparues plus tôt dans l'année Cabass et al. (2022), D'Amico et al. (2022). Cependant, il reste encore beaucoup à faire pour atteindre la précision souhaitée ; en particulier, les effets systématiques des observations ont été peu étudiés jusqu'à présent. Une attention particulière doit être portée à ces effets car la configuration du squeeze (forme triangulaire avec deux grands côtés et un petit : $k_1, k_2 \gg k_3$) peut être contaminée par des effets observationnels à la fois à grande échelle (systématiques d'imagerie) et à petite échelle (détermination du décalage vers le rouge).

En outre, les effets géométriques tels que la fonction de fenêtre sont plus compliqués à modéliser que dans le cas du spectre de puissance Pardede et al. (2022), en particulier pour les multipôles élevés. Les analyses de pointe avec le bispectre Ivanov et al. (2022), Philcox et al. (2022) n'ont utilisé que le monopôle, et une attention particulière serait nécessaire pour les multipôles d'ordre supérieur. Ces effets géométriques peuvent être modélisés à l'aide d'une approche basée sur la simulation afin d'éviter les calculs analytiques lourds.

- **PNG avec successeur DESI** Une fois la cartographie de l'Univers à grand décalage vers le rouge achevée avec les quasars de DESI, le successeur de DESI pourrait être une étude spécifique conçue pour cibler les objets à grand décalage vers le rouge, comme cela a été proposé pour DESI-II (Schlegel et al. 2022). L'analyse précédente pourrait alors être réalisée avec des Lymann-break galaxies (LBG) dans la gamme de redshift ($2.2 < z < 3.6$), sondant un plus grand volume et devant fournir de meilleures contraintes sur les PNGs.

Par exemple, pour contourner la variance cosmique et réduire drastiquement les erreurs, il semble séduisant de procéder à des corrélations croisées entre plusieurs traceurs ayant des biais différents (Seljak 2009, Hamaus et al. 2011). Cette méthode pourrait être appliquée à DESI-II avec deux traceurs principaux potentiels : Les émetteurs Lyman-α (LAEs) et les LBGs. Ces deux sondes ouvrent de grandes opportunités pour contraindre f_{NL}^{loc} avec le biais dépendant de l'échelle et d'autres formes de non-gaussianité avec des analyses de bispectres (Yamauchi et al. 2017).

Des temps passionnants nous attendent.