Detection of Baryon Acoustic Oscillation using Lyman-alpha Forests in DESI/Eboss

Ting Tan

LPNHE
PARIS

THESE DE DOCTORAT

DE SORBONNE UNIVERSITÉ

présentée par

Ting TAN

Pour obtenir le grade de

DOCTEUR DE SORBONNE UNIVERSITÉ

Spécialité :

Physique de l'Univers (STEP'UP - ED 560)

Detection of Baryon Acoustic Oscillation using Lyman-alpha Forests in DESI/Eboss

Soutenue le 21 septembre 2023 devant le jury composé de :

Pr	Christophe	BALLAND	Directeur de thèse
Pr	Alain	BLANCHARD	Rapporteur
Dr	Vanina	RUHLMANN-KLEIDER	Rapporteuse
Pr	Sophie	TRINCAZ-DUVOID	Présidente
Pr	Alexandre	REFREGIER	Examinateur
Pr	Matthew	PIERI	Examinateur
Dr	Stéphanie	ESCOFFIER	Examinatrice
Dr	James	RICH	Co-Directeur de thèse
Dr	Jean-Marc	LE GOFF	Co-Directeur de thèse

cnrs IN2P3 Les deux infinis

SORBONNE UNIVERSITÉ

Centre Pierre Binétruy CNRS • Berkeley

Colophon

Doctoral dissertation entitled "Detection of Baryon Acoustic Oscillation using Lyman-alpha Forests in DESI/Eboss", written by Ting Tan, completed on December 21, 2023, typeset with the document preparation system LaTeX and the yathesis class dedicated to theses prepared in France.

Keywords: cosmologie, structure à grande échelle

Mots clés : cosmology, large-scale structure, bao

This thesis has been prepared at

Laboratoire de physique nucléaire et des haut énergies

Sorbonne Université
Campus Pierre et Marie Curie
4 place Jussieu
75005 Paris
France

☎ +33 1 44 27 42 98
Web Site `https:/http://lpnhe.in2p3.fr/?lang`

Under the heavenly sun, countless flowers bloom on the sea. I see my heart, the flowing water, and the lofty mountains.

Guopan ZENG

Gazing at the bright moon, deeply nostalgic, I lower my head.

Bai LI

Detection of Baryon Acoustic Oscillation using Lyman-alpha Forests in DESI/Eboss

Abstract

Les oscillations acoustiques baryoniques (BAO) sont une sonde puissante permettant de mesurer l'expansion accélérée de l'univers et de fournir des contraintes sur les modèles d'énergie noire. Il peut être mesuré à l'aide de la fonction de corrélation à deux points des traceurs de matière, et le but de cette thèse est de mesurer le BAO à des redshifts $z > 2,1$ élevés en utilisant les forêts Lyman-α (Lyα). Cette thèse utilise des données d'observation spectroscopiques et des catalogues simulés (simulations) de deux grandes enquêtes cosmologiques, eBOSS (DR16) et DESI (EDR). Je présente l'analyse comparative Lyα de ces deux enquêtes et je les trouve cohérentes en termes de qualité et d'ajustement des données. J'ai étudié à la fois sur des simulations et sur des données, l'un des effets systématiques les plus importants de l'analyse Lyα, la présence de systèmes à haute densité de colonnes (HCD). J'ai proposé un modèle empirique et développé un modèle analytique, le modèle Voigt, pour caractériser leur impact sur les fonctions de corrélation Lyα. Le modèle Voigt est bien vérifié sur des simulations et fournit une mesure physique des paramètres de biais et RSD des HCD, ainsi qu'une bonne contrainte sur les paramètres Lyα.

Keywords: cosmologie, structure à grande échelle

Résumé

The Baryon Acoustic Oscillations (BAO) is a powerful probe to measure the accelerated expansion of the universe and provide constraints on dark energy models. It can be measured using the two-point correlation function of matter tracers, and the goal of this thesis is to measure the BAO at high redshifts $z > 2.1$ using Lyman-α (Lyα) forests. This thesis makes use of spectroscopic observation data and simulated catalogs (mocks) from two large cosmological surveys, eBOSS (DR16) and DESI (EDR). I present the comparison Lyα analysis of these two surveys and found them consistent in terms of data quality and fits. I studied on both mocks and data, one of the most important systematic effects of Lyα analysis, the presence of High Column Density Systems (HCDs). I proposed an empirical model and further developed an analytical model, the Voigt model, to characterize their impact on Lyα correlation functions. The Voigt model is well verified on mocks and provides a physical measurement of the bias and RSD parameters of HCDs, and a good constraint on the Lyα parameters.

Mots clés : cosmology, large-scale structure, bao

Laboratoire de physique nucléaire et des hautes énergies
Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France

Acknowledgement

I would like to thank sincerely my three supervisors Christophe BALLAND, James RICH, and Jean-Marc LE GOFF. It was not an easy time during my PhD study, where we suffered from quarantines due to Covid-19 for the first one and a half years. However, your encouragement and support provided me with a continuous passion to carry out my studies. I did not communicate a lot with Jean-Marc, since he left cosmology for environmental science during my second year. However, this results in a non-negligible philosophic impact on me, considering seriously the meaning of science, physics, and cosmology. I really thank Jim for his kind guidance and creative ideas for cosmology, and I really appreciate his spirit in doing serious science. You are not only a supervisor, but also a friend to me. I hope that I can keep collaborating with you in my future research career. I thank Christophe for his enormous support in my frequent visits, conferences, and exchanges. It is very rare to get such generous support during graduate study, and it is very nice of you to help plan my future life and research career.

I thank all the jury members of my defense committee, especially the two referees of my manuscript, Vanina RUHLMANN-KLEIDER, and Alain BLANCHARD. I really appreciate your careful revision and helpful comments. I thank Matthew PIERI, Stéphanie ESCOFFIER, and Alexandre REFREGIER for your participation and comments on my defense. Lastly, I thank sincerely Sophie TRINCAZ-DUVOID, who was not only the president of my defense committee but also my godmother during my three years of study. Thank you for your care and support!

I thank the other colleagues in my office, Julianna STERMER, Enya VAN DEN ABEELE, Svyatoslav TRUSOV, Ugo PENSEC, Thierry SOUVERIN, and others. Thank you for your company and support, making my research life interesting and meaningful. It is very nice being together with you all!

I thank the members of the cosmology group at LPNHE, particularly Pauline ZARROUK, Pierre ASTIER, and Nicolas REGNAULT. Thank you for your help and guidance during my study, and best wishes to Pauline for her baby! It is a very lovely group and I hope everyone will be doing well!

I thank all of my colleagues at LPNHE, particularly the director Marco ZITO, for giving me the chance to study at the lab and gain such a wonderful experience. I thank Marjorie STIEVENART-AMMOUR for her enormous help in making my complex missions and Julien BOLMONT for his support in the preparation of my defense materials. Moreover, I thank Delphine HARDIN for always being a member of my 'comité de suivi' and giving me kind support.

I thank Radek STOMPOR at Centre Pierre Binétruy and Julien GUY at Lawrence Berkeley National Laboratory, for your accepting me as a visiting student twice at Berkeley, under the CNRS-CPB international grant. I also thank my colleagues in the Lyman-alpha working group of DESI, Andreu Font-Ribera, Vid Irsic, Alma González, Corentin Ravoux, Eric Armengaud, Hiram Herrera-Alcantar, Ignasi Pérez-Ràfols, Satya Gontcho A Gontcho, and Solène Chabanier, Calum Gordon, for their support and great collaboration throughout my three years study.

I thank my lovely friends, Kang LIU, Xudong YU, Liding XU, and others, for your nice accompany and encouragement over these years. It is not always easy to against all odds on the path of the academy, but it is much easier to become a legend at Summoner's Rift.

I would like to thank my family, my friends, and my homeland. It has been five years since we last met, back when I was still pursuing my Bachelor's degree. And now, I am coming back, having finished my PhD study. How many five years do we have in our lives? I miss everything so hard.

Finally, I would like to thank myself, for being courageous and being kind. I would like to finish with my favorite Chinese poem:

Under the heavenly sun, countless flowers bloom on the sea.
I see my heart, the flowing water, and the lofty mountains.

Introduction (English version)

Modern cosmology was developed after the discovery of Hubble (Hubble's law, 1929), arguing that extragalactic nebulas are moving away from each other and as an observation that was interpreted as evidence of the expansion of the universe. 70 years later, the measurement of the acceleration of the expanding universe with type Ia supernovae further suggests a dynamic universe currently well-modeled by the so-called standard cosmological model, the ΛCDM model.

The Baryon Acoustic Oscillations (BAO) is a powerful probe to measure the Hubble parameter and provide constraints on the ΛCDM model. It imprints the density fluctuations of baryons and photons, which propagated in the form of sound waves the in the early universe baryon-photon plasma. The BAO can be measured as a peak in the two-point correlation function of matter tracers, and the goal of this thesis is to study this probe using Lyman-α (Lyα) forests.

The Lyα forests are present as absorption lines in the quasar's spectrum, caused by the Lyα transitions of photons passing through the intergalactic medium (IGM) in the universe. Using Lyα forests provides the measurement of the BAO peak at high redshift $z > 2$.

In this Ph.D. manuscript, I give a brief introduction to the cosmology model in Chapter 1. Chapter 2 presents the two cosmology surveys that my analyses benefit from, the extended Baryon Acoustic Oscillation (eBOSS), and the Dark Energy Spectroscopic Instrument (DESI). I provide a description of the synthetic Lyα data, the so-called Lyα mocks, in Chapter 3. These mocks are used to validate our Lyα analysis pipeline, which is described in Chapter 4.

I contributed as an active member in the DESI collaboration to the comparison analyses of Lyα forest BAO using mocks and data from both eBOSS and DESI. This part of my work is presented in Chapter 5, and is included in several collaboration publications of DESI.

I present in Chapter 6 the most important contribution of this thesis to the DESI collaboration, the analysis of High Column Density systems (hereafter HCDs), one of the most important systematic effects for Lyα forest BAO. I developed a theoretical model, which I call the Voigt model, to explain the non-local damping effect of HCDs on the Lyα correlation functions. This model also provides a physical ground to explain the phenomenological models that were used in previous studies.

Introduction (French version)

La cosmologie moderne s'est développée après la découverte de Hubble (loi de Hubble, 1929), arguant que les nébuleuses extragalactiques s'éloignent les unes des autres et comme une observation interprétée comme une preuve de l'expansion de l'univers. 70 ans plus tard, la mesure de l'accélération de l'univers en expansion avec des supernovae de type Ia suggère encore un univers dynamique actuellement bien modélisé par le modèle cosmologique dit standard, le modèle λCDM. Le Baryon Acoustic Oscillations (BAO) est une sonde puissante pour mesurer le paramètre de Hubble et fournir des contraintes sur le modèle λCDM. Il imprime les fluctuations de densité des baryons et des photons, qui se sont propagés sous forme d'ondes sonores dans le plasma baryon-photon de l'univers primordial. Le BAO peut être mesuré comme un pic dans la fonction de corrélation à deux points des traceurs de matière, et le but de cette thèse est d'étudier cette sonde à l'aide de forêts de Lyman-α (Lyα). Les forêts Lyα sont présentes sous forme de raies d'absorption dans le spectre du quasar, causées par les transitions Lyα des photons traversant le milieu intergalactique (IGM) dans l'univers. L'utilisation de forêts Lyα fournit la mesure du pic BAO à un décalage vers le rouge élevé $z > 2$. Dans ce doctorat. manuscrit, je donne une brève introduction au modèle de cosmologie. Le chapitre 2 présente les deux études de cosmologie dont bénéficient mes analyses, l'oscillation acoustique baryonique étendue (eBOSS) et l'instrument spectroscopique à énergie noire (DESI). Je fournis une description des données synthétiques Lyα, les soi-disant simulations Lyα, au chapitre 3. Ces simulations sont utilisées pour valider notre pipeline d'analyse Lyα, qui est décrit au chapitre 4. J'ai contribué en tant que membre actif de la collaboration DESI aux analyses comparatives du BAO de la forêt Lyα en utilisant des simulations et des données provenant à la fois d'eBOSS et de DESI. Cette partie de mon travail est présentée dans le chapitre 5, et est reprise dans plusieurs publications collaboratives du DESI. Je présente dans le chapitre 6 la contribution la plus importante de cette thèse à la collaboration DESI est l'analyse des systèmes à haute densité de colonne (ci-après HCD), l'un des effets systématiques les plus importants pour le BAO de la forêt Lyα. J'ai développé un modèle théorique, que j'appelle le modèle de Voigt, pour expliquer l'effet d'amortissement non local des HCD sur les fonctions de corrélation Lyα. Ce modèle fournit également une base physique pour expliquer les modèles phénoménologiques qui ont été utilisés dans les études précédentes.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction to Cosmology

Cosmology is the field of science studying the structure and evolution of the Universe. For centuries, observational tools and theoretical models have been developed to explain the observed universe. In the past century, modern cosmology was developed and greatly changed our understanding of the universe. In this chapter, I describe the theoretical foundations of modern cosmology, as well as the observational probes to constrain these models. The notions introduced in this chapter will be used throughout this thesis.

General Relativity (GR) provides a theoretical ground to develop cosmological models. According to observations, GR was modified several times. In Einstein's time, GR was used to describe a static universe with a cosmological constant. This was changed sooner after the observation of an expanding universe (Hubble 1929), where the astronomy community removed this constant to interpret a dynamical universe. In recent decades, observations of distant Type Ia supernovae (SNIa) indicated the acceleration of the expansion of the universe (Riess et al. 2000 ; Perlmutter et al. 1999), and the cosmological constant was reconsidered as Dark energy to explain the acceleration.

Besides Dark energy, Dark matter and baryons also play an important role in the evolution of the universe. The standard cosmological model is then built based on a geometrical setup of GR and an energy-matter setup of these energy components.

The evolution of the universe is interpreted based on the cosmological principle that it is homogeneous and isotropic (Bondi et Gold 1948 ; Hoyle, Burbidge et Narlikar 1993). The structures of the universe, both at large and smaller scales, offer different cosmological probes to study the matter distribution and time-dependent evolution of the universe.

In this chapter, I describe Baryon Acoustic Oscillations (BAO), one of the fundamental probes to study the expanding universe. Moreover, I introduce the Lyα forest, seen as absorptions in quasar spectra, that can be used to detect BAO at high redshift (see Chapter 5). Lyα forest of distant quasars are the main probe the work of this thesis is based on.

This chapter is inspired from reference books (Peter et Uzan 2009 ; Dodelson et F. Schmidt 2020) and previous theses (De Sainte Agathe 2019 ; Stermer 2022 ; Zarrouk 2018a ; Ravoux 2022 ; Chabanier, Palanque-Delabrouille et al. 2019).

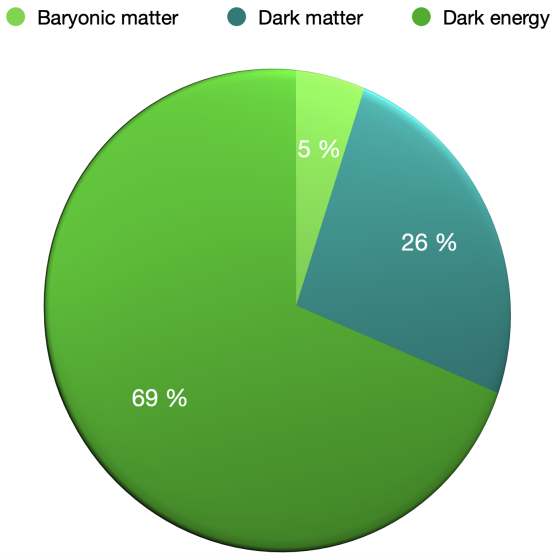

FIGURE 1.1 : Matter-energy contents of the Universe (AGHANIM et al. 2020).

1.1 The standard cosmology model

The observation of Hubble (Hubble's law, 1929) was interpreted as evidence for the expansion of the universe. It suggests a dynamic universe, currently well-modeled by the so-called standard cosmological model, the ΛCDM model. The ΛCDM model was tested with great success using the Cosmic Microwave Background (CMB), and the measurement of the acceleration of the expanding universe with type Ia supernovae (see Section 1.2 RIESS et al. 2000 ; PERLMUTTER et al. 1999). In this model, the universe was born after a 'Big Bang' around 13.8 billion years ago and evolved to its current state after several evolution phases, dominated by different contents.

1.1.1 The ΛCDM model

In this section, I describe the ΛCDM cosmological model with its parametrization, as well as the constraints from the CMB.

The content of the universe

According to the ΛCDM model, the universe is made of the following components (as shown in Figure 1.1) :

- Baryonic matter that can be observed by direct detections, such as stars and gas in galaxies, the intergalactic medium (IGM), etc. However, today only a small fraction (5%) of the total energy density is composed of baryonic matter, while most of the matter necessary to explain the dynamical evolution of galaxies and clusters remain undiscovered.
- Cold dark matter (CDM), the dominant matter contents (25%) of the universe today. CDM particles are non-relativistic, pressureless, and only interact through gravitational effects. CDM has a dominant impact on the large-scale structure (LSS) of the universe, as well as the formation of sub-structures at the galaxy scale. Potential candidates of CDM

particles cover a wide range of energy scales, such as Weakly interacting massive particles (WIMPs, Jungman, Kamionkowski et Griest 1996), axions (Donnelly, Freedman, Lytel, Peccei et Schwartz 1978 ; J. E. Kim et Carosi 2010), etc. However, up to today, no significant direct evidence has been found for these DM candidates with the past and ongoing experiments.

- Radiation and relativistic matter contents, such as photons and hot dark matter particles (HDM, such as relativistic neutrinos). These particles play an essential role in the evolution of the early universe, while also influencing the formation of LSS.
- Dark energy, the dominant contribution to the energy density of the universe today, which is the possible explanation of its late expansion. It is identified with a cosmological constant Λ in the standard cosmological model based on General Relativity (GR). Some extended theories suggest a dynamical dark energy equation of state, that can be well constrained using various probes, such as the Baryon Acoustic Oscillations (BAO, see Section 1.2.1), which is the purpose of this thesis.

Dark energy and CDM comprise most of the energy density of the universe today ($\sim 95\%$). The ΛCDM model and its extensions can be constrained by different cosmological probes, such as the CMB, the BAO, Weak Lensing (WL), Gravitational Waves (GW), Type Ia SNIa, etc.

The Cosmic Microwave Background

The CMB refers to the photon radiation, produced at the epoch of recombination (nearly $380,000$ years after the 'Big Bang'), at redshift $z \sim 1100$. Before recombination, photons, electrons, and protons were tightly coupled with each other by Compton and Coulomb scatterings, and formed a baryon-photon plasma. In this plasma, electrons, neutrons, and protons could not combine into Hydrogen atoms since the mean energy of photons was higher than the ionization level of atoms. At the epoch of recombination, the plasma had cooled enough for atoms to form. After recombination, photons can move freely in the universe without scattering with other particles. These photons are redshifted today to the microwave wavelength range and form the Cosmic Microwave Background. The anisotropic energy distribution of CMB photons can be further used to measure the matter density fluctuations in the baryon-photon plasma, and constrain cosmological models. The CMB was first detected in the 1960s (Penzias et R. W. Wilson 1965 ; Penzias et R. W. Wilson 1979), and the latest constraints were provided by the Planck satellite (Aghanim et al. 2020). Figure 1.2 shows the anisotropic temperature distribution of CMB photons, which imprints the matter fluctuation in the early universe before recombination. The red regions, i.e., hotter photons, trace over-densities, and the blue regions are related to lower-density regions.

To give a more detailed description of the temperature anisotropies, the relative temperature fluctuations are projected on a two-dimensional spherical space with two angular coordinates θ and ϕ, using a spherical harmonics basis Y_{lm} :

$$\frac{\delta T}{T}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta, \phi). \tag{1.1}$$

Here a_{lm} are the amplitudes of multipoles l and m.

The two-point correlation function is often used to characterize the statistical properties of the temperature fluctuations, and is defined as :

$$C_l = \langle |a_{lm}|^2 \rangle_m, \tag{1.2}$$

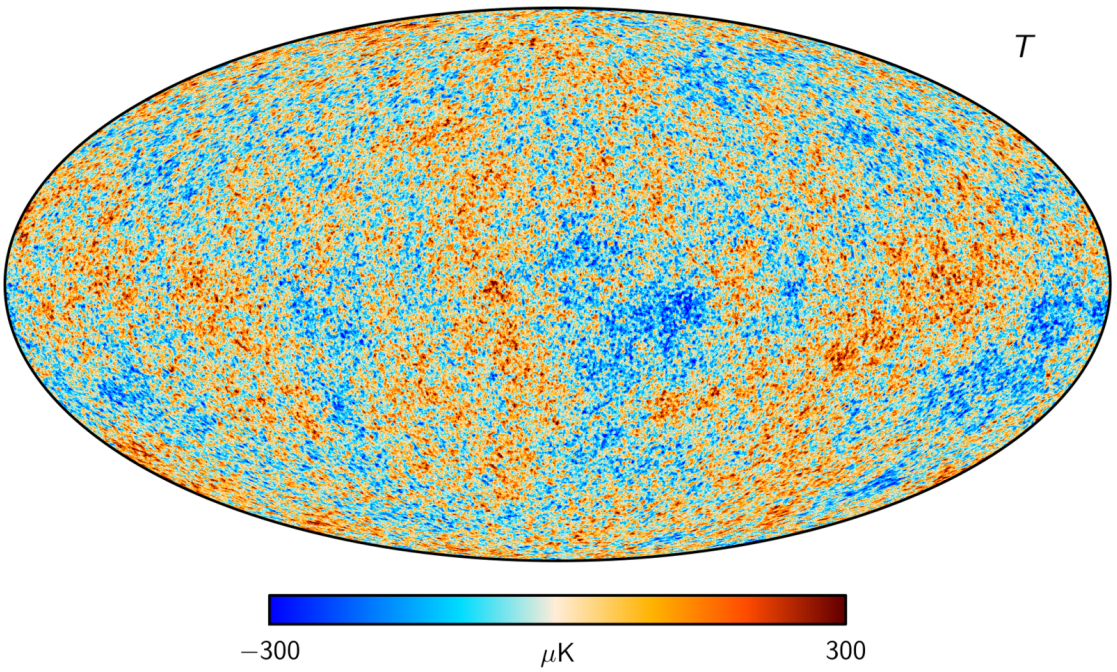

FIGURE 1.2 : Anisotropic temperature distribution of CMB photons, measured by AGHANIM et al. 2020. Blue regions show lower temperatures and red regions show higher temperatures.

where $\langle\rangle$ denotes an average over multipoles m. Figure 1.3 shows the power spectrum $D_l^{TT} = l(l+1)C_l/2\pi$ as a function of multipole l in angular space, which can be roughly divided into three regions :

- The large scales where $l < 100$ corresponds to the early universe, that can be used to detect the primordial fluctuations generated at inflation.
- The region between $l = 100$ and $l = 1000$ corresponds to the Baryon Acoustic Oscillations (BAO, see Section 1.2.1), and shows a series of acoustic peaks.
- The smaller scales for $l > 1000$ correspond to the last baryon-photon scattering surface during recombination.

Constraining the ΛCDM model

CMB is one of the fundamental cosmological probes to constrain the ΛCDM model. This model is parameterized by 6 parameters [1] :

- n_s : the spectral index of primordial scalar perturbations.
- $\Omega_b h^2$: the baryon density parameter today. Here $h = \frac{H_0}{100\ \mathrm{km{\cdot}s^{-1}{\cdot}Mpc^{-1}}}$, H_0 is the Hubble constant, and Ω_b is the baryon density as a fraction of the critical energy density of the universe today (see definition in Equation 1.32).

[1] Density parameters are functions of time and the ΛCDM model is parametrised as a function of their values as of today

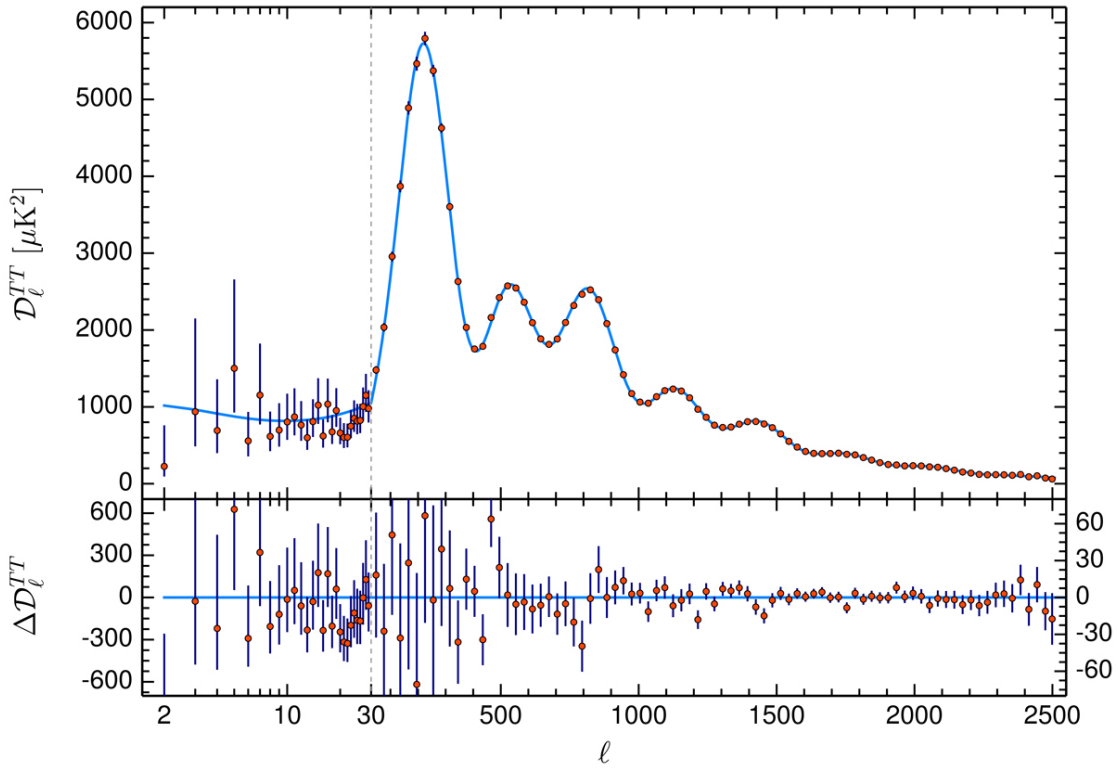

FIGURE 1.3 : Power spectrum of CMB photons measured by AGHANIM et al. 2020. The two-point correlation function C_l is multiplied by $l(l+1)/2$ to better see the oscillations.

TABLEAU 1.1 : Parameter constraints on the ΛCDM model using CMB data from AGHANIM et al. 2020.

Parameters	Constraints
n_s	0.9652 ± 0.0042
$\Omega_b h^2$	0.02233 ± 0.00015
$\Omega_b c^2$	0.1198 ± 0.0012
τ	0.0540 ± 0.0074
$\ln(10^{10})A_s$	3.043 ± 0.014
$100\theta_{MC}$	1.04089 ± 0.00031
$\Omega_m h^2$	0.1428 ± 0.0011
$H_0[\ \mathrm{km \cdot s^{-1} \cdot Mpc^{-1}}]$	67.37 ± 0.54
Age[Gyr]	13.801 ± 0.024
σ_8	0.8101 ± 0.0061
$S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$	0.830 ± 0.013
z_{Re}	7.64 ± 0.74
$100\theta_*$	1.04108 ± 0.00031
r_{drag}[Mpc]	147.18 ± 0.29

- $\Omega_c h^2$: the cold dark matter density parameter today. Here Ω_c is the cold dark matter density as a fraction of the critical energy density of the universe (see definition in Equation 1.32).
- τ : the optical depth at the epoch of reionization. Reionization refers to the time when the earliest galaxies formed, and emitted photons that reionized the surrounding Neutral Hydrogen atoms. This effect is mostly seen at high l.
- $\ln(10^{10})A_s$: where A_s is the amplitude of the primordial matter power spectrum.
- $100\theta_{MC}$: where θ_{MC} is the apparent angle of the first acoustic peak's position.

The latest measurement from the Planck satellite constrained the ΛCDM parameters with a 1% precision, as shown in Table 1.1. The first six parameters are ΛCDM parameters, and the other ones are derived parameters :

- $\Omega_m h^2$: where Ω_m is the total matter density today in the universe (see definition in Equation 1.32), as a fraction of the critical energy density, and multiplied by h^2.
- $H_0[\ \mathrm{km \cdot s^{-1} \cdot Mpc^{-1}}]$: Hubble constant.
- Age[Gyr] : age of the universe.
- σ_8 : the normalization of matter fluctuations today, averaged over spheres of radius $8h^{-1}$Mpc.
- S_8 : a derived parameter using σ_8 and Ω_m, taking into account their correlations.
- z_{Re} : redshift of reionisation.
- $100\theta_*$: numerical result for the acoustic scale angle $100\theta_{MC}$ (see definition above).
- r_{drag} : the comoving scale of the BAO sound wave at the epoch of recombination (see Section 1.2.1).

In the following sections, I will describe the theoretical basis of the ΛCDM model, as well as the equations of state of the different energy density components.

Redshifts in cosmology

Redshift is one of the fundamental concepts in cosmology since most observations capture signals coming from distant astrophysical objects. It is commonly defined as :

$$1 + z = \frac{\lambda_{\rm obs}}{\lambda_{\rm RF}}, \tag{1.3}$$

where $\lambda_{\rm obs}$ and $\lambda_{\rm RF}$ refer to the wavelength in the observation frame and the rest frame, respectively.
There are two notions of redshift used in cosmology : the one due to Doppler shift, and the cosmological redshift. The cosmological redshift (related to the expansion of the universe) is defined in Equation 1.25, as a function of the cosmology scale factor a. In particular, $z = 0$ corresponds to the present epoch for which $a = a_0$. The Doppler redshift of an astrophysical object is defined as the Doppler shift of its signal due to its peculiar velocity v :

$$1 + z \approx \sqrt{\frac{1 + \frac{v}{c}}{1 - \frac{v}{c}}}. \tag{1.4}$$

This formula can be approximated as $z \approx \frac{v}{c}$ for low-redshift ($z \ll 1$) astrophysical objects, which is the method that Hubble used to measure the distance and velocities of a few nearby galaxies, and proposed Hubble's law (see Figure 1.4) :

$$v = H_0 D. \tag{1.5}$$

Here H_0 is the Hubble constant, and D is the distance between the observer and the galaxy.

1.1.2 General relativity in cosmology

The ΛCDM model is based on an isotropic and homogeneous universe described by General Relativity (GR). For simplicity, I use in the following sections the usual convention of units :

$$c = \hbar = k_B = 1. \tag{1.6}$$

Einstein's theory of GR assumes that a massive object placed in a gravitational field is affected by the same gravitational strength, independently of its nature. This assumption is called the Equivalence Principle. In Einstein's theory, this gravitational effect is a consequence of curved space-time, thus enabling a geometrical equivalence of gravitational interaction.

The geometry of the space-time

In general, the invariant quantity ds^2 of a 4-dimensional space-time is described by a metric-tensor $g_{\mu\nu}$:

$$ds^2 = g_{\mu\nu}(\mathbf{x}) dx^\mu dx^\nu, \tag{1.7}$$

where $\mathbf{x} = (x^0, x^1, x^2, x^3)$ with x^0 representing the time coordinate, and x^1, x^2, x^3 referring to the space coordinates. For a Minkowski space (considering no curvature), $g_{\mu\nu}$ is equivalent to the Minkowski metric $\eta_{\mu\nu}$, which is a pseudo-tensor :

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{1.8}$$

Einstein's equation

Einstein's equation of GR describes gravity using an equivalent form of the space-time geometry and the energy density of objects. On one hand, the Einstein tensor $G_{\mu\nu}$ can be expressed in terms of geometric tensors :

$$G_{\mu\nu} = R_{\mu\nu} - \frac{R}{2} g_{\mu\nu}, \tag{1.9}$$

where $R_{\mu\nu} = R^k_{\mu k\nu}$ is the Ricci tensor, and $R = R^\mu_\mu$ is the Ricci scalar. Here $R^k_{\mu k\nu}$ is the Riemann curvature tensor, defined as :

$$R^i_{jkl} = \frac{\partial \Gamma^i_{lj}}{\partial x^k} - \frac{\partial \Gamma^i_{kj}}{\partial x^l} + (\Gamma^i_{kp}\Gamma^p_{lj} - \Gamma^i_{lp}\Gamma^p_{kj}). \tag{1.10}$$

Here Γ is the Christoffel symbol, defined by considering the transformation of a dynamical metric :

$$\Gamma^m_{ij} = \frac{1}{2} g^{mk}\left(\frac{\partial}{\partial x^j} g_{ki} + \frac{\partial}{\partial x^i} g_{kj} - \frac{\partial}{\partial x^k} g_{ij}\right) = \frac{1}{2} g^{mk}(g_{ki,j} + g_{kj,i} - g_{ij,k}). \tag{1.11}$$

On the other hand, Einstein's equations relate the geometric expression of Equation 1.9 with the energy content of the universe :

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.12}$$

Here $T_{\mu\nu}$ is the energy-momentum tensor which describes the energy content of the universe. The combination of Equation 1.9 and Equation 1.12 indicates that the existence of massive objects changes the curvature of space-time, and all the other objects move following a geodesics path in this curved space-time (Equivalence Principle) :

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{ij} \frac{dx^i}{ds} \frac{dx^j}{ds} = 0, \tag{1.13}$$

where s is a scalar parameter of motion, e.g., in Minkowski space it is the proper time.

Distances in cosmology

There are several important notions of distance in cosmology, including :

- The comoving distance : the distance between two objects that is independent of time. It has different expressions along and across the line-of-sight.

  Along the line-of-sight, it is defined as :

$$D_{\mathrm{C}} = \int_0^z \frac{c dz'}{H(z')}, \tag{1.14}$$

  which is an integration of the inverse of the Hubble parameter over redshift from the observer at $z = 0$ to the astrophysical object at redshift z.

  Along the transverse direction across the line-of-sight of a distant astrophysical object, the comoving distance is defined as :

$$D_{\mathrm{M}} = \begin{cases} D_{\mathrm{H}} \frac{1}{\sqrt{\Omega_k}} \sinh\left(\sqrt{\Omega_k} \frac{D_{\mathrm{C}}}{D_{\mathrm{H}}}\right), & \Omega_k > 0 \\ D_{\mathrm{C}}, & \Omega_k = 0 \\ D_{\mathrm{H}} \frac{1}{\sqrt{\Omega_k}} \sin\left(\sqrt{\Omega_k} \frac{D_{\mathrm{C}}}{D_{\mathrm{H}}}\right), & \Omega_k < 0 \end{cases} \tag{1.15}$$

  Here Ω_k is the curvature energy density (see definition in Equation 1.32), $D_{\mathrm{H}} = \frac{c}{H_0}$ is the Hubble distance, and one can figure out that $D_{\mathrm{M}} = D_{\mathrm{C}}$ for a flat universe ($k = 0$).

- The angular diameter distance D_{A} : the angular diameter distance of an astrophysical object with an actual size d and an angular size $\Delta\theta$ is defined as $D_{\mathrm{A}} = \frac{d}{\Delta\theta}$. In cosmology, it interprets the physical distance between two astrophysical objects. It relates to D_{M} by :

$$D_{\mathrm{A}} = \frac{D_{\mathrm{M}}}{1+z}, \tag{1.16}$$

  where z is the redshift of the astrophysical object.

- The luminosity distance D_{L} : the distance from an astrophysical source with a measured luminosity L (the emitted total electromagnetic energy per unit of time) from the total flux F (the total energy that crosses a unit area per unit time), defined as :

$$D_{\mathrm{L}} = \sqrt{\frac{L}{4\pi F}}. \tag{1.17}$$

A general relation between these three distances is :

$$D_{\mathrm{L}} = (1+z) D_{\mathrm{M}} = (1+z)^2 D_{\mathrm{A}}. \tag{1.18}$$

1.1.3 An expanding universe

The expansion of the universe was first discovered thanks to the observation of HUBBLE 1929. Figure 1.4 shows the measured velocities of galaxies as a function of their distances from the observer. The slope of the fitted line shows the proportionality of galaxy peculiar velocities with their distances from the observer, and is the so-called Hubble constant (H_0 in Equation 1.5).

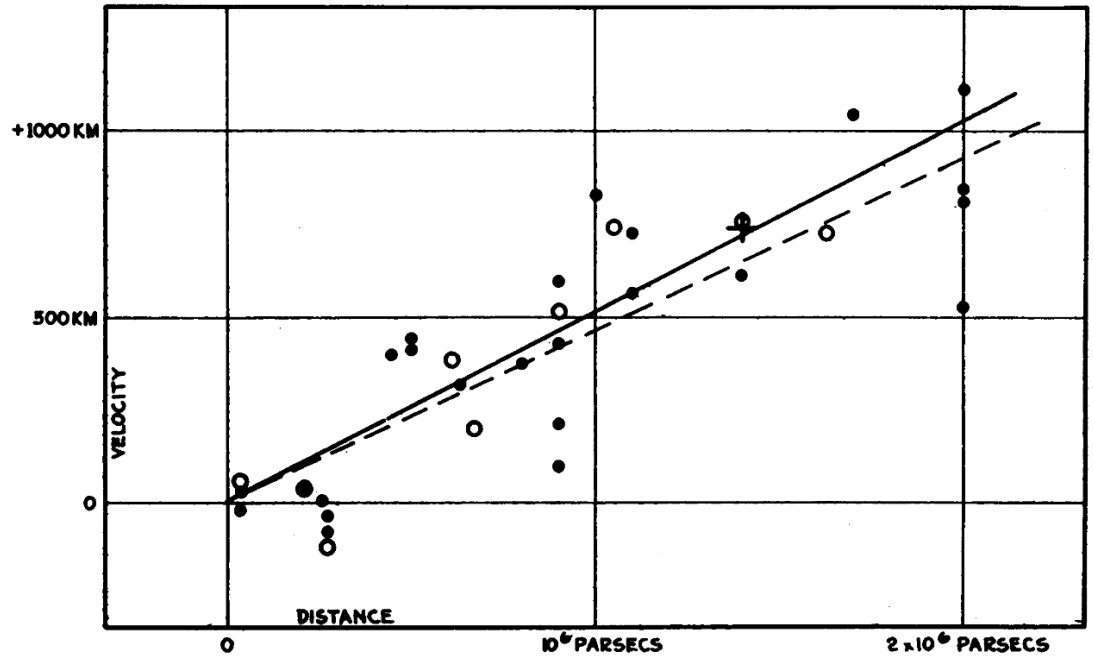

FIGURE 1.4 : Hubble diagram measured using neaby galaxies (HUBBLE 1929). It shows the velocities of galaxies as a function of their distances from the observer, indicating that the universe is expanding.

The metric of a dynamical universe

Based on Einstein's equations and knowing from Hubble's law that the universe is expanding, I describe in this subsection the geometry of an expanding universe. There are two fundamental principles to describe the observed matter distribution of our universe :

- The universe is homogeneous, meaning that the matter distribution is uniform, seen from a much larger scale than the galaxy scale.
- The universe is isotropic, seen from all directions.

These two principles form the cosmological principle and allow the universe to be described by a so-called Friedmann-Robertson-Walker (FRW) metric (ROBERTSON 1936 ; WALKER 1937) with only two parameters $a(t)$ and k :

$$ds^2 = c^2dt^2 - a(t)^2[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2(\theta)d\phi^2)], \tag{1.19}$$

with a metric tensor :

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{a^2}{1-kr^2} & 0 & 0 \\ 0 & 0 & -a^2r^2 & 0 \\ 0 & 0 & 0 & a^2r^2\sin^2\theta \end{pmatrix}. \tag{1.20}$$

Here $a(t)$ is the time-dependent scale factor of the universe, and $k = \frac{R}{6}$ is the curvature of the universe. $k = 0$ describes a flat and infinite universe, $k > 0$ a spherical and finite universe, and $k < 0$ an hyperbolic and infinite one.

Dynamics in the universe

Apart from geometry, the cosmological principle of an expanding universe also implies an energy-momentum tensor considering the universe as a perfect fluid in thermodynamical equilibrium :

$$T_{\mu\nu} = (\rho + P)u_\mu u_\nu - P g_{\mu\nu}. \tag{1.21}$$

Here P is the pressure of the fluid, ρ is the energy density, and u is the fluid velocity four-vector.

By solving Einstein's equation with the FRW metric and the energy-momentum tensor in Equation 1.21, two independent dynamic equations can be obtained :

$$(\frac{\dot{a}}{a}) = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1.22}$$

and

$$(\frac{\ddot{a}}{a}) = \frac{4\pi G}{3}(\rho + 3P). \tag{1.23}$$

Here the dots denote time derivatives. Equation 1.22 tells how fast the universe is expanding. The Hubble parameter is thus defined as :

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.24}$$

The scale factor of today's universe is noted a_0, and the cosmological redshift is equal to :

$$1 + z = \frac{a_0}{a}. \tag{1.25}$$

Since a is a function of time, this equation provides an equivalence of distance, time and redshift : low z indicates a closer distance and time from the current observer, while high z the opposite.

The evolution of the universe

Combining Equation 1.22 and 1.23, one obtains a derived equation :

$$\dot{\rho} + \frac{\dot{a}}{a}[\rho + P] = 0. \tag{1.26}$$

Considering the equation of state of fluid as $P = \rho w$ (w is the equation of state parameter of a perfect fluid), the integration of Equation 1.26 yields :

$$\rho(t) = \rho(t_0)(\frac{a_0}{a(t)})^{3(w+1)}. \tag{1.27}$$

Different density contents (see Section 1.1.1) are described by different w values, thus implying different structure evolution :

- Baryons and CDM : are pressureless, so that $P \ll \rho$. The density evolution is thus :

$$\rho_m(a) \propto a^{-3}. \tag{1.28}$$

- Radiation : the equation of state is $P = \frac{\rho}{3}$, which yields

$$\rho_r(a) \propto a^{-4}. \tag{1.29}$$

- Cosmological constant : The cosmological constant Λ can be considered as a fluid of constant density with negative pressure $P = -\rho$. In this case, the dark energy density will be independent of time and expansion of the universe :

$$\rho_\Lambda(a) = \text{constant}. \tag{1.30}$$

The total energy density in the universe can be further defined as the sum of all these contents :

$$\Omega_{\text{total}} = \Omega_{\text{m}} + \Omega_{\text{r}} + \Omega_\Lambda = 1 - \Omega_k. \tag{1.31}$$

Here each energy density content is normalised by the critical density of the universe $\rho_{\text{crit}} = \frac{3H^2}{8\pi G}$, that is :

$$\begin{aligned} \Omega_i &= \frac{\rho_i}{\rho_{\text{crit}}},\ i = \text{m, r, } \Lambda, \\ \Omega_k &= \frac{-kc^2}{a^2H^2}. \end{aligned} \tag{1.32}$$

The first Friedmann equation 1.22 can therefore be rewritten as :

$$H^2(t) = H_0^2[\Omega_{\text{r},0}a(t)^{-4} + \Omega_{\text{m},0}a(t)^{-3} + \Omega_{k,0}a(t)^{-2} + \Omega_{\Lambda,0}], \tag{1.33}$$

where subscript 0 denotes present-time values. This is visualized in Figure 1.5 for a flat universe ($k = 0$) as a function of time (redshift). One can see from the plot that radiation dominated the evolution of the early universe, matter dominated the universe between $0.5 < z < 3600$, and the current universe is governed by Dark Energy (here a cosmological constant).

1.2 The accelerated expansion of the universe

In the initial formulation of Einstein's equation, a cosmological constant was introduced as a correction of space curvature to ensure a static universe. This constant was removed after the measurement of Hubble's law (Hubble 1929), allowing GR to interpret an expanding universe. In this case, a homogeneous universe as described in previous sections will slow down its expansion with the evolution of gravity.

However, around 70 years later, two independent research groups (Riess et al. 2000 ; Perlmutter et al. 1999) discovered that the universe expansion was accelerating using Type Ia supernovae (SNIa) up to redshift $z \sim 0.7$. These measurements were then explained by the ΛCDM model with a flat universe $\Omega_k = 0$ and $\Omega_\Lambda = 0.7$ at present time, considering the cosmological constant Λ again in GR, as an energy component.

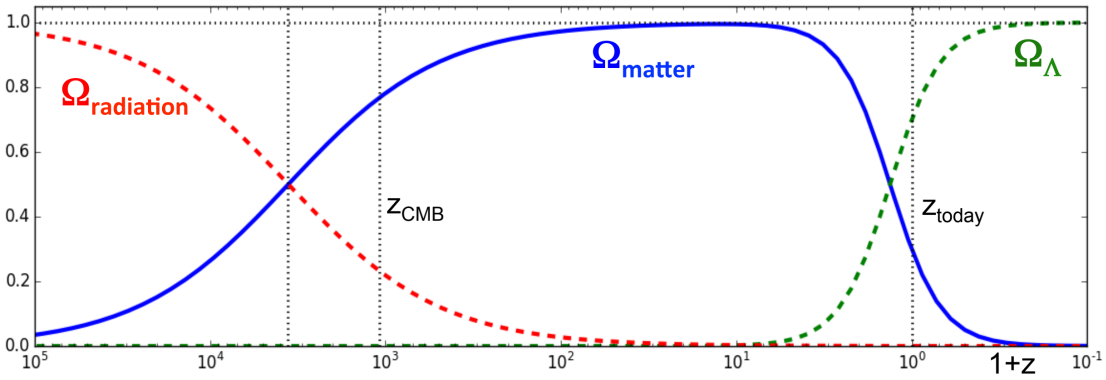

FIGURE 1.5 : The evolution of the energy density for a flat universe ($k = 0$), as a function of redshift, for different components of the universe : baryons and CDM (blue), cosmological constant (green), and radiation (red). Credits : ZARROUK 2018b.

Type Ia supernovae

SNIa are interpreted as extreme explosions of white dwarfs that accrete matter from neighboring stars (YOON et LANGER 2004 ; MAZZALI, ROPKE, BENETTI et HILLEBRANDT 2007) to a critical mass limit [a].
SNIa provides precise measurements of the luminosity distance D_L (see previous section). These luminosity distances can be used to constrain Ω_m and Ω_Λ using the ΛCDM model. However, D_L do not directly measure H_0, and the constraints on H_0 depend on both SNIa and distance ladder (nearby distances measured using Cepheid stars) (MÖRTSELL, GOOBAR, JOHANSSON et DHAWAN 2022). This distance ladder constraint on H_0 has a discrepancy of several σ compared to the results from other probes (see Figure 1.6, DI VALENTINO et al. 2021), e.g., CMB. This discrepancy is called the Hubble tension, and is still under investigation by the astronomy community.

[a] The mass limit is the so-called Chandrasekhar limit (BETHE, BROWN et WORPOLE 1985 ; MAZZALI, ROPKE, BENETTI et HILLEBRANDT 2007)

In order to understand the expansion history of the universe and forecast its future, It is essential to study the nature of Dark energy and its time-dependent equation of state with some extended models. In next section, I describe another independent probe besides SNIa, the Baryon Acoustic Oscillations (BAO), to measure the expansion of the universe at different redshifts.

1.2.1 The Baryon Acoustic Oscillations

The Baryon Acoustic Oscillations (BAO) is another powerful probe to measure the expansion of the universe and provide constraints on the ΛCDM model. Studying this probe is the goal of my thesis, and I will describe in this subsection the physics of the BAO.

The inflation of the universe right after the 'Big Bang' generated density fluctuations of baryons, Dark matter, and photons, due to the quantum fluctuations of initial fields. These inhomogeneities of the matter density field evolved into overdense and underdense regions, and turned into the large-scale structure of the universe, as a result of gravitational clustering.

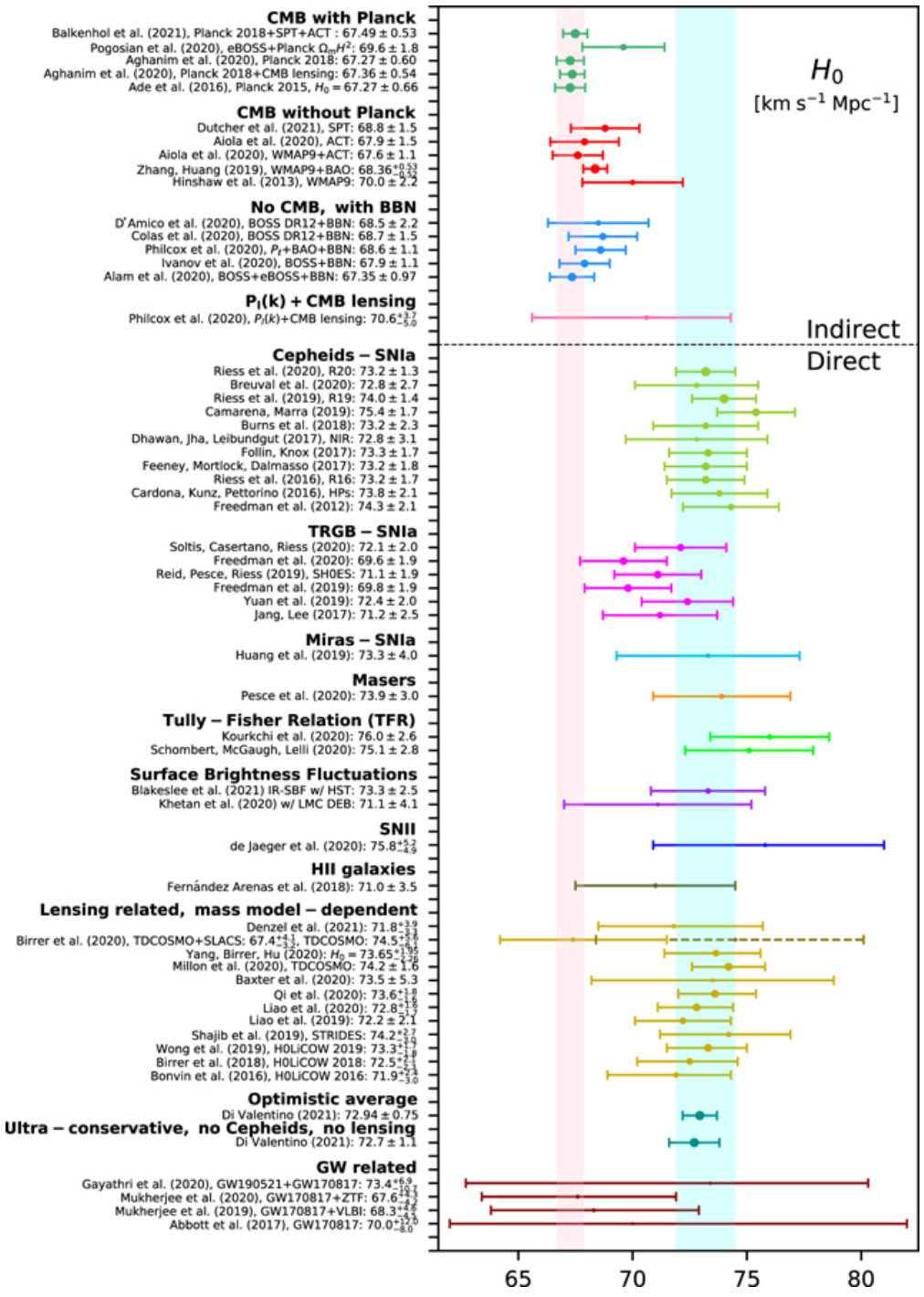

FIGURE 1.6 : The constraints on H_0 from different probes and different experiments. Credits : DI VALENTINO et al. 2021.

Before recombination, the primordial density fluctuations generated during inflation propagated in the form of sound waves in the baryon-photon plasma, and left a shell at the sound horizon $r_{\rm d}$, after the baryon-photon decoupling. The sound horizon is therefore defined as the maximum length of this sound wave that traveled until decoupling :

$$r_{\rm d} = \int_{z_{\rm drag}}^{\infty} \frac{c_{\rm s}}{H(z)}. \tag{1.34}$$

Here $z_{\rm drag}$ refers to the drag epoch where the photons were fully decoupled from baryons. Figure 1.7 (D. J. Eisenstein, H.-j. Seo, Sirko et Spergel 2007) shows the mass profile of the different components of the baryon-photon plasma before and after recombination. Before recombination (two upper plots), baryons and photons were tightly coupled so they followed the same mass profile. The primordial fluctuations generated by inflation then propagated as a sound wave. After recombination, photons decoupled from the plasma and left the previous fluctuations as a peak in the correlation functions of matter tracers. One can see in Figure 1.7 the BAO peak from baryons and CDM, at a scale $\sim 100\ h^{-1}$ Mpc.

1.2.2 The two-point correlation function

As mentioned in the previous section, the BAO peak is seen in the fluctuations of baryons and CDM. In this regard, statistical approaches such as the two-point correlation function of matter tracers can be used to measure this peak.

Consider the perturbation of the matter density field ρ at a position $\vec{x}$ as the density contrast (δ refers to the density field) :

$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\bar{\rho}} - 1. \tag{1.35}$$

The two-point correlation function (2PCF) at a separation $\vec{r}$ is expressed as :

$$\xi(\vec{r}) = \langle \delta(\vec{x})\delta(\vec{x}+\vec{r}) \rangle, \tag{1.36}$$

where the brackets $\langle\rangle$ denote the mean average over $\vec{x}$.

The associated power spectrum is then derived as the Fourier Transform of the two-point correlation function :

$$P(\vec{k}) = \int_{-\infty}^{+\infty} \xi(\vec{r}) e^{(-i\vec{k}\cdot\vec{r})} d^3\vec{r}. \tag{1.37}$$

1.2.3 Matter tracers

The BAO signal can be detected as a peak in the 2PCF using discrete matter tracers (galaxies, quasars, voids, etc) or continuous matter tracers (Lyα forests). These matter tracers are usually considered biased matter density fields in the first- or second-order approximations.

The bias

Consider the matter tracers $\delta_{\rm T}$ at a redshift z as a biased tracer of the underlying dark matter density field :

$$\delta_{\rm T}(\vec{r}, z) = b_{\rm T}(z)\delta_{\rm matter}(\vec{r}, z). \tag{1.38}$$

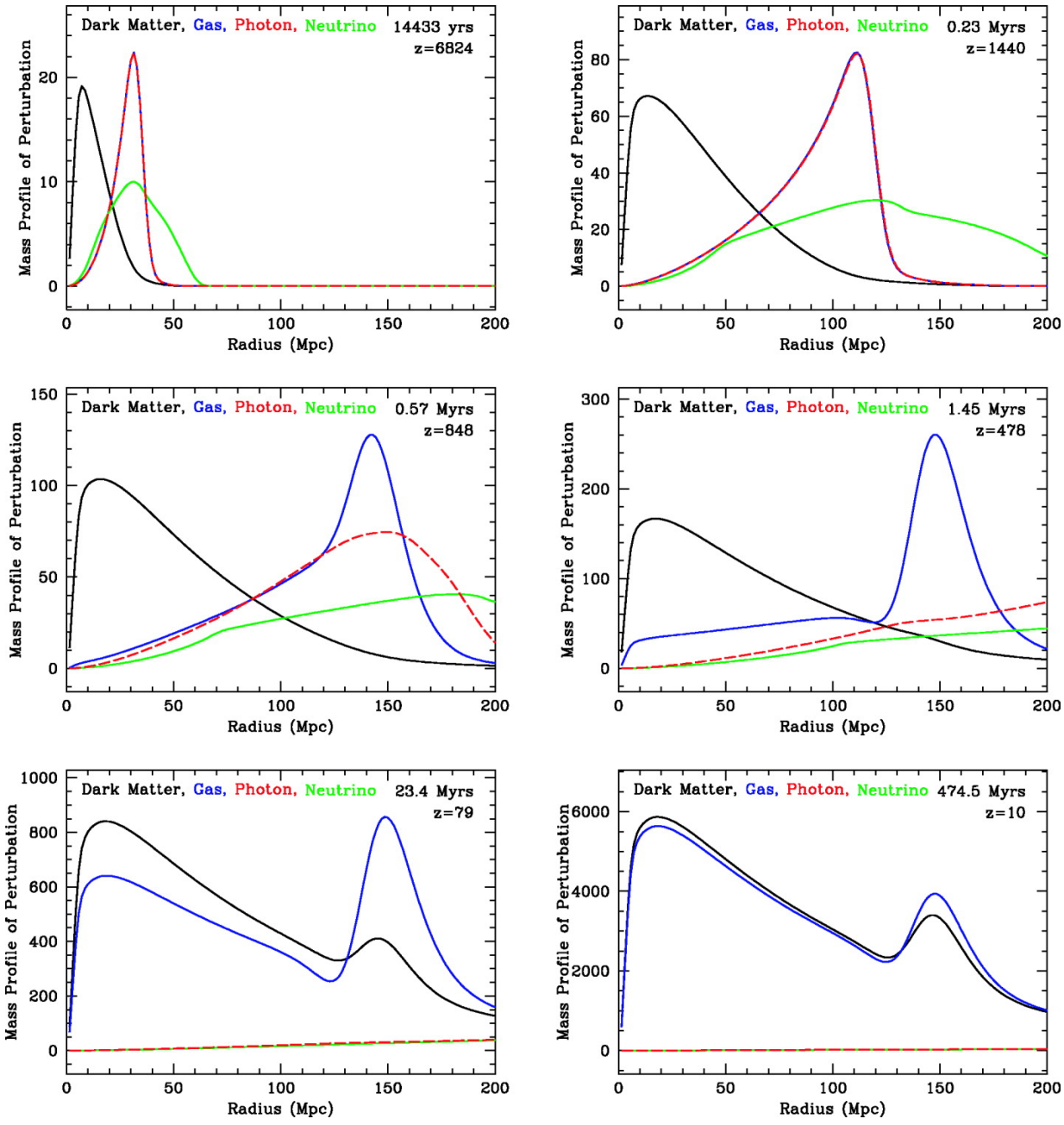

FIGURE 1.7 : The mass profile of different matter contents in the baryon-photon plasma before and after the recombination. Credits : D. J. EISENSTEIN, H.-j. SEO, SIRKO et SPERGEL 2007.

The 2PCF is then derived as :

$$\xi_{\mathrm{T}}(\vec{r}, z) = b_{\mathrm{T}}^2(z)\xi_{\mathrm{matter}}(\vec{r}, z). \tag{1.39}$$

Redshift distortions

Due to gravitation, matter in galaxy clusters is attracted by over-density regions and has a peculiar velocity directed towards these regions. As a result, matter on opposite sides of an over-density regions along or across a line-of-sight will have opposite contributions to the measured redshift.

In practice, the measured redshifts of galaxy clusters are affected by their peculiar velocities, which implies :

$$z_{\mathrm{obs}} = z_{\mathrm{cosmo}} + z_v, \tag{1.40}$$

where z_{cosmo} refers to the cosmological redshift and z_v is a redshift due to peculiar velocities (the so-called redshift space distortions effect, RSD). These two redshifts can not be distinguished from measurements, and thus the RSD need to be modeled. Figure 1.8 details different cases of this effect : the upper plots show a flattened measurement (when peculiar velocity is smaller than the cluster scale) of a cluster in redshift space, which is interpreted as a squashing or Kaiser effect (Kaiser 1987) ; the bottom plots present the case where the galaxy distribution is elongated along the line-of-sight (when peculiar velocity is larger than the cluster scale), and is referred to as the 'Finger of God' (FoG) effect (Jackson 1972).

Considering the redshift-dependent density field in Fourier space :

$$\hat{\delta}(\vec{k}, z) = \int \hat{\delta}(\vec{x}, z) e^{(-i\vec{k}\cdot\vec{x})} d^3\vec{x}, \tag{1.41}$$

then the RSD effect can be modeled as :

$$\hat{\delta}_{\mathrm{RSD}}(\vec{k}, z) = (1 + f(z)\mu_k^2)\hat{\delta}(\vec{k}, z), \tag{1.42}$$

which is the simplest RSD model, that is used for Lyα analysis. Here $\mu_k = \frac{\vec{k}}{||\vec{k}||} \cdot \vec{\mu}$ ($\vec{\mu}$ is the direction along the line-of-sight). f is the linear growth rate as a function of the scale factor a, and is defined as (A. Hamilton 2001) for the ΛCDM model :

$$f(a) = \frac{d\ln(D(a))}{d\ln a} = -1 - \frac{\Omega_{\mathrm{m}}}{2}(1 - \frac{5a}{D(a)}) + \Omega_\Lambda, \tag{1.43}$$

with

$$D(a) = \frac{2\Omega_{\mathrm{m}}}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3}. \tag{1.44}$$

1.2.4 The Lyman-α forest

I describe in this subsection the physics of quasars and Lyman-α (Lyα) forests, seen as series absorption lines in the quasar spectrum (between the Lyβ and Lyα emission peak).

Quasars

Quasars, or quasi-stellar objects, are the most luminous objects in the universe. They were first observed in 1963 by Matthews et Sandage 1963 and their spectra are measured by

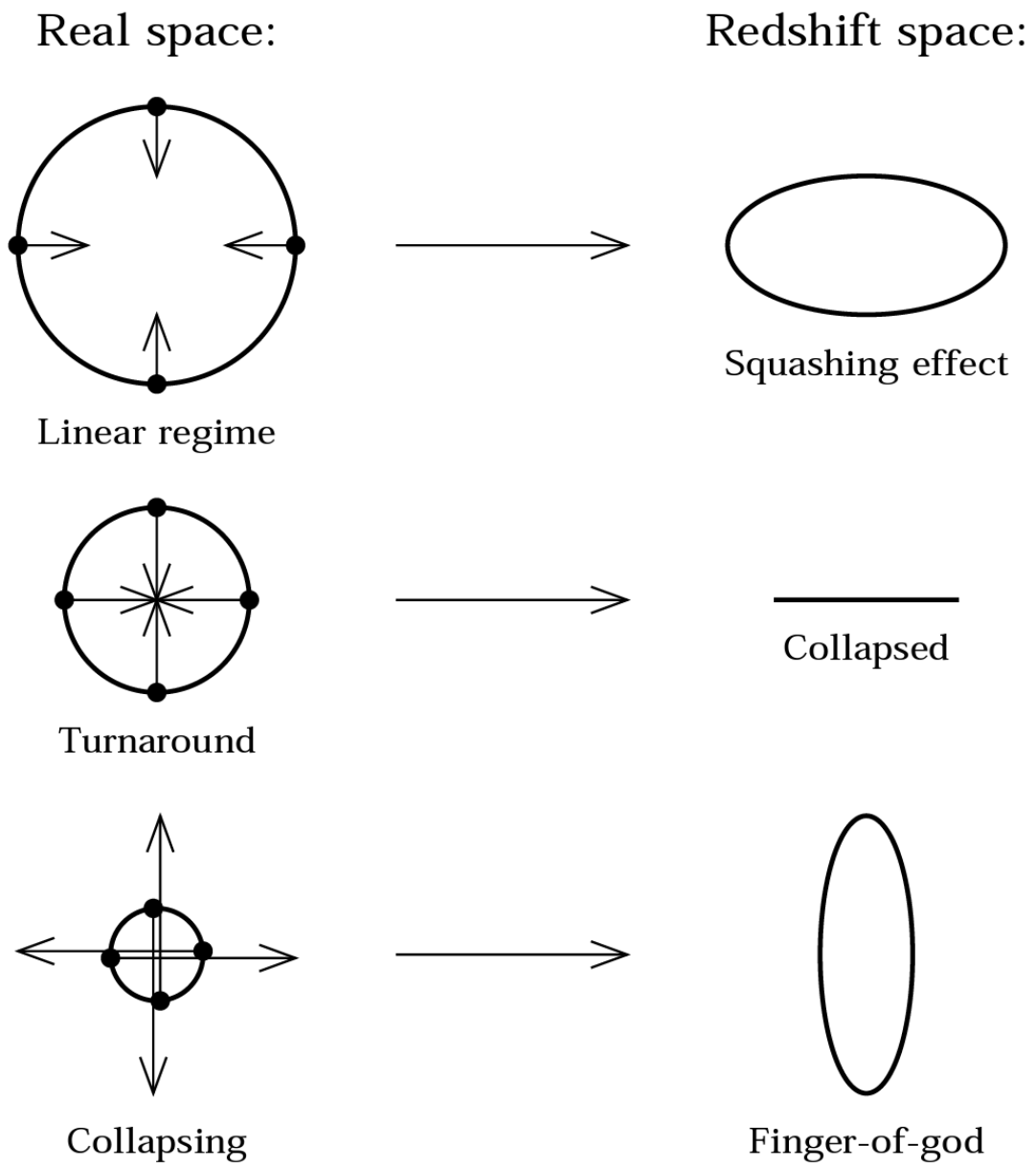

FIGURE 1.8 : The example of the impact of peculiar velocities on the redshift measurement distortions. Credits : A. HAMILTON 1998.

Region	$\lambda_{\rm RF}^{\rm min}$[Å]	$\lambda_{\rm RF}^{\rm max}$[Å]
Lyα	1040	1216
Lyβ	920	1020

TABLEAU 1.2 : The wavelength intervals of Lyα and Lyβ regions in rest frame.

GREENSTEIN et M. SCHMIDT 1979. Quasars are point-like objects with a continuum, very broad emission lines, and some absorptions (see an example in Figure 1.9). They are sub-classes of active galactic nuclei (AGN), with supermassive black holes highly active at the galaxy center, and accreting the surrounding gas or dust into a disk. AGNs eject their matter as luminous jets, and quasars are AGNs with jets directly pointing at the observer.

Quasars are characterized by strong emissions in radio, infrared, optical, ultraviolet, X-ray, and gamma-ray wavelength ranges. The emission spectral lines of different atoms (Hydrogen, helium, carbon, magnesium, iron, oxygen, etc.) are broadened due to the large velocities and heating of the rotating matter close to the quasar (Doppler broadening).

The strong Lyα emission line of quasars makes it possible to study the absorber features along the line-of-sight, and I will introduce Lyα forest in the next sub-section, which can be used to probe cosmological models.

Lyα forests

When photons emitted by quasars go across the intergalactic medium (IGM) in the universe and meet with gases of Neutral Hydrogen atoms, they have a probability to excite these atoms and produce a Lyα (wavelength of 1215.67 Å in the rest frame) absorption in the quasar spectra. This effect generated by IGM at different redshifts shifts the absorption in quasar spectra at an observing wavelength $\lambda_{\rm obs} = (1 + z_{\rm IGM})\lambda_{\rm Ly\alpha}$. The collection of all shifted Lyα absorption lines by hydrogen along the line-of-sight is called the Lyα forest. Figure 1.9 shows a quasar spectrum ($z_{\rm QSO} = 2.96$) observed by DESI. One can see the Lyα forest range $\lambda \in [1040, 1216]$ Å rest-frame between the Lyα and Lyβ peaks. The black curve shows the estimated quasar continuum, which is the quasar spectrum without any Lyα absorption. The Lyα absorption lines also exist on the left side of the Lyβ peak, and are mixed with the Lyβ absorption lines. Lyα forests can be used as continuous matter tracers to detect the BAO at high redshift, $z > 2$. The combined analysis of Lyα forests in both the Lyα region ($\lambda_{\rm rf} \in [1040, 1216]$ Å see Table 1.2) and Lyβ ($\lambda_{\rm rf} \in [920, 1020]$ Å)region is discussed in DES BOURBOUX, RICH et al. 2020, and I will not focus on this in my thesis. Consider these photons with total flux F going through the IGM across length dL and Neutral Hydrogen column density $n_{\rm HI}$. The absorbed flux dF is then expressed as :

$$dF = n_{\rm HI}\sigma_{\rm Ly\alpha}F dL, \tag{1.45}$$

where $\sigma_{\rm Ly\alpha}$ is the Lyα cross-section. Integrating this equation one get :

$$F = F_0 e^{-\tau}, \tag{1.46}$$

where τ is defined as the optical depth :

$$\tau = \sigma_{\rm Ly\alpha} \int n_{\rm HI} dL. \tag{1.47}$$

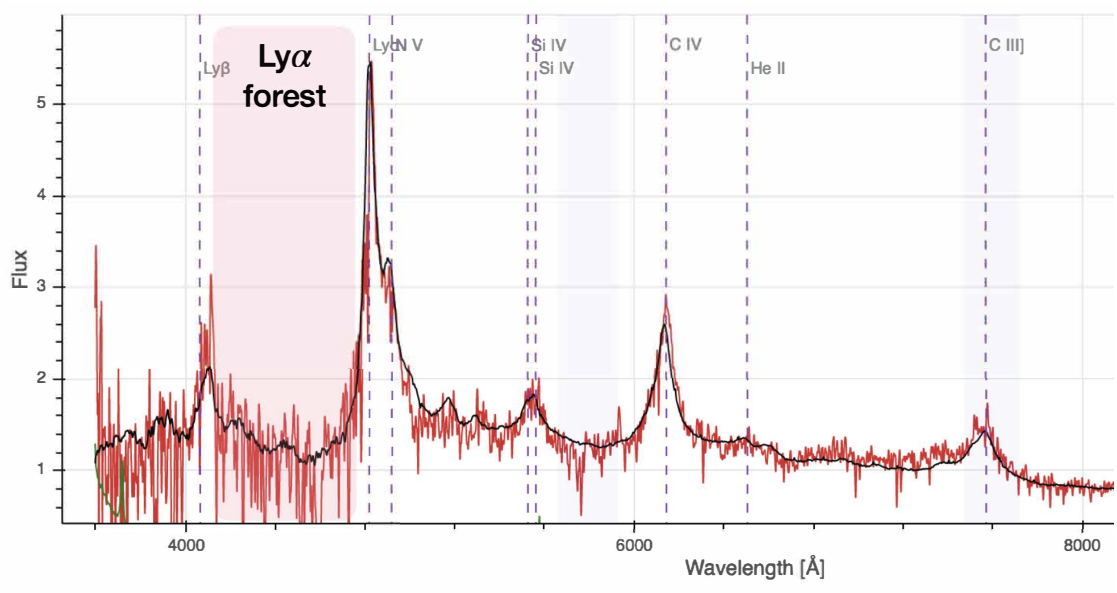

FIGURE 1.9 : An example of a quasar spectrum ($z_{\rm QSO} = 2.96$) observed by DESI.

The flux fluctuations of $\delta_{\rm Ly\alpha}(\vec{x}, z) = \frac{F_{\rm Ly\alpha}(\vec{x},z)}{\bar{F}} - 1$ can then be used as continuous biased matters to detect the BAO. I will further describe the technical details of measuring the BAO using Lyα forests in Chapter 5, which is the main goal of this thesis.

High Column Density systems

In the IGM, bound gas concentrations of Neutral Hydrogen (hereafter HI) atoms with HI column densities $10^{20.3}{\rm cm}^{-2} > N_{\rm HI} > 10^{17.2}{\rm cm}^{-2}$ are called Lyman limit systems (LLS), and those with $N_{\rm HI} > 10^{20.3}{\rm cm}^{-2}$ are called Damped Lyman-alpha systems (DLAs). In this manuscript I will call the combination of these two categories of systems the High Column Density systems (HCDs) with $N_{\rm HI} > 10^{17.2}{\rm cm}^{-2}$. HCDs are seen as strong absorptions with damping wings in the Lyα forests (see Figure 1.10 a DLA centered at $z_{\rm DLA} = 3.286$.). Their damping wings are usually parametrized by Voigt profiles (see Section 6.2.3), which is a convolutional product of a Gaussian

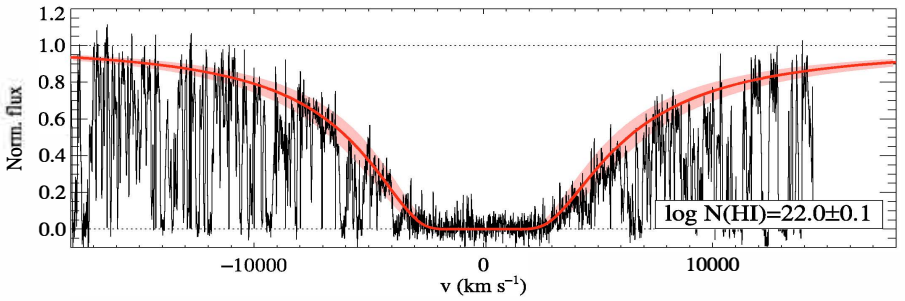

FIGURE 1.10 : An example of the Voigt profile fitting of a DLA centered at $z_{\rm DLA} = 3.286$ (J. X. PROCHASKA et HERBERT-FORT 2004).

profile and a Lorenztian profile, corresponding to the thermal Doppler effect and the collisional cross-section of Neutral Hydrogen atoms. Figure 1.11 shows a comparison of a Gaussian profile, a Lorentzian profile, and a Voigt profile. The Lorentzian profile characterizes the damping wings while the Gaussian profile determines the shape of the trough.

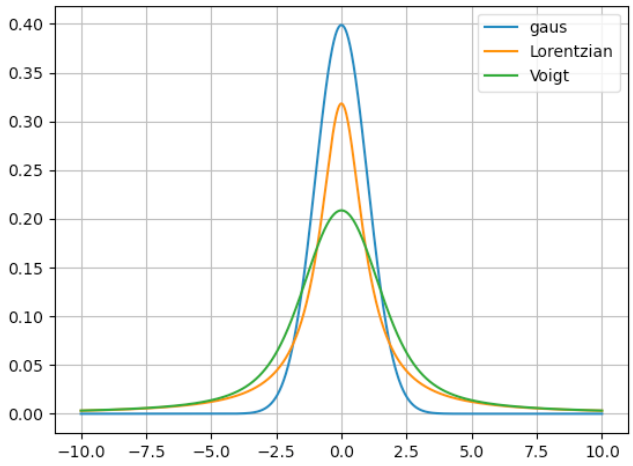

FIGURE 1.11 : A schematic diagram for a Gaussian profile, a Lorentzian profile, and a Voigt profile.

The presence of these HCDs has a similar effect on the Lyα power spectrum as the 'fingers of God' effect of galaxies (see Section 1.2.3), which both affect along the line-of-sight and on small scales. However, the 'fingers of God' effect is local and affects the redshift estimation, while the damping wings of HCDs extend out to all absorption lines in the Lyα forest along the line-of-sight. The cross-correlations between quasars or Lyα forests with these HCDs with damping wings, will then result in a suppression (at the scale of HCD widths) on the Lyα power spectrum after Fourier Transform. Figure 1.12 presents an illustration of Voigt profiles with different HI column densities and their Fourier Transform.

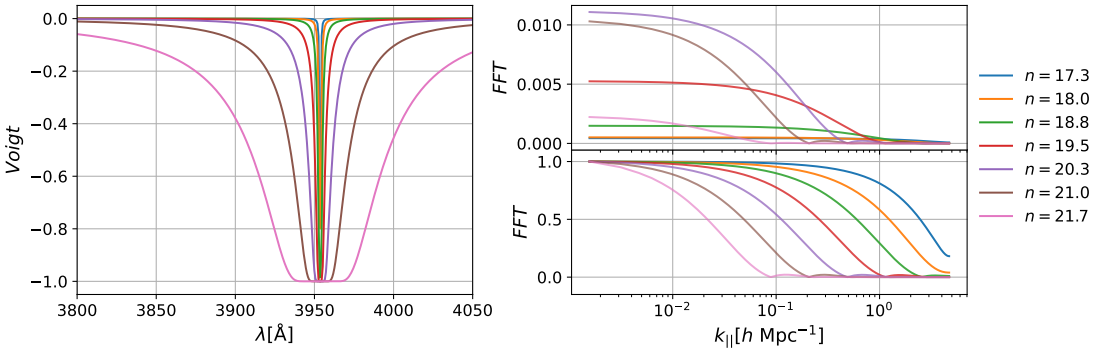

FIGURE 1.12 : Examples for Voigt profiles with different HI column densities, and their associated Fourier Transform : unnormalized (upper) and normalized (bottom).

Bibliographie du présent chapitre

Hubble, E. (1929). "A relation between distance and radial velocity among extra-galactic nebulae". In : *Proceedings of the national academy of sciences* 15.3, p. 168-173.

Robertson, H. P. (1936). "Kinematics and World-Structure III." In : *Astrophysical Journal, vol. 83, p. 257* 83, p. 257.

Walker, A. G. (1937). "On Milne's theory of world-structure". In : *Proceedings of the London Mathematical Society* 2.1, p. 90-127.

Bondi, H. et T. Gold (1948). "The steady-state theory of the expanding universe". In : *Monthly Notices of the Royal Astronomical Society* 108.3, p. 252-270.

Matthews, T. A. et A. R. Sandage (1963). "Optical Identification of 3C 48, 3C 196, and 3C 286 with Stellar Objects." In : *Astrophysical Journal, vol. 138, p. 30* 138, p. 30.

Penzias, A. A. et R. W. Wilson (1965). "Measurement of the Flux Density of CAS a at 4080 Mc/s." In : *The Astrophysical Journal* 142, p. 1149.

Jackson, J. (1972). "A critique of Rees's theory of primordial gravitational radiation". In : *Monthly Notices of the Royal Astronomical Society* 156.1, 1P-5P.

Donnelly, T., S. Freedman, R. Lytel, R. Peccei et M. Schwartz (1978). "Do axions exist ?" In : *Physical Review D* 18.5, p. 1607.

Greenstein, J. L. et M. Schmidt (1979). "The quasi-stellar radio sources 3c 48 and 3c 273". In : *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, p. 811-818.

Penzias, A. A. et R. W. Wilson (1979). "A measurement of excess antenna temperature at 4080 MHz". In : *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, p. 873-876.

Bethe, H., G. Brown et I. Worpole (1985). "How a Supernova Explodes". In.

Kaiser, N. (1987). "Clustering in real space and in redshift space". In : *Monthly Notices of the Royal Astronomical Society* 227.1, p. 1-21.

Hoyle, F., G. Burbidge et J. V. Narlikar (1993). "A quasi-steady state cosmological model with creation of matter". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 410, no. 2, p. 437-457.* 410, p. 437-457.

Jungman, G., M. Kamionkowski et K. Griest (1996). "Supersymmetric dark matter". In : *Physics Reports* 267.5-6, p. 195-373.

Hamilton, A. (1998). "Linear redshift distortions : A Review". In : *The Evolving Universe : Selected Topics on Large-Scale Structure and on the Properties of Galaxies*, p. 185-275.

Perlmutter, S. et al. (1999). "Measurements of Ω and Λ from 42 high-redshift supernovae". In : *The Astrophysical Journal* 517.2, p. 565.

Riess, A. G. et al. (2000). "Tests of the accelerating universe with near-infrared observations of a high-redshift type Ia supernova". In : *The Astrophysical Journal* 536.1, p. 62.

Hamilton, A. (2001). "Formulae for growth factors in expanding universes containing matter and a cosmological constant". In : *Monthly Notices of the Royal Astronomical Society* 322.2, p. 419-425.

Prochaska, J. X. et S. Herbert-Fort (2004). "The sloan digital sky survey damped Lyα survey : data release 1". In : *Publications of the Astronomical Society of the Pacific* 116.821, p. 622.

Yoon, S.-C. et N. Langer (2004). "Presupernova evolution of accreting white dwarfs with rotation". In : *Astronomy & Astrophysics* 419.2, p. 623-644.

Eisenstein, D. J., H.-j. Seo, E. Sirko et D. N. Spergel (2007). "Improving cosmological distance measurements by reconstruction of the baryon acoustic peak". In : *The Astrophysical Journal* 664.2, p. 675.

Mazzali, P. A., F. K. Ropke, S. Benetti et W. Hillebrandt (2007). "A common explosion mechanism for type Ia supernovae". In : *Science* 315.5813, p. 825-828.

Peter, P. et J.-P. Uzan (2009). *Primordial cosmology*. Oxford University Press.

Kim, J. E. et G. Carosi (2010). "Axions and the strong C P problem". In : *Reviews of Modern Physics* 82.1, p. 557.

Zarrouk, P. (2018a). "Clustering Analysis in Configuration Space and Cosmological Implications of the SDSS-IV eBOSS Quasar Sample". Thèse de doct. Université Paris-Saclay (ComUE).

— (2018b). "Clustering Analysis in Configuration Space and Cosmological Implications of the SDSS-IV eBOSS Quasar Sample". Thèse de doct., p. 17.

Chabanier, S., N. Palanque-Delabrouille et al. (2019). "The one-dimensional power spectrum from the SDSS DR14 Lyα forests". In : *Journal of Cosmology and Astroparticle Physics* 2019.07, p. 017.

De Sainte Agathe, V. (2019). "Mesure de la position du pic d'oscillations acoustiques baryoniques dans les forêts Lyα et Lyβ des spectres des quasars du relevé eBOSS-SDSS IV". Thèse de doct. Sorbonne université.

Aghanim, N. et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In : *Astronomy & Astrophysics* 641, A6.

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Dodelson, S. et F. Schmidt (2020). *Modern cosmology*. Academic press.

Di Valentino, E. et al. (2021). "In the realm of the Hubble tension—a review of solutions". In : *Classical and Quantum Gravity* 38.15, p. 153001.

Mörtsell, E., A. Goobar, J. Johansson et S. Dhawan (2022). "The Hubble tension revisited : additional local distance ladder uncertainties". In : *The Astrophysical Journal* 935.1, p. 58.

Ravoux, C. (2022). "One-and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman-α forest samples". Thèse de doct. Université Paris-Saclay.

Stermer, J. (2022). "Utilisation de catalogues simulés pour les analyses BAO Lyman-alpha du relevé eBOSS". In.

Chapitre 2

The spectroscopic surveys eBOSS and DESI

For the past centuries, astrophysical observations were carried out by traditional optical telescopes and manual data collection. However, with the advancements in optoelectronic systems over the past few decades, galaxy surveys conducted using large telescopes have emerged as the primary approach for observational cosmology research.

This thesis benefits from spectroscopic observation data from two large cosmological surveys, eBOSS and DESI. The survey validation program of DESI was started almost at the same time as this thesis, thus enabling me to get involved in the data quality checking, target selection pipeline test, collection of the main survey data, and scientific analysis of DESI.

In this chapter, I describe the scientific goals, survey designs, target selection procedures, instrument setup, and observation enrollments of these two surveys.

2.1 SDSS

The realization of large cosmology surveys were facilitated with the development of high precision optical instruments (e.g., Charged Coupled Devices CCDs and multi-object fiber spectroscopy) and computational resources, thus enabling the exploration of cosmology into a new phase. The first large cosmology survey, the Sloan Digital Sky Survey (SDSS), was motivated and proposed by York et al. 2000 in the 1980s. Organized by a large astronomy collaboration, it aimed at mapping the universe from the local group to the largest clustering of galaxies. After a decade of preparation and construction, SDSS took its first testing light in 1998, and started observations in 2000. It used a 2.5m Ritchey-Chrétien telescope installed at Apache Point Observatory (APO) in New Mexico, USA.

The scientific goal of SDSS was to get a better understanding of the large scale structure of the Universe and the growth of structure. Therefore, measuring the distribution of matter and studying the nature of dark energy using BAO became the core approaches.

2.1.1 Survey design

SDSS launched different phases of both photometric and spectroscopic surveys every few years. In this section I will describe these different generations of surveys.

SDSS-I, II, and III

Equipped with two spectrographs connected to an imaging camera and a fiber plate respectively, the first observation program SDSS-I ($2000 - 2005$) provided a photometric survey with ~ 8000 deg^2 footprint and a spectroscopic survey with ~ 5700 deg^2. The famous detection of BAO using luminous red galaxies (LRGs) (D. J. Eisenstein, Zehavi et al. 2005), was announced during SDSS-II ($2005 - 2008$), which provided a photometric survey with ~ 11500 deg^2 footprint and a spectroscopic survey with ~ 7500 deg^2.

Following by SDSS-II, the third observation program SDSS-III ($2008 - 2014$) was carried out with upgraded spectrographs, and four different surveys were completed for different scientific goals :

- The Baryon Oscillation Spectroscopic Survey (BOSS), a six-year spectroscopic survey dedicated to the measurement of BAO using extragalactic targets. The entire survey collected ~ 1.5 million LRGs (z between 0.2 and 0.75) and $150,000$ high-reshift quasars (QSOs with $z > 2$), covering ~ 10000 deg^2. Thanks to those high-redshift QSOs, a sufficient number of Lyα forests were observed from the QSOs' spectra, and the BAO peak was detected from both the Lyα auto-correlation function (J. E. Bautista et al. 2017) and the Lyα-QSO cross-correlation function (Delubac, Rich et al. 2013).

- The Sloan Extension for Galactic Understanding and Exploration 2 (SEGUE-2), designed to observe $\sim 120,000$ stars in our galaxy halo, and better understand the sub-structures of the Milky Way.

- The Apache Point Observatory Galactic Evolution Experiment 1 (APOGEE-1), which measured $\sim 100,000$ red giant stars with precise peculiar velocities in our galaxy.

- The Multi-object APO Radial Velocity Exoplanet Large-area Survey (MARVELS), a spectroscopic survey planned to observe $\sim 11,000$ bright stars in our galaxy, and study the formation and evolution of planet systems with high-precision.

SDSS-IV and eBOSS

Having achieved fruitful and remarkable scientific results based on three successive generations of surveys, the fourth program of SDSS was proposed to further investigate more near-field and extragalactic targets. SDSS-IV (2014−2019)[1] (Blanton et al. 2017) provided the latest available data release DR16 (Ahumada et al. 2020) and DR17 (Abdurro'uf et al. 2022), and is composed of three surveys :

- The APO Galactic Evolution Experiment 2 (APOGEE-2), aimed at exploring the history and evolution of the Milky Way by targeting an enormous number of nearby stars.
- The Mapping Nearby Galaxies at APO (MaNGA), observed stars and gas in $10,000$ nearby galaxies.
- The extended Baryon Oscillation Spectroscopic Survey (eBOSS) (K. S. Dawson et al. 2016), was designed as an extension of BOSS to detect the BAO with percent-level precision in the redshift range $0.6 < z < 3.5$, by collecting large catalogs of different types of extragalactic galaxies and QSOs. I will further give more details about the eBOSS survey in the next section (see Section 2.1.3).

2.1.2 The instrument

The 2.5m Ritchey-Chrétien telescope was used for the data collection of both BOSS and eBOSS surveys, with both photometric and spectroscopic programs. The instrument was therefore designed for both modes : the photometric mode with an imaging camera composed by an array of 6 columns of 5 CCDs equipped with different filters u, g, r, i, z, with central wavelengths of $3590, 4810, 6230, 7640, 9060$ Å (Gunn, M. Carr et al. 1998) ; the spectroscopic mode used 1000 fibers (fiber diameter of 120 μm) placed on a fiber cartridge on the focal plane. The fiber cartridge or the camera can be connected to 2 identical spectrographs (Smee et al. 2013). Figure 2.1 gives a graphic of this structure. This gives at the end a field of view (FOV) of 7 deg^2.

2.1.3 eBOSS

In this section, I will describe in detail the eBOSS survey, which provided the largest sample of high-redshift extragalactic targets among the different cosmological surveys carried out in SDSS. In this survey, four samples of targets were observed (see Figure 2.3), including a large sample of high-redshift LRGs, a new sample of emission line galaxies (ELGs, blue star-forming galaxies) in the redshift range $0.6 < z < 1.2$, a low-redshift QSOs sample (hereafter tracer QSOs, with $z < 2$), and a high-redshift QSOs sample (hereafter Lyα QSOs). A visualization of the eBOSS QSO footprint is shown in Figure 2.2, which includes all the QSO targets in SDSS-I, II, BOSS, and eBOSS. Using these data, the first detection of the BAO peak using tracer QSOs ($z < 2$) (Ata et al. 2018) was made possible, thus enabling percent-level measurements of the BAO peak position over a wide redshift range (see Figure 2.4). These measurements provided an unprecedentedly strong constraint on the ΛCDM model and its extended theories (Alam et al. 2021). Figure 2.4 shows the ratio of the BAO measurement compared to the Planck 2018 result using CMB with the ΛCDM model (Aghanim et al. 2020). One can also tell from Figure 2.4 that Lyα forests give the highest redshift constraints of BAO, which is the goal of my PhD study, i.e., using the eBOSS/DESI data to get higher redshift BAO measurement.

[1] `https://www.sdss.org/`

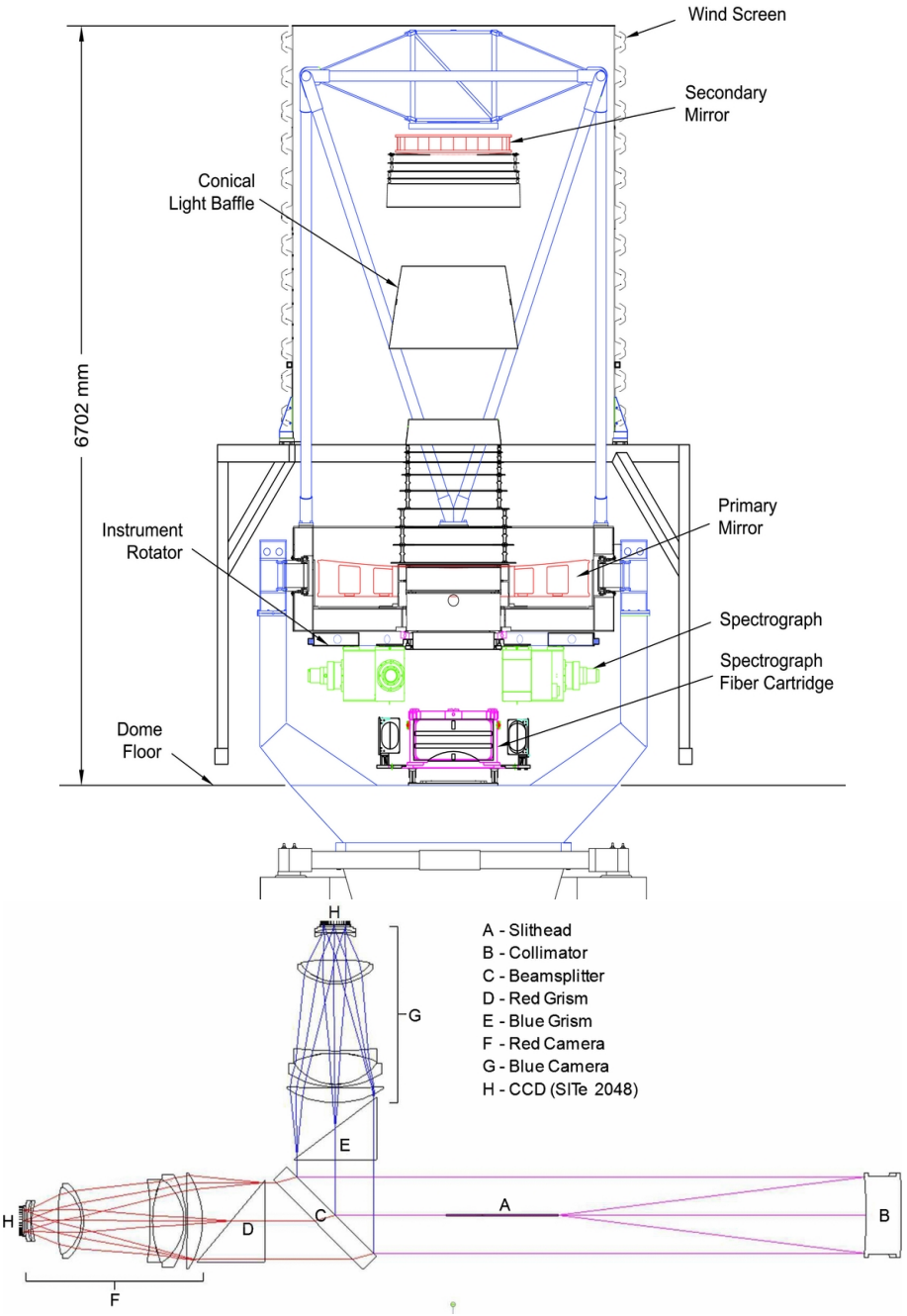

FIGURE 2.1 : The upper plot presents a description of the SDSS/eBOSS instrument. In spectroscopic mode, the fiber cartridge was connected to the 2 spectrographs. The lower plot shows the schematic structure of the spectrographs used in eBOSS. The input light will be transformed into a beam by a collimator, and the splitted beam will be collected from the red and the blue cameras, each of which contains 8 lenses and a CCD with 4096×4096 pixels.

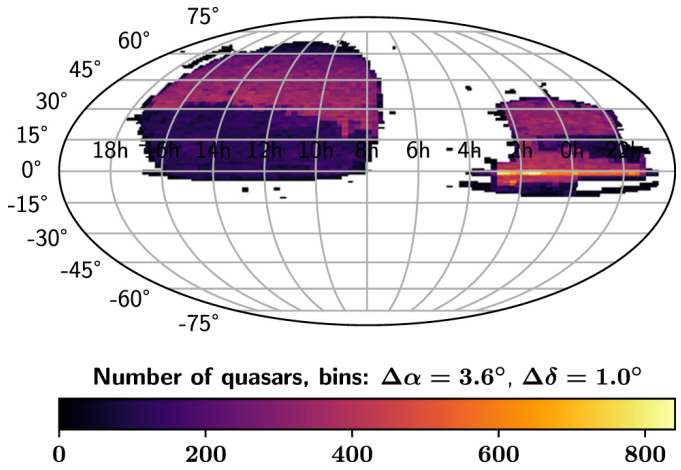

FIGURE 2.2 : Footprint of eBOSS DR16 QSOs (LYKE et al. 2020) (including all the QSO targets from SDSS-I, II, BOSS, and eBOSS).

eBOSS Target selection

eBOSS made a target selection of 4 classes of extragalactic galaxies and quasars. Figure 2.3 presents the combined BOSS and eBOSS targets :

- Luminous red galaxies (LRGs), $377,458$ spectra in the redshift range $0.6 < z < 1.0$ (J. BAUTISTA 2020).
- Emission line galaxies (ELGs), $69,243$ in the redshift range $0.6 < z < 1.2$ (RAICHOOR, COMPARAT et al. 2017).
- Tracer Quasars (QSOs), in the redshift range $0.9 < z < 2$ (ALAM et al. 2021).
- Lyα Quasars (QSOs), 0.7 million Lyα quasars in the redshift range $2 < z$. The final DR16 (DES BOURBOUX, RICH et al. 2020) catalog collects $200,000$ Lyα QSOs.

Quasar target selection

My PhD research mostly makes use of the high-redshift Lyα QSOs catalog in DR16 (DES BOURBOUX, RICH et al. 2020). I hereby describe the target selection and classification pipeline used in eBOSS.

The classification of targets was performed firstly using the photometric data of each target. At low redshifts, stars and quasars can be distinguished (RICHARDS et al. 2002) by looking at their measured magnitudes in the four-dimensional color band (u, g, r, i). Moreover, the spectra of quasars can be easily distinguished from those of stars, that roughly follow a black body spectrum. However, at higher redshifts $z \sim 2.8$, the distribution of quasars and stars overlap in the $u - g$ color band regions, making it difficult to distinguish these classes efficiently. The XDQSOz method (CROOM et al. 2009 ; BOVY et al. 2011) was used to improve this selection process, which is a probabilistic target-selection technique that uses density estimation of quasars in flux space. Moreover, additional information for the mid-infrared color band was added using the data from WISE (Wide-field Infrared Survey Explorer) (WRIGHT et al. 2010), to help the

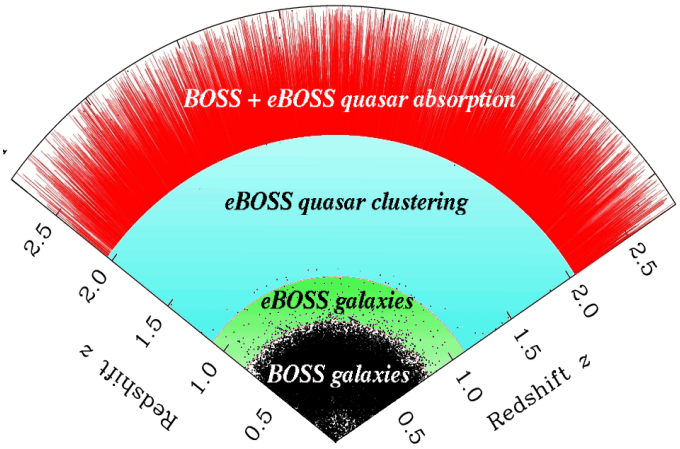

FIGURE 2.3 : Clustering of targets collected in BOSS and eBOSS, over a wide range of redshift. Credits : `https://www.sdss.org/surveys/eboss/`.

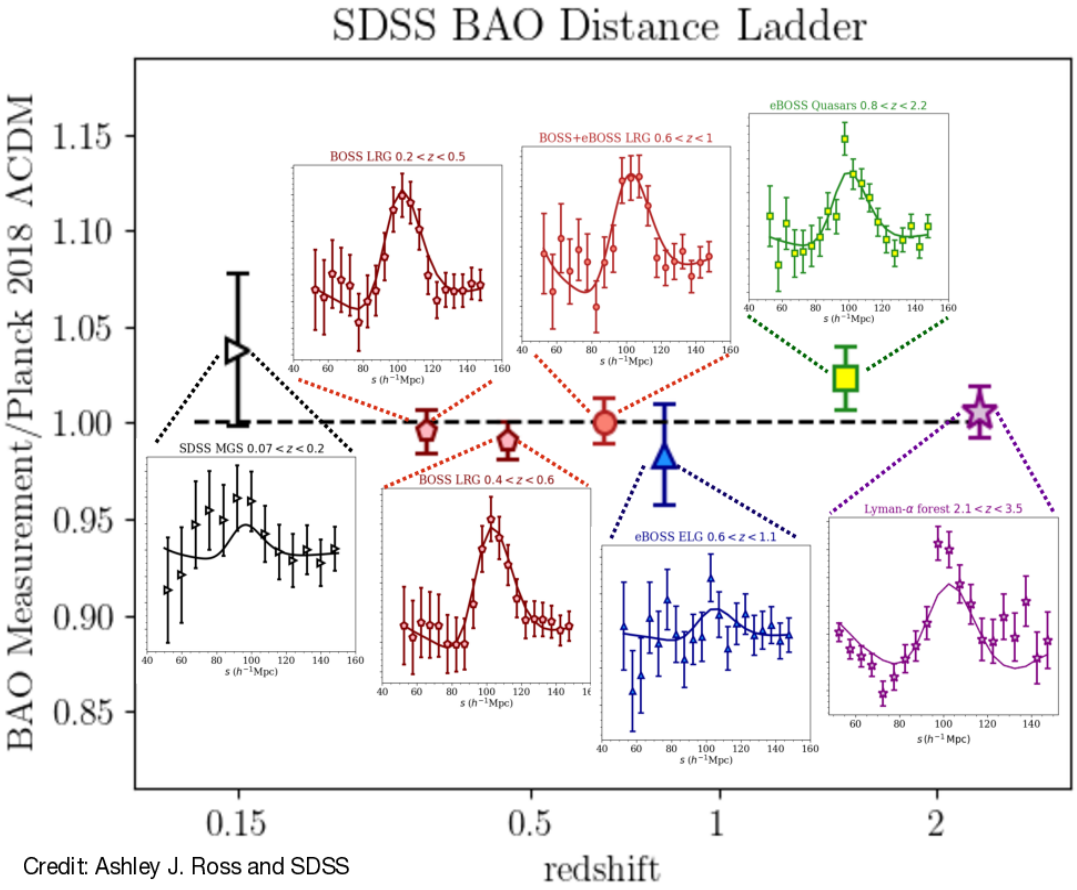

FIGURE 2.4 : BAO measurement from SDSS, BOSS, and eBOSS using different tracers of galaxies and quasars, over a wide range of redshift. Credit : Ashley J.Ross and SDSS.

selection of QSOs (the complete QSOs selection strategy described in A. D. Myers, Palanque-Delabrouille et al. 2015).

The determination of QSOs redshifts is essential for further two-point analysis, where the redshift error will contribute as a non-negligible systematic effect for the detection of the BAO peak (Youles et al. 2022). The eBOSS pipeline will do the classification and redshift estimation by fitting each target spectrum with templates and Principal Component Analysis (PCA). Then a least-squares minimization is performed by comparing each spectrum to all the templates, and the final classification and redshift are made by determining the lowest χ^2 (A. S. Bolton, D. J. Schlegel et al. 2012).

A Visual Inspection (VI) program was conducted to ensure the purity of the quasar sample and the accuracy of the estimated redshift. For each spectrum to be visually inspected, a power law fit of the continuum and all the emission lines is displayed and can be displaced by inspectors to reach the maximum fitting. The MgII line is usually used for redshift estimation as it is not strongly affected by systematic shifts of quasar outflows (Shen et al. 2016). The CIV line was used if the MgII line was not available or hard to identify.

2.2 Dark Energy Spectroscopic Instrument

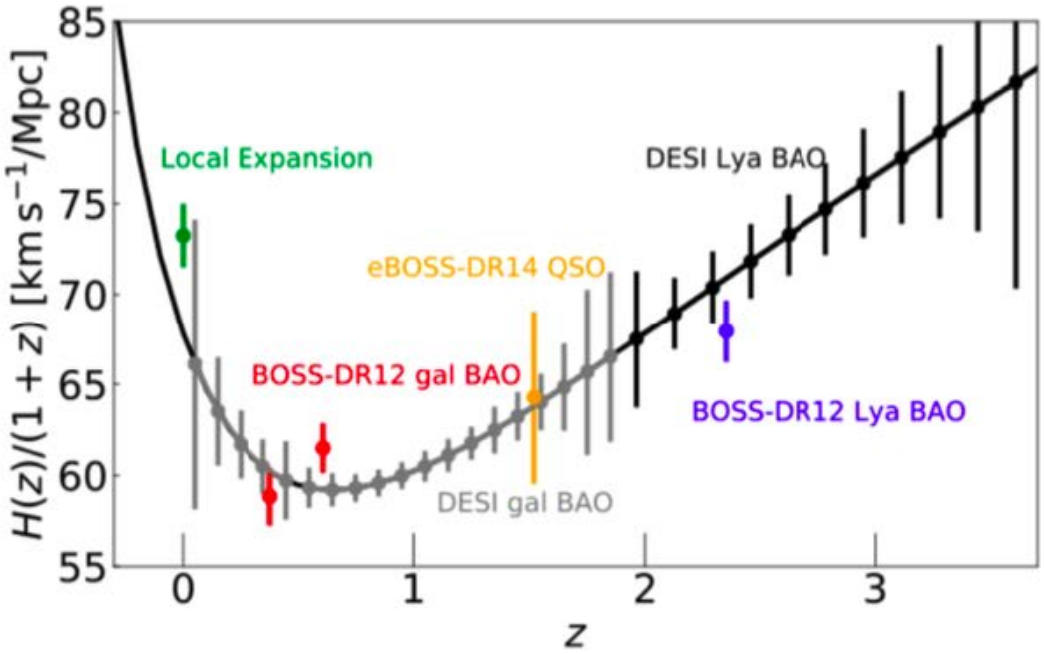

FIGURE 2.5 : DESI's forecast measurement of the Hubble diagram in function of redshift using BAO (AGHAMOUSA et al. 2016a). Credits : DESI key figures for external communication.

The eBOSS survey has made great success in verifying the ΛCMD model. As a continuation of eBOSS, the Dark Energy Spectroscopic Instrument (DESI) (M. LEVI et al. 2013), was proposed in 2012 and the construction was started in 2015. It collected its first light in October 2019, while the commissioning phase and survey validation program was postponed to late 2020 due to a 7-month shutdown during the Covid-19 pandemic. Affected by this particular pandemic, the 5-year observation commission was started in May 2021 and is entirely remote for supporting observing scientists during 2021-2022 (in-person shifts are available since 2023). DESI was deployed with the 4-meter Mayall telescope at Kitt Peak National Observatory in Arizona, USA. It will collect the largest galaxy catalog to date with more than 40 million spectra of galaxies and quasars (AGHAMOUSA et al. 2016a) from the nearby universe to beyond redshift $z > 3.5$, over 14,000 deg^2 of the sky.

The DESI survey was designed as a stage-IV cosmology survey. Its scientific goal is to further explore the nature of dark energy through the measurement of BAO with a more precise determination of the matter distribution of the local Universe, and to probe modifications of general relativity by measuring the growth of structure through Redshift Space Distortions (RSD). The enormous amount of DESI data will not only constrain the ΛCDM model with unprecedented statistical precision, but will also be useful for the understanding of neutrino masses and extension theories beyond ΛCDM. With DESI data, BAO will be used as a probe to measure the isotropic cosmic distance scale to 0.28% precision in the redshift bin $0 < z < 1.1$ and to 0.39% precision in the redshift bin $1.1 < z < 1.9$. It will also measure the Hubble parameter

at $1.9 < z < 3.7$ to 1.05%. More details about the scientific goals for each tracer are described in Abareshi et al. 2022. A forecast of DESI's measurement of the Hubble parameter using different tracers is presented in Figure 2.5. Lyα forest BAO is expected to give the constraints at the highest redshift, which is the main goal of this thesis.

The main analysis of this thesis relies mainly on the DESI Lyα quasar data. Figure 1.9 in Chapter 1 shows an example of an observed quasar spectrum in DESI. I contributed as an active member to the analysis of DESI data quality, and its comparison with the eBOSS DR16 data. Moreover, simulated mocks for both DESI/eBOSS surveys are also made to validate and test the analysis pipeline. I will further describe these Lyα analyses in Chapter 5.

2.2.1 Survey design

In order to meet its scientific purpose, DESI is designed to cover 14,000 deg^2 of the sky area and observe more than 40 million galaxies and quasars in its five-year mission time. The footprint of DESI survey is present in Figure 2.6, which is designed by the DESI Legacy Imaging Surveys (A. Dey et al. 2019). It is constructed from a combination of a few surveys, including The Mayall z-band Legacy Survey (MzLS [2]), the Beijing-Arizona Sky Survey (BASS [3]), the Dark Energy Camera Legacy Survey (DECaLS [4]) that provide optical imaging data in the g, r, and z bands, and the Wide-field Infrared Survey Explorer (WISE [5]) satellite that provides all-sky mid-infrared imaging in the 3.4 μm and 4.6 μm WISE bands. It is composed of two regions, one in the North Galactic Cap (NGC) covering 9900 deg^2 and one in the South Galactic Cap (SGC) covering 4400 deg^2. In addition to the above-mentioned surveys, the Dark Energy Survey (DES [6]) is also used as an external source to complete the SGC footprints. Figure 2.6 shows all the surveys that contribute to the DESI Legacy Imaging Surveys.

2.2.2 Target selection

The observing targets of DESI are provided by the photometric datasets of DESI Legacy Imaging Surveys. Then an integrated pipeline (A. D. Myers, Moustakas et al. 2022) is applied for the DESI target selection (TS), to obtain the desired number of targets for each class of galaxies or quasars. Those targets will be categorized into 5 classes (see Figure 2.7 for a visualization), and the TS is detailed in C. Hahn et al. 2023 for the Bright galaxy sample (BGS), Zhou et al. 2023 for Luminous red galaxies (LRGs), Raichoor, Moustakas et al. 2023 for Emission line galaxies (ELGs), and Chaussidon et al. 2023 for QSOs :

- BGSs, 10 million in the redshift range $0 < z < 0.4$.
- LRGs, 6 million in the redshift range $0.4 < z < 1.0$. The average density should be at least 300 deg^{-2}, and the redshift completeness (see definition in Equation 2.1) should be larger than 95% for each pointing averaged over all fibers that observe objects.
- ELGs, 17 million in the redshift range $0.6 < z < 1.6$. The average density should be at least 1280 deg^{-2}, and the redshift completeness should be larger than 90% for each pointing averaged over all targets above the O II flux limit.

[2] https://www.legacysurvey.org/mzls/
[3] https://www.legacysurvey.org/bass/
[4] https://www.legacysurvey.org/decamls/
[5] https://www.nasa.gov/mission_pages/WISE/main/index.html
[6] https://www.darkenergysurvey.org/

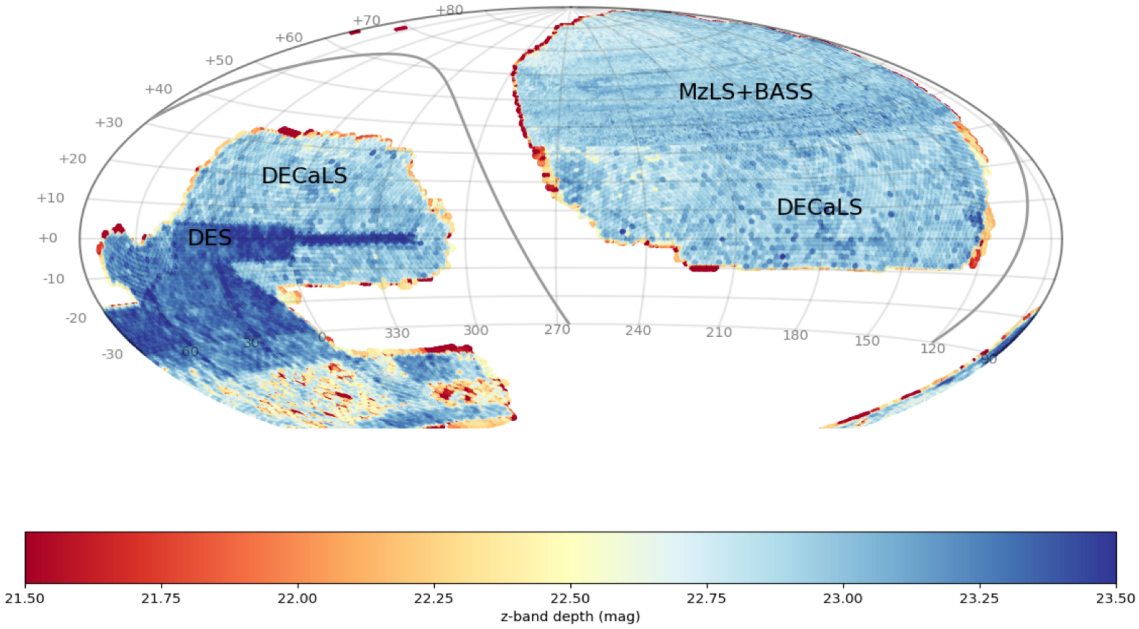

FIGURE 2.6 : Footprint of the DESI Legacy Imaging Surveys, and all the imaging surveys that it is composed of. MzLS/BASS contributes to the northern part of NGC (9900 deg^2). DECaLS contributes to the southern part of NGC and all of the SGC (4400 deg^2). DES is used as an external source to complete the SGC footprints. Credits : DESI key figures for external communication.

- Tracer Quasars (QSOs), 1.7 million tracer quasars in the redshift range $0.9 < z < 2.1$. The average density should be at least 120 deg^{-2}, and the redshift completeness should be larger than 90% for each pointing averaged over all fibers that observe objects.
- Lyα Quasars (QSOs), 0.7 million Lyα quasars in the redshift range $2.1 < z$. The average density should be at least 50 deg^{-2} in this redshift range and $r < 23.5$ mag.

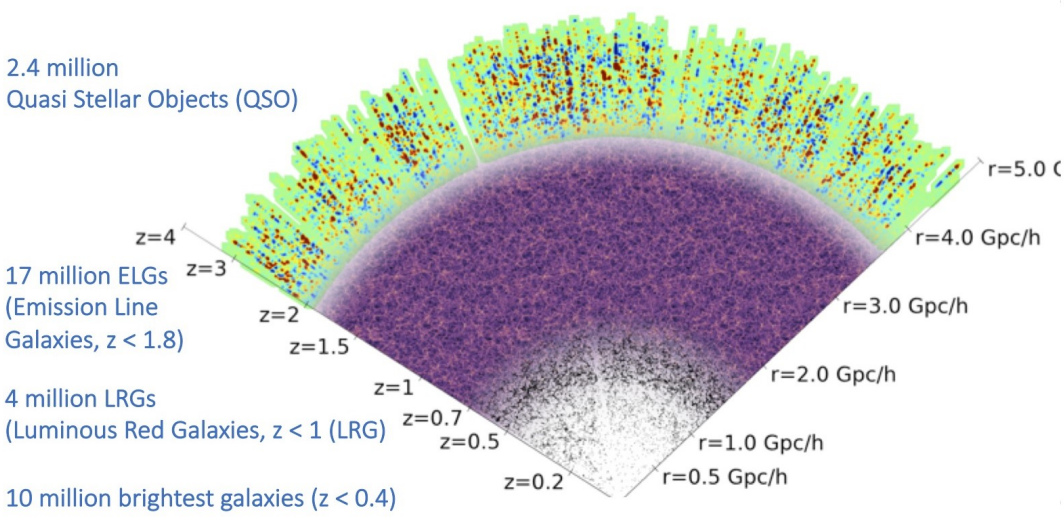

FIGURE 2.7 : Different expected target catalogs of DESI's five-year plan. Note that the expected number of LRGs has updated to 6 million, and overall, DESI is observing more targets than expected. Credits : DESI key figures for external communication.

The QSO selection procedure is mainly composed of two steps :

- The TS pipeline : selecting QSOs apart from stars based on photometric data of g, r, z optical bands and infrared bands W. Figure 2.8 shows this selection where red points are classified as stars, and points from blue to yellow are classified as QSOs at different redshifts. A Random Forests (RF) machine learning algorithm (YÈCHE et al. 2020 ; CHAUSSIDON et al. 2023) is used to ensure the efficiency of this selection.
- The classification pipeline : after the TS, a classification pipeline is applied to the QSO candidates to construct a true QSO catalog. It is a combination of three algorithms used to further classify QSOs and provide precise redshift estimations :
  - The DESI pipeline classifier Redrock (RR (BAILEY in preparation)), which is a template-based fitting classifier. It applies a PCA decomposition for the target spectrum, and determines the lowest χ^2 compared to all the template spectra. In the end, the best-fitted target class (star, galaxy, or QSO) and an estimated redshift are provided. This algorithm gives an accurate estimation of the redshift, while the completeness (see definition in Equation 2.1) is not sufficient. Therefore, we use it at first to pre-select the QSO targets to be retained in the final catalogs.
  - A broad Mg II line finder (Mg II), which uses the Mg II broad line for the classification, since some quasars will have broad Mg II line and thus be falsely missed by RR. It uses

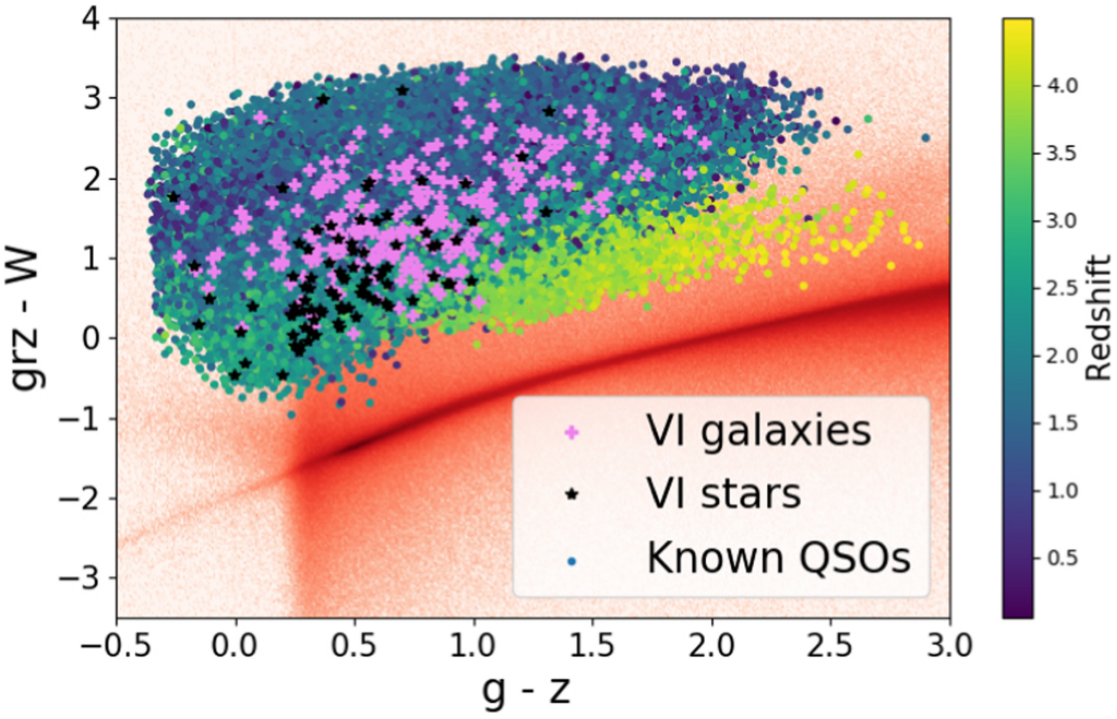

FIGURE 2.8 : QSO TS based on photometric data of g, r, z optical bands and infrared bands W. Red points are classified as stars, while points from blue to yellow are classified as QSOs at different redshifts. Credits : CHAUSSIDON et al. 2023.

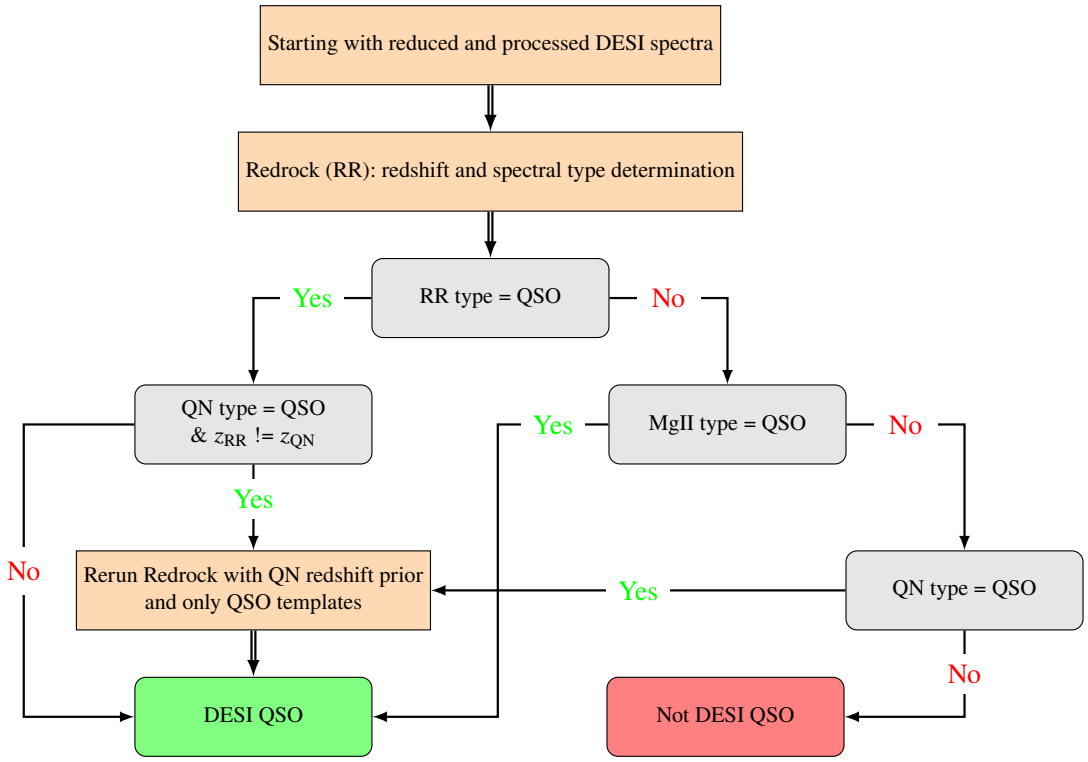

FIGURE 2.9 : Workflow of the QSO classification pipeline to create a QSO catalog.

Label	Ground False 0	Ground Truth 1
Prediction 0	True Negative (TN)	False Negative (FN)
Prediction 1	False Positive (FP)	True Positive (TP)

TABLEAU 2.1 : Confusion Matrix for quasar classification.

the RR output as an input and fits a Gaussian centered at the Mg II line determined by RR. The Mg II line will be considered as a broad line if the χ^2 is improved by 16, then the target will be considered as a QSO. This method does not modify the estimated redshift by RR but is used as an afterburner to discover the QSOs missed by RR.

– A machine learning classifier QuasarNET (BALLAND et al. 2018 ; FARR, FONT-RIBERA et PONTZEN 2020) based on a convolutional neural network (CNN) (see Figure 2.10). This algorithm searches for six emission lines : Ly α, C IV, C II, Mg II, H α, H β. A target will be classified as a QSO if one of these six emission lines' confidence level is larger than 0.5 (see Figure 2.11).

A true QSO catalog collected by visual inspection (D. M. ALEXANDER et al. 2023) is used to validate the classification algorithms. A confusion matrix is built to summarize the classification by comparing the predicted results with the true catalog, as shown in Table 2.1. The overall classification efficiency is then defined by purity and completeness as

$$\begin{aligned} \text{Purity} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Completeness} &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \tag{2.1}$$

Here TP, FP, FN are defined in Table 2.1. A workflow of these three algorithms is designed to maximize the classification purity and completeness, which is visualized in Figure 2.9. With this strategy, a target is considered a QSO once classified by one of these algorithms in the order RR→Mg II→QuasarNET. In the end, an overall 99.3% purity and 94.0% efficiency are achieved (see Figure 2.12).

During my PhD, I contributed to the implementation of the QSO classification pipeline into the DESI working environment and performed the first test of the QSO catalog. These contributions are summarised in CHAUSSIDON et al. 2023.

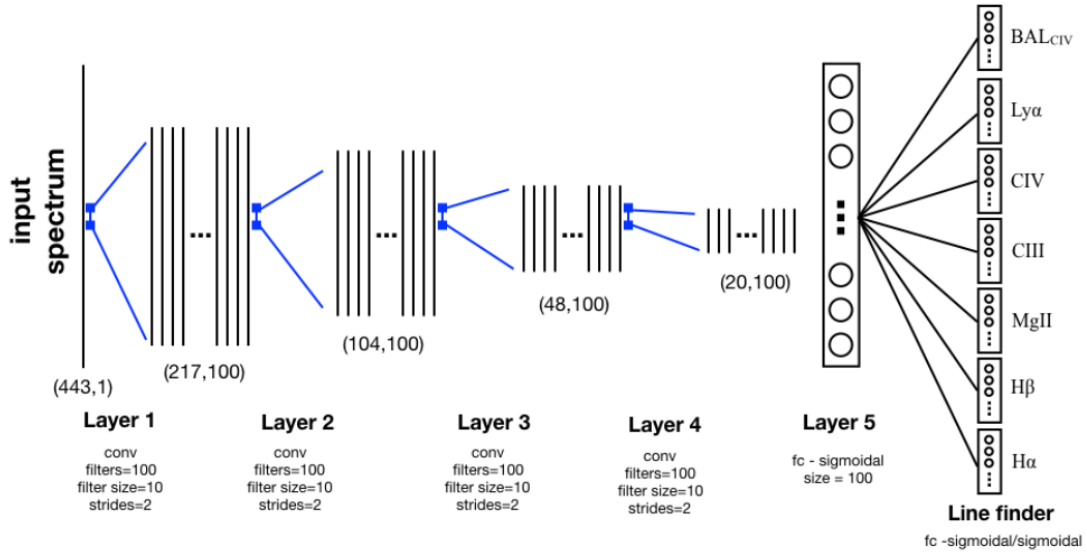

FIGURE 2.10 : The CNN structure of QuasarNET, composed of 4 convolutional layers and a connected layer. The input spectrum is down-sampled to 443 pixels, while the final classification relies on a multi-task classification for 6 emission lines.

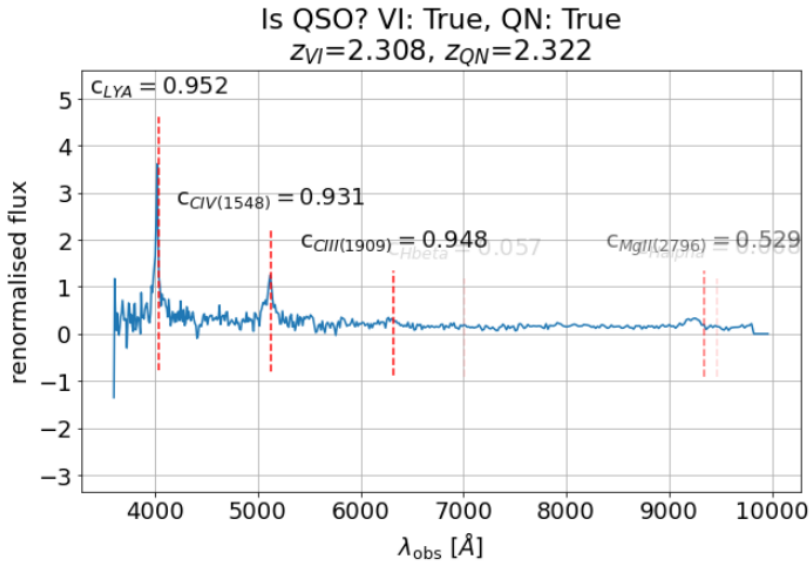

FIGURE 2.11 : A QSO classified by QuasarNET. The four emission lines with confidence level larger than 0.5 are : Ly α, C IV, C II, Mg II.

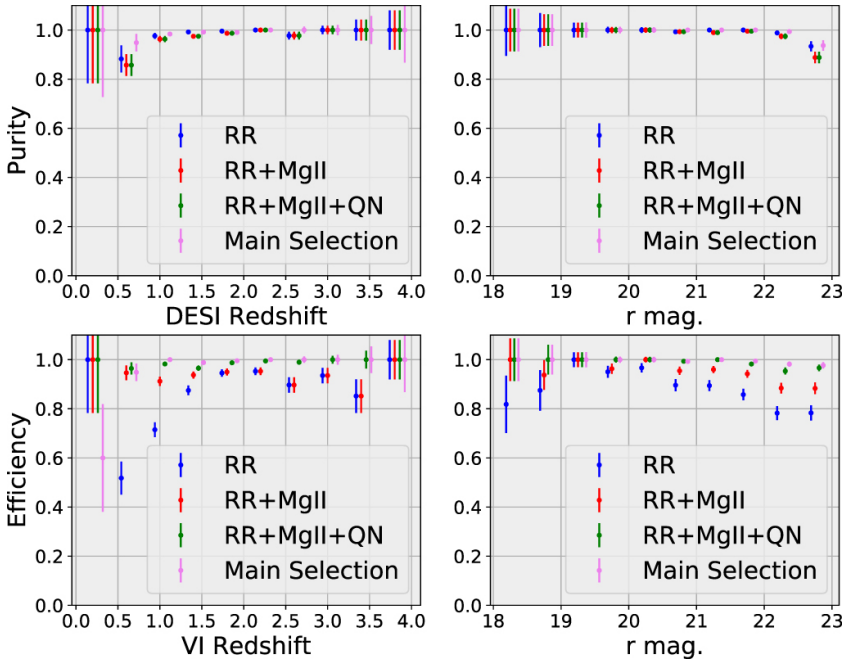

FIGURE 2.12 : Efficiency and purity of the DESI QSO classification, using the visual inspection data, with a combination of three algorithms (RR, Mg II and QuasarNET). The performance is presented as a function of redshift and r magnitude. The main selection is performed following the procedure shown in Figure 2.9.

2.2.3 The instrument

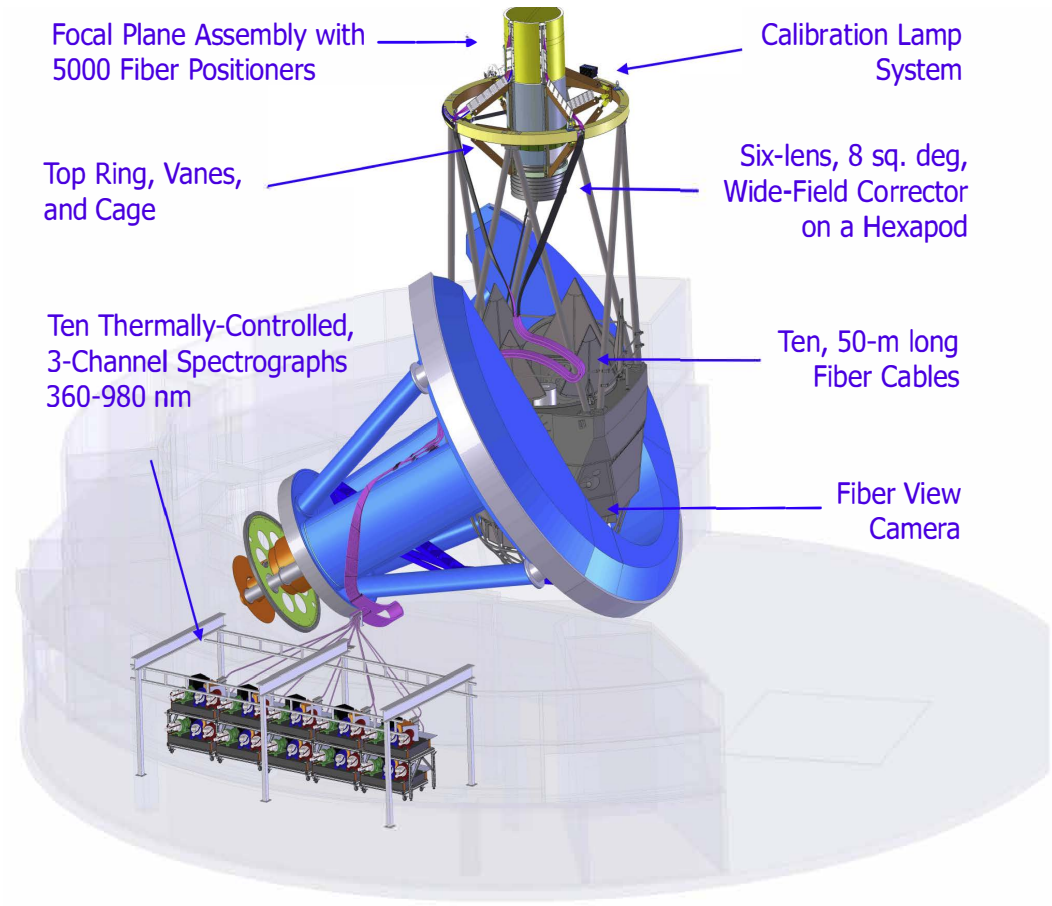

FIGURE 2.13 : Main structure of the DESI 4-meter Mayall telescope at Kitt Peak in Arizona, USA (ABARESHI et al. 2022).

DESI uses a multi-object system (AGHAMOUSA et al. 2016b ; ABARESHI et al. 2022) composed of a corrector, a focal plane with 5000 fiber robots, ten 3-arm spectrographs in the Coudé room (a special room used for instrument setups and hosting the thermal enclosure, to ensure the necessary thermal and humidity stability) and ten fiber cables that connect the focal plane with the spectrographs. The high stability and performance of these instruments guarantee the efficiency and precision of the survey. Figure 2.13 shows the main structure of the instrument.

The corrector

The corrector (SHOLL et al. 2012 ; T. MILLER et al. 2022, in prep ; T. N. MILLER et al. 2018) system is designed to maintain the optical alignment for all the lenses (to make sure that the combination of lenses works correctly to collect light). It has six lenses each one with 1-meter in diameter, that in total provide the telescope focal ratio around $f/2.8 - f/3.9$ (a unitless number defined by dividing the focal length of the telescope by the aperture), and a 3.2° field of view. Four of the lenses are fused silica and two of them have one aspheric surface, as shown in Figure 2.14.

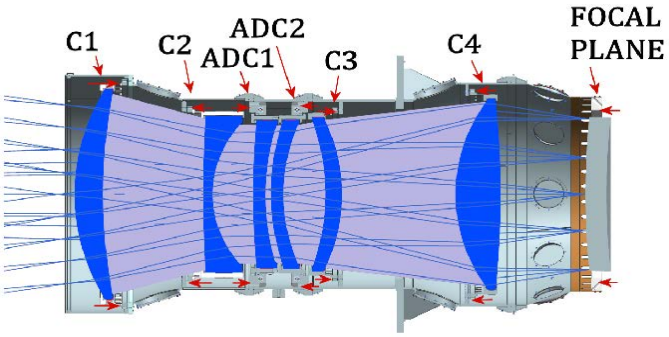

FIGURE 2.14 : The corrector for DESI (ABARESHI et al. 2022), that is composed of six lenses, four of which are fused silica (C1, C2, C3, C4), and two of them have one aspheric surface (ADC1, ADC2).

The two aspheric lenses compose the atmospheric dispersion compensator (ADC), which is used to correct the spectral spread of light when passing by the atmosphere. This ADC system then needs to be adjusted according to the actual atmosphere situation.

Focal plane

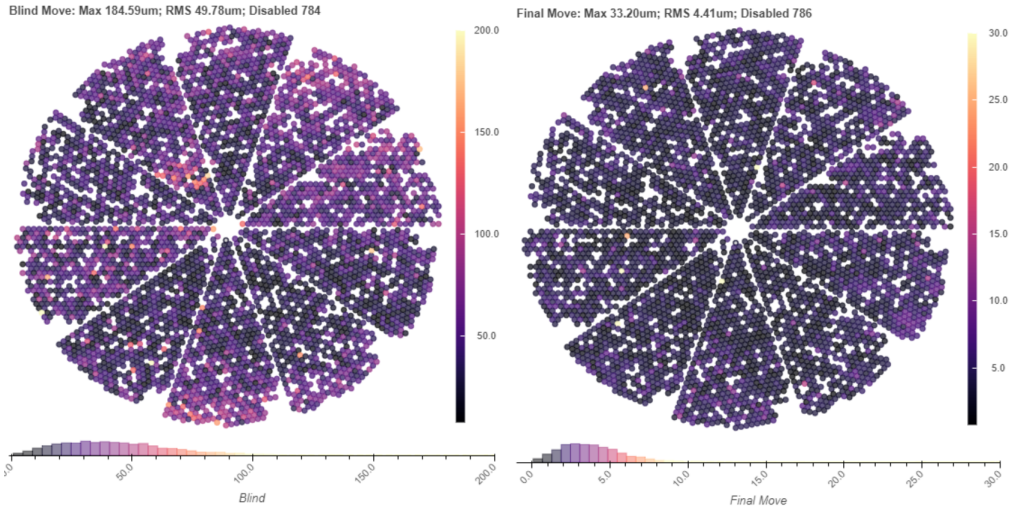

FIGURE 2.15 : The two-step positioning move : the first blind move (left) and a second corrected move (right) with the help of the FVC imaging. Histograms of all the positioner accuracies for both cases are also shown at the bottom. One can tell from the plots that the RMS (root mean square) of positioner accuracies drops from ~ 50 μm (blind move) to < 5 μm (final move).

The DESI focal plane is composed of 5000 fiber robots, segmented into 10 petals. Each petal is connected with a single spectrograph, two additional fibers that connect to the sky monitor system, and a Guide/Focus/Alignment (GFA) detector system (six of them are used for the guidance of reference stars, and four of them are used to maintain the optical alignment between

the prime focus corrector and the primary mirror (Silber et al. 2022)). Figure 2.16 shows an overview of the fibers and their positioners. A fiber view camera (FVC) placed in an opening at the center of the mirror is used to obtain an accurate positioning of the fibers. The positioning consists of two steps : a first blind move and a second move corrected with the FVC imaging. This operation increases the accuracy of the positioners at their desired locations from around 50μm rms to less than 10μm rms (see Figure 2.15), which is reasonably small compared to the 107μm diameter of the fibers. As a result of this two-step positioning procedure, the loss of collected light from each fiber is much reduced.

Spectrographs

Each petal in the focal plane is connected with one 3-arm spectrograph (Poppett et al. 2022, in prep ; Jelinsky et al. 2022, in prep) in the Coudé room, covering different optical wavelength ranges for blue ($3600-5930$Å), red ($5600-7720$Å), and NIR ($7470-9800$Å). Each spectrograph channel utilizes a cryostat that hosts a 4096×4096 CCD, with 15 μm pixels, four readout channels, and a readout rate of 100 kHz. This yields a resolution $\lambda/\Delta\lambda$ of 2000-3000 for the blue spectrograph, 3500-4500 for the red, and 4000-5500 for the near-infrared, respectively.

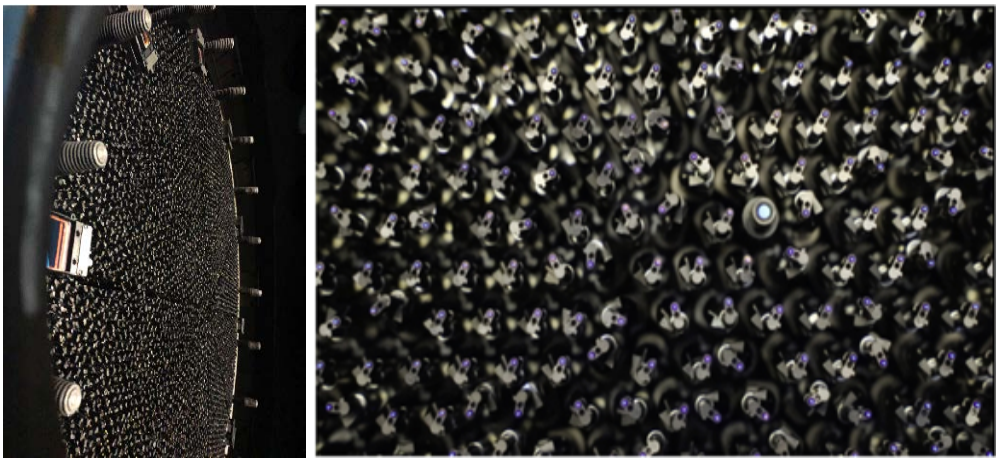

Figure 2.16 : The left plot shows an overview of the focal plane and the right plot shows the positioners (Abareshi et al. 2022).

2.2.4 Observing with DESI

During my PhD, I contributed as a Support Observing Scientist (SO) remotely for more than one month. In this section, I will describe the observation with DESI with the help of a complex monitoring system, that is composed of the DESI Instrument Control system (ICS), Data System, and a collaborated observation schedule. All of these systems are highly automated and all the associated parameters are adjusted according to observing conditions. For example, the exposure time and the region (tile) to observe are optimized depending on the observing conditions, which makes the observation highly efficient.

System description

The Sky Monitor System connected with the focal plane petals through fibers, is composed of a photometry camera near the Coudé room. These fibers are attached to specific positioners that are pointed at the blank sky in order to measure the brightness of the sky through an exposure. The expected Signal to Noise can be calculated together with the point-spread function (PSF) of guide stars, measured by the GFA. Comparing the expected exposure with the effective exposure time measured from the spectrograph, we are able to validate the status of the telescope, and adjust in time in case of any problem, such as weather issues or instrumental errors.

The DESI ICS contains all the monitoring infrastructures for instrument operations, data acquisition, and system maintenance, including the DESI Online System (DOS), and the Observation Control System (OCS). The DOS is a software built on Pyro2 (Harpole, Zingale, Hawke et Chegini 2019), that contains all the user interfaces for controlling the dashboard, telemetry, and image previews. The OCS is used to control the acquisition, flow, and storage of data.

The duty of the DESI Data System is to monitor consistently the pipeline of target selection, transfer, archiving, and distribution of data. The targets are selected according to the DESI Legacy Imaging surveys, and the correct targets are assigned to fibers with the help of the Next Field/Tile Selection (NFS/NTS). Calibrations, spectra extraction and sky subtraction are executed one after another after the completion of target selection.

Weather monitoring

Every night before observation, it is essential to check the weather forecast and monitor its status during the night. The required attributes are as follows :

- Humidity. Observers should track the trend of humidity and make sure that it is still within safe margins ($< 90\%$).
- Wind. Observers should track the trend, direction of the wind, and make sure that it is still within safe margins (< 45 mph).
- Cloud cover. Observers should check the status of the cloud cover, and if there are any holes for observing.
- Upcoming weather/Clouds.

A survey simulation was performed using 10-years of Mayall weather history in order to ensure the weather conditions. The real-time and forecast of the weather can be found on KPNO weather page [7]. Figure 2.17 gives an example of a screenshot from this weather monitoring website.

Observation scheduling

Daily observation plans are scheduled for every night and are announced at the Afternoon meeting (a regular meeting organized at 17h pm everyday to make daily observing plans). There will be one Observing Associate (OA), one Lead Observer (LO), and two Support Observing Scientists (SO) responsible for different tasks. The OA is present at the telescope, responsible for its operation, safety, and performance. The task for SO is mainly for monitoring the data pipeline by assessing the data quality and instrument performance using a software named `NightWatch` [8].

[7] `http://www-kpno.kpno.noirlab.edu/Info/Mtn_Weather/`

[8] `https://desi.lbl.gov/trac/wiki/DESIOperations/NightWatch`

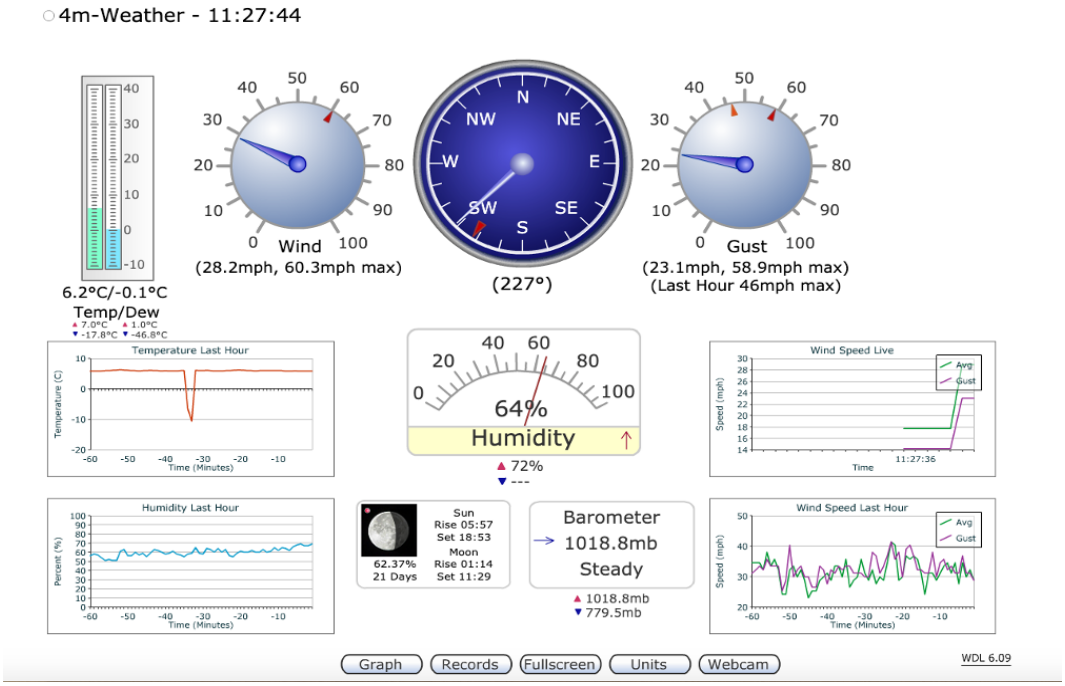

FIGURE 2.17 : Weather monitoring tool on KPNO weather page, that is useful to determine the current/forecast wind and humidity at the telescope.

Before starting observation, the LO needs to set up the system with the following major steps :

- Setup focal plane. We need to distribute all positioners in a starting position and identify the positioners that were frozen due to any issues.
- Denoise GFAs. We need to subtract the additional noise caused by GFAs.
- Home the ADCs (see definition in Section 2.2.3). The ADCs need to be set at their original position, and are adjusted according to the amount of atmosphere that light passes through (airmass).
- Focus. DESI uses an auto-focus pipeline to determine the best setting to focus the instrument, which is applied at the start of each exposure. It is realized by looking at images from the focused GFAs (take an exposure for 60 seconds) and the donut analysis (stars appear as donuts in the images. To achieve a good focus, we need to see holes in the middle of each donut. See Figure 2.18 as an example) in order to set up the best focus settings.

After the system setup, the SO will complete spectrograph calibrations with the following operations :

- Make sure that the system is ready and the ICS instance has been restarted, then run Zeros (CCD intrinsic readout noise measurement) and Darks (CCD electronic noise during observations, this is measured with the shack and the dome being dark) to test if any issues arise in these two cases. These two noise will be subtracted in the data processing pipeline.

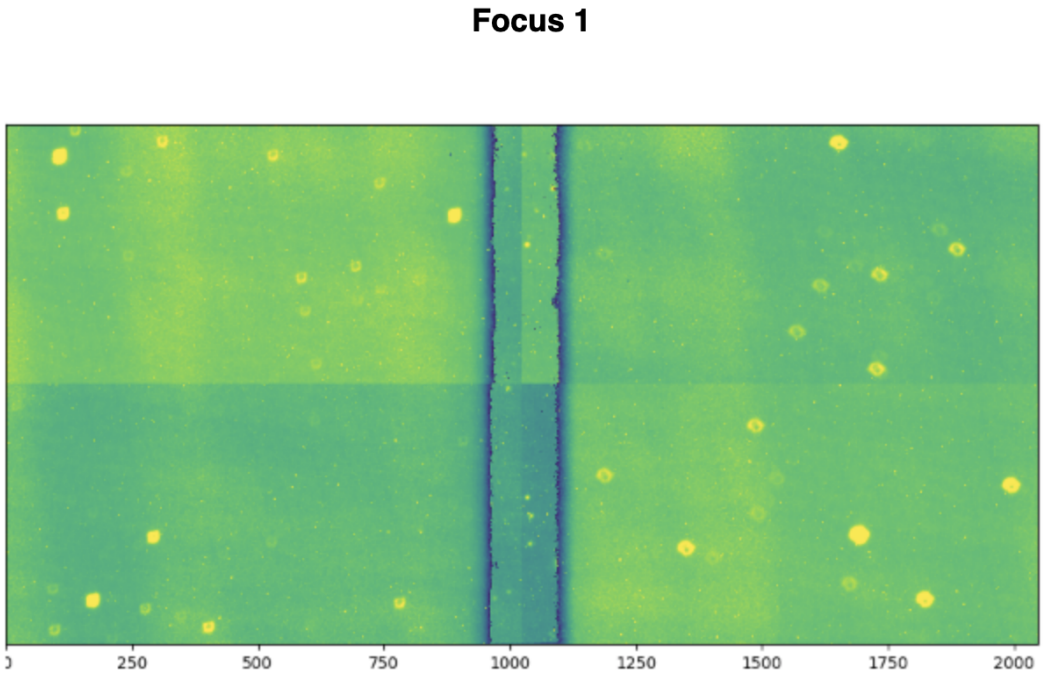

FIGURE 2.18 : An example image from the focus GFAs. Stars appear as donuts in the plot. To achieve a good focus, we need to see holes in the middle of each donut.

- After the electronic maintenance (EM, engineering group) or the OA moves the telescope to the desired position. Confirm that the Zeros are complete, the dome is dark, the mirror covers are opened, and the mirror cooling is on.
- Run calibration tests (with the dome being dark, mirror covers opened, and the telescope pointing at the white spot initial position) and check the images at NightWatch.

Observation will usually start at 12° twilight and end at 12° dawn, where the sun is 12° below the horizon. It is controlled and monitored by the ICS, and can be visualized by a set of software in the DOS (see Figure 2.19). Figure 2.19 shows an example of the all-sky camera taken from one of my observation shifts, and it can be used to visualize the cloud status by eye.

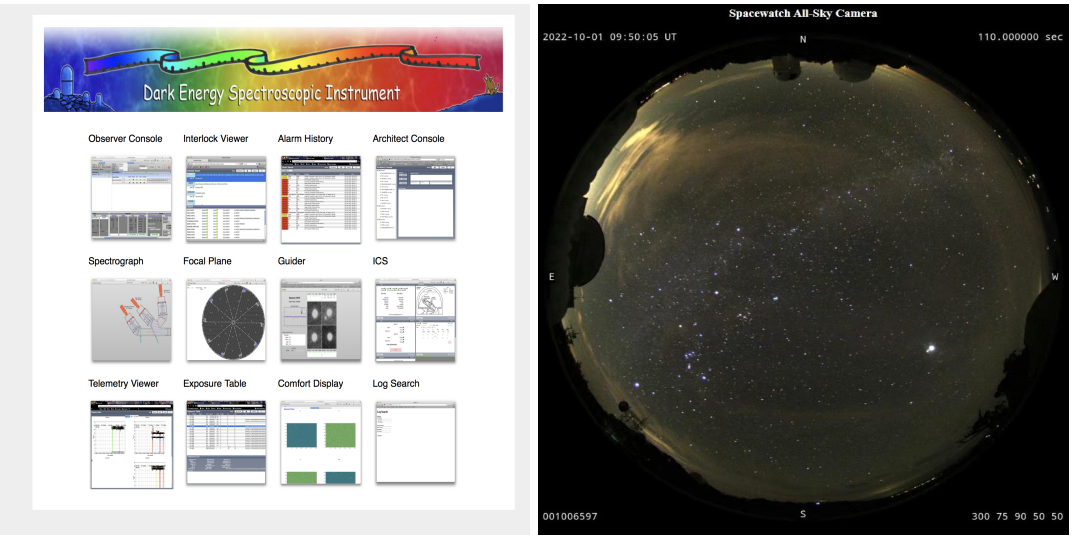

FIGURE 2.19 : The left plot shows all the software in DOS used to control the instrument. The right plot gives an example of the all-sky camera image, taken from one of my observation shifts. There were a few clouds, and some light pollution at the edge of the camera. The overall weather condition and the survey speed were good.

The LO and SO will use the DOS to operate the observations. Figure 2.20 shows the screenshot of the Observer Console GUI that contains most of the operations in the DOS, including the ICS status, the System control, the Exposure control, the Request queue, the details of the current exposure, the request history, the DOS message, the setup, the positioning, the observing conditions, etc.

Data collection

DESI collected its first testing data in October 2019, and completed the Survey Validation data collection (SV also named Fuji, LAN et al. 2022; D. M. ALEXANDER et al. 2023) in December 2020 - June 2021. This dataset is also called the 'One-Percent Survey' since it covers 140 $\deg^2$ of the sky ($\sim$ 1% of the DESI footprint) with typical exposures of the main survey. The SV data is used to characterize the performance of DESI operation, improve the data quality, and validate the scientific requirements described in AGHAMOUSA et al. 2016a.

After SV, the first two months of main survey data collected from mid-May to mid-July 2021 is named Guadalupe. The whole data release combining Fuji, Guadalupe, and one year of main

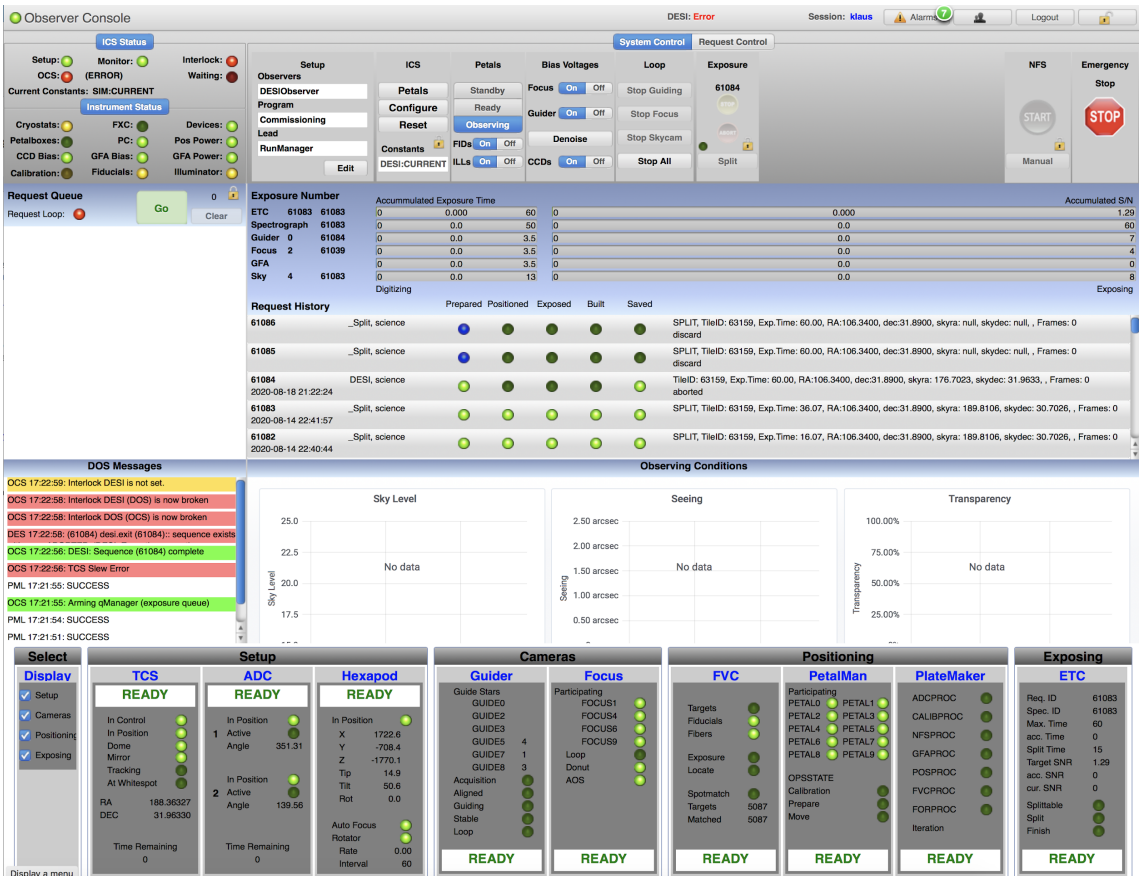

FIGURE 2.20 : Observer Console GUI that contains most of the operations in the DOS.

survey data is named Himalayas/Iron. The DESI collaboration released Guadalupe + Fuji as the Early Data Release (EDR, (Adame et al. 2023 ; Ramirez-Pérez, Pérez-Ràfols et al. 2023)) in early 2023, and plans Iron as Year 1 Data Release (DRY1) about one year later.

Observation status

The main survey started officially on May 2021, and was shut down for several months from June 2022 to September 2022 because of the forest fire near Kitt Peak [9]. The forest fire has left unforeseen implications on the instrument and the services around it. The usage of utility power and internet were restored after 3 months of the fire, and it has caused the warming-up of all 30 cryostats, and irreversible damage to 2 CCDs. However, thanks to the excellent performance of the DESI instrument and pipeline, DESI is functioning with an underestimated survey speed and is still 3 months ahead of schedule for the dark-time survey (observing the main targets of DESI at dark time, mainly for LRGs, ELGs, QSOs). Figure 2.21 shows the status of the survey, with up to November 2022, 35.3% completion of the dark-time survey and 49.7% completion of the bright-time survey (observing BGS and the Milky Way Survey at bright time). The whole survey is expected to be completed in April 2024 for the bright-time program and in May 2025 for the dark-time program.

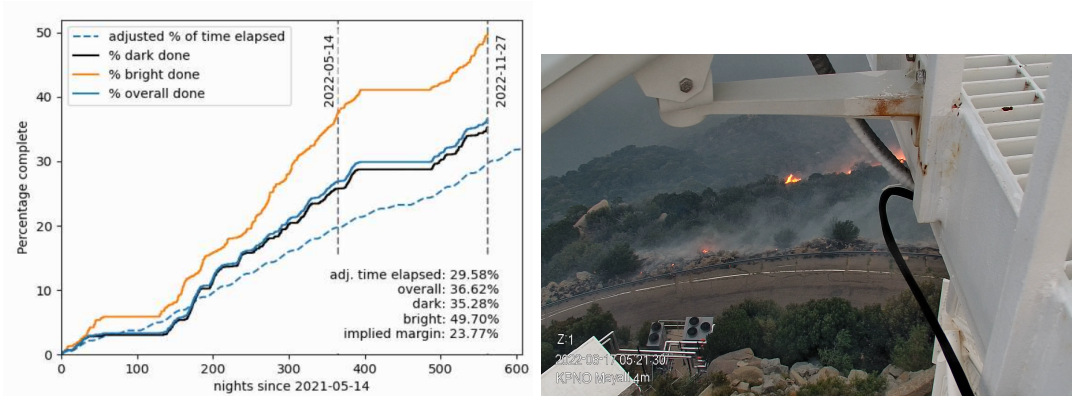

Figure 2.21 : The left plot gives the status of the DESI observation schedule, credits : Michael Levi, DESI collaboration meeting Dec 2022. The right plot shows a side view of the forest fire near the telescope in June 2022, credits : Michael Levi, DESI collaboration meeting June 2022.

2.2.5 DESI spectroscopic pipeline

The raw CCD images collected using the 10 spectrographs of DESI are processed into flux-calibrated spectra with a spectroscopic pipeline (Guy et al. 2023). The post-processed spectra cover a wavelength range from near-UV (3600 Å) to near-infrared (9800 Å) with a pixel width of 0.8 Å. This pipeline is composed of the following main steps :

- CCD calibration, corrections on raw CCD images using calibration images.
- Spectroscopic extraction, a monitoring software nightwatch is used to assess spectra quality during the observation, and an algorithm based on "spectroperfectionism" methodology (A. S. Bolton et D. J. Schlegel 2010) is applied for spectra extraction.

[9] https://noirlab.edu/public/news/noirlab2213/

- Spectral calibration, which is performed using the spectra of calibration stars.
- Co-adding exposures, which co-adds the multiple spectra observed for the same targets.
- Noise evaluation, different sources coming from the CCD are estimated, e.g., Poisson noise, readout noise (measured during the Zeros and Darks), etc.

The sky subtraction and spectra calibration steps mentioned in this section will be modeled for the Lyα analysis, that will be described in Section 3.2.3.

2.2.6 DESI-Ib and DESI-II

Since the DESI instrument has shown excellent performance in its first-year observation, the DESI collaboration is planning to launch a continuous project as an extension to the current survey (detailed in Michael Levi, DESI collaboration meeting Dec 2022). Based on different updates on the instrument and scientific goals, it is designed into two potential proposals, that are still under discussion by the direction board.

DESI-Ib is designed to be a continued commission with three or more years with the existing instrument. The potential surveys can be added directly to the current survey. With an increased completeness and survey area ($3000-4000\text{deg}^2$), the planned dark-time survey can improve systematic errors in the LRG and ELG samples, and the bright-time survey can improve systematic errors for BGS samples.

DESI-II would upgrade the instrument with twice the number of fibers, which needs a more complicated collaboration plan in 8 years in the future. It will be able to explore more science topics, e.g. collecting samples of Lyman Alpha Emitters (LAEs, star-forming galaxies with significant emission peaks that can be easily found by narrow-band detection) and Lyman Break Galaxies (LBGs, high-redshift galaxies with Lyα regions fully absorbed) to have a better understanding of the high redshift Universe, improving dark matter science by mapping the Milky way and a clearer prescription of the local Universe.

2.3 Summary and prospects

In this chapter, I presented an overview of both the scientific and operational designs of the SDSS/eBOSS and DESI survey. These two surveys provide an enormous amount of data for cosmology studies, facilitating particularly the measurement of the BAO peak position with percent-level precision.

This thesis makes use of Lyα forests collected in eBOSS and DESI to measure the BAO. In the next chapter, I will describe in detail the Lyα analysis pipeline from data collection to cosmological interpretation, and the associated model for fitting. This pipeline is then tested and validated using a set of simplified simulations, the so-called mocks, that will be introduced in Chapter 4. I will further present the analysis results on mocks and real data in Chapter 5, and discuss one of the most important systematic effects of this analysis, the HCDs, in Chapter 6.

Bibliographie du présent chapitre

Gunn, J. E., M. Carr et al. (1998). "The Sloan digital sky survey photometric camera". In : *The Astronomical Journal* 116.6, p. 3040.

York, D. G. et al. (2000). "The sloan digital sky survey : Technical summary". In : *The Astronomical Journal* 120.3, p. 1579.

Richards, G. T. et al. (2002). "Spectroscopic target selection in the sloan digital sky survey : The quasar sample". In : *The Astronomical Journal* 123.6, p. 2945.

Eisenstein, D. J., I. Zehavi et al. (2005). "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies". In : *The Astrophysical Journal* 633.2, p. 560.

Croom, S. M. et al. (2009). "The 2dF-SDSS LRG and QSO Survey : the spectroscopic QSO catalogue". In : *Monthly Notices of the Royal Astronomical Society* 392.1, p. 19-44.

Bolton, A. S. et D. J. Schlegel (2010). "Spectro-perfectionism : an algorithmic framework for photon noise-limited extraction of optical fiber spectroscopy". In : *Publications of the Astronomical Society of the Pacific* 122.888, p. 248.

Wright, E. L. et al. (2010). "The Wide-field Infrared Survey Explorer (WISE) : mission description and initial on-orbit performance". In : *The Astronomical Journal* 140.6, p. 1868.

Bovy, J. et al. (2011). "Think outside the color box : Probabilistic target selection and the SDSS-XDQSO Quasar targeting catalog". In : *The Astrophysical Journal* 729.2, p. 141.

Bolton, A. S., D. J. Schlegel et al. (2012). "Spectral classification and redshift measurement for the SDSS-III baryon oscillation spectroscopic survey". In : *The Astronomical Journal* 144.5, p. 144.

Sholl, M. J. et al. (2012). "BigBOSS : a stage IV dark energy redshift survey". In : *Ground-based and Airborne Instrumentation for Astronomy IV*. T. 8446. SPIE, p. 1902-1913.

Delubac, T., J. Rich et al. (2013). "Baryon acoustic oscillations in the Lyα forest of BOSS quasars". In : *Astronomy & Astrophysics* 552, A96.

Levi, M. et al. (2013). "The DESI Experiment, a whitepaper for Snowmass 2013". In : *arXiv preprint arXiv :1308.0847*.

Smee, S. A. et al. (2013). "The multi-object, fiber-fed spectrographs for the sloan digital sky survey and the baryon oscillation spectroscopic survey". In : *The Astronomical Journal* 146.2, p. 32.

Myers, A. D., N. Palanque-Delabrouille et al. (2015). "The SDSS-IV extended Baryon oscillation spectroscopic survey : Quasar target selection". In : *The Astrophysical Journal Supplement Series* 221.2, p. 27.

Aghamousa, A. et al. (2016a). "The DESI experiment part I : science, targeting, and survey design". In : *arXiv preprint arXiv :1611.00036*.

— (2016b). "The desi experiment part ii : Instrument design". In : *arXiv preprint arXiv :1611.00037*.

Dawson, K. S. et al. (2016). "The SDSS-IV extended Baryon Oscillation Spectroscopic Survey : overview and early data". In : *The Astronomical Journal* 151.2, p. 44.

Shen, Y. et al. (2016). "The Sloan Digital Sky Survey reverberation mapping project : velocity shifts of quasar emission lines". In : *The Astrophysical Journal* 831.1, p. 7.

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z= 2.3 with SDSS DR12 Lyα-Forests". In : *Astronomy & Astrophysics* 603, A12.

Blanton, M. R. et al. (2017). "Sloan digital sky survey IV : Mapping the Milky Way, nearby galaxies, and the distant universe". In : *The Astronomical Journal* 154.1, p. 28.

Raichoor, A., J. Comparat et al. (2017). "The SDSS-IV extended Baryon Oscillation Spectroscopic Survey : final emission line galaxy target selection". In : *Monthly Notices of the Royal Astronomical Society* 471.4, p. 3955-3973.

Ata, M. et al. (2018). "The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample : first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2". In : *Monthly Notices of the Royal Astronomical Society* 473.4, p. 4773-4794.

Balland, C. et al. (2018). "QuasarNET : Human-level spectral classification and redshifting with Deep Neural Networks". In : *arXiv e-prints*, arXiv-1808.

Miller, T. N. et al. (2018). "Fabrication of the DESI corrector lenses". In : *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III*. T. 10706. SPIE, p. 256-264.

Dey, A. et al. (2019). "Overview of the DESI legacy imaging surveys". In : *The Astronomical Journal* 157.5, p. 168.

Harpole, A., M. Zingale, I. Hawke et T. Chegini (fév. 2019). *pyro : a framework for hydrodynamics explorations and prototyping*. Version 3.1.

Aghanim, N. et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In : *Astronomy & Astrophysics* 641, A6.

Ahumada, R. 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". In : *The Astrophysical Journal Supplement Series* 249.1, p. 3.

Bautista, J. (2020). "Spectral Reductions, Redshifts, and Catalogs for Cosmology". In : *American Astronomical Society Meeting Abstracts# 235*. T. 235, p. 413-02.

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Farr, J., A. Font-Ribera et A. Pontzen (2020). "Optimal strategies for identifying quasars in DESI". In : *Journal of Cosmology and Astroparticle Physics* 2020.11, p. 015.

Lyke, B. W. et al. (2020). "The Sloan Digital Sky Survey Quasar Catalog : Sixteenth Data Release". In : *The Astrophysical Journal Supplement Series* 250.1, p. 8.

Yèche, C. et al. (2020). "Preliminary Target Selection for the DESI Quasar (QSO) Sample". In : *Research Notes of the AAS* 4.10, p. 179.

Alam, S. et al. (2021). "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey : Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". In : *Physical Review D* 103.8, p. 083533.

Abareshi, B. et al. (2022). "Overview of the instrumentation for the Dark Energy Spectroscopic Instrument". In : *The Astronomical Journal* 164.5, p. 207.

Abdurro'uf, N. et al. (2022). "The seventeenth data release of the Sloan Digital Sky Surveys : Complete release of MaNGA, MaStar, and APOGEE-2 data". In : *The Astrophysical Journal. Supplement Series* 259.2.

Lan, T.-W. et al. (2022). "The DESI Survey Validation : Results from Visual Inspection of Bright Galaxies, Luminous Red Galaxies, and Emission Line Galaxies". In : *arXiv preprint arXiv :2208.08516*.

Myers, A. D., J. Moustakas et al. (2022). "The Target Selection Pipeline for the Dark Energy Spectroscopic Instrument". In : *arXiv preprint arXiv :2208.08518*.

Silber, J. H. et al. (2022). "The Robotic Multi-Object Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)". In : *arXiv preprint arXiv :2205.09014*.

Youles, S. et al. (2022). "The effect of quasar redshift errors on Lyman-α forest correlation functions". In : *Monthly Notices of the Royal Astronomical Society* 516.1, p. 421-433.

Adame, A. et al. (2023). "The Early Data Release of the Dark Energy Spectroscopic Instrument". In : *arXiv preprint arXiv :2306.06308*.

Alexander, D. M. et al. (2023). "The DESI Survey Validation : Results from Visual Inspection of the Quasar Survey Spectra". In : *The Astronomical Journal* 165.3, p. 124.

Chaussidon, E. et al. (2023). "Target Selection and Validation of DESI Quasars". In : *The Astrophysical Journal* 944.1, p. 107.

Guy, J. et al. (2023). "The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument". In : *The Astronomical Journal* 165.4, p. 144.

Hahn, C. et al. (2023). "The DESI Bright Galaxy Survey : Final Target Selection, Design, and Validation". In : *The Astronomical Journal* 165.6, p. 253.

Raichoor, A., J. Moustakas et al. (2023). "Target Selection and Validation of DESI Emission Line Galaxies". In : *The Astronomical Journal* 165.3, p. 126.

Ramirez-Pérez, C., I. Pérez-Ràfols et al. (2023). "The Lyman-alpha forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release". In : *arXiv preprint arXiv :2306.06312*.

Zhou, R. et al. (2023). "Target Selection and Validation of DESI Luminous Red Galaxies". In : *The Astronomical Journal* 165.2, p. 58.

Bailey, S. (in preparation). *Redrock : Spectroscopic Classification and Redshift Fitting for the Dark Energy Spectroscopic Instrumen*.

Jelinsky, P. et al. (2022, in prep). In.

Miller, T. et al. (2022, in prep). In.

Poppett, C. et al. (2022, in prep). In.

Chapitre 3

The Lyα forest correlation function

In this chapter, I describe the pipeline to measure the two-point correlation function of Lyα forests, and their cross-correlation function with quasars. This pipeline was applied for the eBOSS DR16 analysis, and is updated for DESI data with minor developments. As introduced in Section 1.2.4, Lyα forests are seen as a series of absorption lines in quasar spectra. In order to measure their fluctuations, the first important step is to estimate the unabsorbed quasar continuum. This is achieved based on quasar spectra templates and fitting algorithms. The fitting procedure (`Continuum Fitting`) takes into account the instrumental effects, the redshift dependency of Lyα forests, and quasar diversity. Moreover, a distortion effect is introduced by `Continuum Fitting`, thus a distortion matrix (see Section 3.1.1) is applied to each Lyα delta field (fluctuations of Lyα flux absorption in contrast to the unabsorbed quasar continuum) to correct this systematic effect. We measure the correlations of each Lyα delta field pair (or forest-quasar pair) in the angular redshift space (positions parametrized by the angular position θ and redshift z) and make a coordinate transformation to directions along and across the line-of-sight. The correlation functions can then be expressed in a two-dimensional map of $\{r_{||}, r_{\perp}\}$.

To model the Lyα correlation function, an analytical model is developed for its Fourier transform, i.e., the Lyα power spectrum (see Section 3.2.1). We divide the correlation function into two parts : a smooth part that does not contain the BAO peak, and a 'peak-only' part. We model the correlations from all the other sub-dominant contaminants as additional correlations to the total Lyα correlation function, e.g., metals, HCDs (see Section 1.2.4), sky subtraction, and quasar radiation (see Equation 3.22 and Equation 3.23).

Since the cosmological simulations of Lyα forests are computationally expensive to provide an accurate covariance matrix, we use an estimated sub-sample covariance matrix (see Section 3.1.1) for the fitting of Lyα correlation functions. The fitting results will be further discussed in Chapter 5.

During my thesis, I studied several systematic effects of the Lyα analysis. I tested the parameter convergence of the `Continuum Fitting` procedure using mocks and real data (see Section 3.1.1), and proposed a new method to ensure convergence (see Section 3.1.1). Moreover, I analyzed the binning effect and the modeling of HCDs on the Lyα correlation function. In this respect, I will describe a new model, the `Voigt` model, in Chapter 6, which is the main contribution of this thesis to the DESI collaboration.

3.1 Measuring the Lyα correlation function

In this section, I describe the steps of measuring the Lyα correlation functions from mocks or real data, for the auto-correlation function in Section 3.1.1, and for the cross-correlation function in Section 3.1.2. During my thesis, I contributed to the fitting pipeline to estimate the unabsorbed quasar spectra continuum and analyzed the parameter convergence issue (see the following section on continuum convergence).

3.1.1 The Lyα auto-correlation function

The Lyα auto-correlation function is measured following the analysis pipeline detailed in Des Bourboux, Rich et al. 2020. Given a catalog of quasar spectra, we first measure the flux-transmission field $\delta_q(\lambda)$ (the so-called delta field) as :

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{C_q(\lambda)\bar{F}(\lambda)} - 1. \tag{3.1}$$

Here $f_q(\lambda)$ is the observed flux in each quasar line-of-sight q at wavelength λ, $C_q(\lambda)$ is the quasar continuum without absorptions, estimated using a continuum fitting pipeline (see the next subsection). It is thus different for each quasar spectrum. $\bar{F}(\lambda)$ is the mean transmissions averaged over all quasars and is therefore the same for all of them.

Continuum fitting

The spectrum of photons emitted from a quasar will be redshifted according to the quasar redshift. A quasar spectrum without any absorption features is called a quasar continuum, and needs to be fitted in order to study only foreground Lyα absorption. For Lyα mocks, the unabsorbed quasar continuum $C_q(\lambda)$ can be obtained in two ways : either from transmission fields in the raw mocks directly, the so-called `True Continuum` method, or from a normalization fitting using synthetic quasar spectra generated from the routine `quickquasars` (described in Section 4.2), the so-called `Continuum Fitting` method. For real data, only `Continuum Fitting` could be used since the true transmissions are unknown.

This normalization continuum fitting was developed in Delubac, Rich et al. 2013, where the quasar continuum is expressed as a first-order polynomial expansion :

$$C_q(\lambda)\bar{F}(\lambda) = \bar{C}(\lambda_{\rm rf})(a_q + b_q log(\lambda)). \tag{3.2}$$

Here $\bar{C}(\lambda_{\rm rf})$ is the mean of all quasar continua as a function of the rest-frame wavelength $\lambda_{\rm rf}$. In practice, the product of quasar continuum and the mean of transmission $C_q(\lambda)\bar{F}(\lambda)$ is fitted, instead of the continuum itself. a_q and b_q are two parameters accounting for quasar diversity, that are fitted separately for the Lyα ($\lambda_{\rm RF} \in [1040, 1200]$ Å, $\lambda_{\rm RF}$ is the restframe wavelength) and Lyβ wavelength ranges ($\lambda_{\rm RF} \in [920, 1020]$ Å), for each quasar spectrum, by maximizing the likelihood function :

$$\ln L = -\frac{1}{2}\Big(\sum_i \frac{(f_q(\lambda_i) - \bar{F}(\lambda_i)C_q(\lambda_i, a_q, b_q))^2}{\sigma_q^2(\lambda_i)} + \ln(\sigma_q^2(\lambda_i))\Big). \tag{3.3}$$

This likelihood is fitted for each quasar spectrum, and after several runs (by default 20 iterations)

to get convergence. It takes into account the flux variances for each spectrum pixel, computed as

$$\sigma_q^2(\lambda) = \eta(\lambda)\sigma_{\mathrm{pip},q}^2(\lambda) + \sigma_{\mathrm{LSS}}^2(\lambda)(\bar{F}(\lambda)C_q(\lambda))^2 + \epsilon(\lambda)\frac{(\bar{F}C_q(\lambda))^2}{\sigma_{\mathrm{pip},q}(\lambda)}. \tag{3.4}$$

The three components of $\sigma_q^2(\lambda)$ are :

- The instrumental noise $\sigma_{\mathrm{pip},q}(\lambda)$ from the flux uncertainty, which is at first estimated by the analysis pipeline, then corrected by a free parameter $\eta(\lambda)$ after iterations as described below.
- The redshift-dependent variance of the Lyα absorption fields, σ_{LSS}. It is associated with the density of absorbers at a certain redshift, thus is called the large-scale structure variance.
- The quasar diversity variances $\epsilon(\lambda)$, which increases at high SNR. Therefore, this component is re-scaled by $\frac{1}{\sigma_{\mathrm{pip},q}(\lambda)}$, which is proportional to the SNR.

Continuum convergence

The nuisance parameters $\bar{C}$, η, σ_{LSS}, and ϵ are evaluated using the likelihood mentioned in Equation 3.3 with several iterations, which is time-consuming. In order to ensure convergence in a limited number of iterations, I performed an analysis of these parameters for a total number of iterations $n = 20$. The test results on eBOSS DR16 mocks (described in Section 4.1.1) are shown in Figure 3.1. I plot the differences between these parameters ($\bar{C}$, η, σ_{LSS}) of each iteration compared to those of the last (20_{th}) iteration, which is assumed to be the best converged. All iterations are separated by the red dashed lines. Inside a given iteration, all of these parameters are plotted as a function of the wavelength λ ($\lambda \in [3600, 5500]$Å). It can be seen that η and σ_{LSS} converge after 5 iterations. On the other hand, the mean continuum $\bar{C}$ still iterates after 10 iterations. One can tell from Figure 3.1 and Figure 3.2 that the convergence of parameters is slower on data than on mocks. This is because the diversities of quasar spectra in data are more complicated than in mocks, making the continuum fitting harder to convergent.

To improve the convergence speed, I suggested a modified iteration approach :

$$\Phi'_{n+1} = \Phi_n + k(\Phi_{n+1} - \Phi_n), \tag{3.5}$$

where $\Phi \in \{\eta, \sigma_{\mathrm{LSS}}, \epsilon, \bar{C}\}$ is one of the parameters, Φ_n is the fitted parameter at the n_{th} iteration, Φ'_{n+1} is the value estimated at the $n + 1_{\mathrm{th}}$ iteration with the new method, while Φ_{n+1} is the value estimated using the original method. k is an *ad hoc* parameter to adjust the convergence process : $k = 1$ recovers the default original method. I tested $k = 0.5$ (new method, blue lines) on the analysis for the eBOSS DR16 data, and found a significant improvement compared to $k = 1$ (original method, black lines), as shown in Figure 3.2 : $\bar{C}$ converges quickly after 5 iterations with the new method. A further investigation of this k parameter can be explored in future analyses.

The distortion matrix

For each quasar spectrum, the `Continuum Fitting` process introduced in Equation 3.2 estimates the quasar continuum using the observed flux $f_q(\lambda)$ of all the pixels in the wavelength range of interest. This fitting method introduces an additional correlation for the continuum at each wavelength λ_i, from the observed flux at all other wavelengths λ_j, and in the end results in a distortion effect of the Lyα correlation function, as shown in J. E. Bautista et al. 2017. This additional correlation on each measured flux-transmission field $\hat{\delta}_q(\lambda_i)$ (the hat denotes the

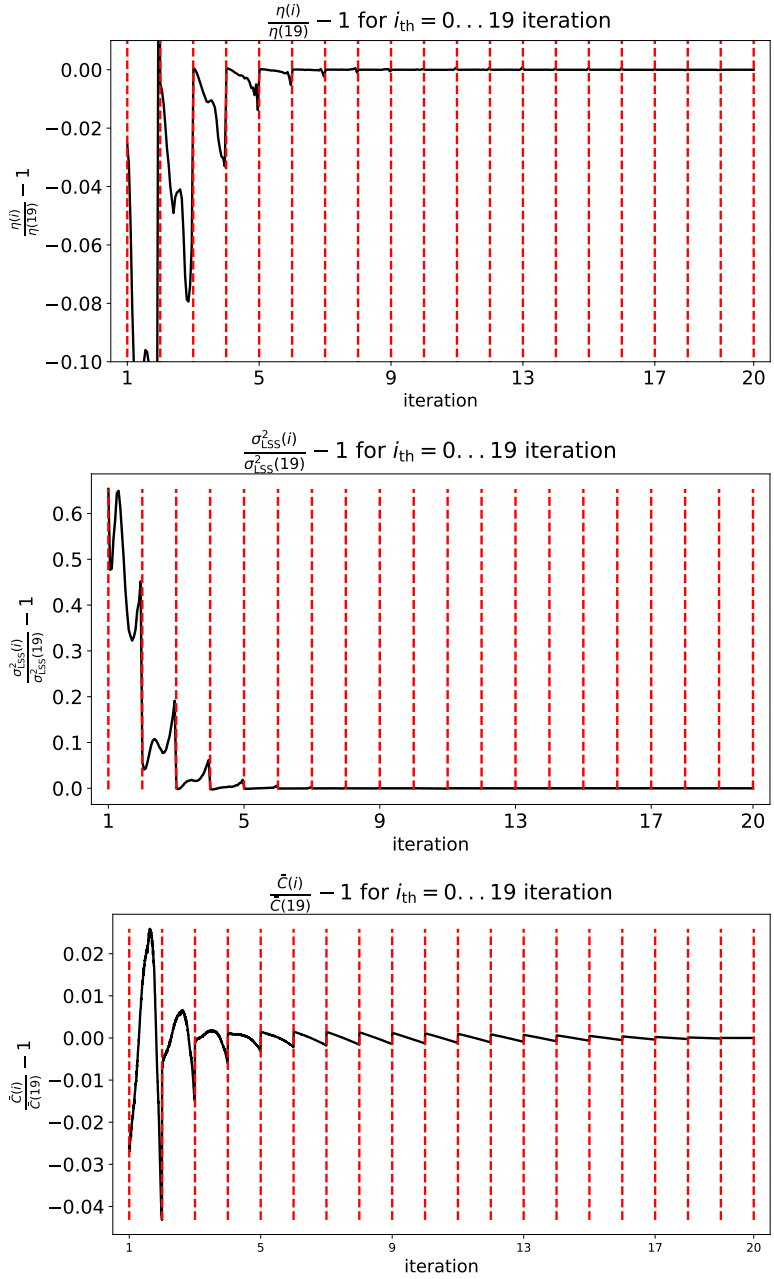

FIGURE 3.1 : The convergence of the parameters $\{\eta, \sigma_{\rm LSS}, \bar{C}\}$ for a Saclay mock (see Section 4.1.1) in 20 iterations. The differences between these parameters of each iteration are shown compared to those of the last ($20_{\rm th}$) iteration, which is assumed to be the best converged. All iterations are separated by the red dashed lines. Inside a given iteration, all of these parameters are plotted as a function of the wavelength λ ($\lambda \in [3600, 5500]$Å). For example, the black curve between the first and the second red dashed lines in the bottom plot shows $\frac{\bar{C}_0(\lambda)}{\bar{C}_{19}(\lambda)} - 1$ with $\lambda \in [3600, 5500]$Å, which indicates the difference between the first and the last iteration of $\bar{C}$.

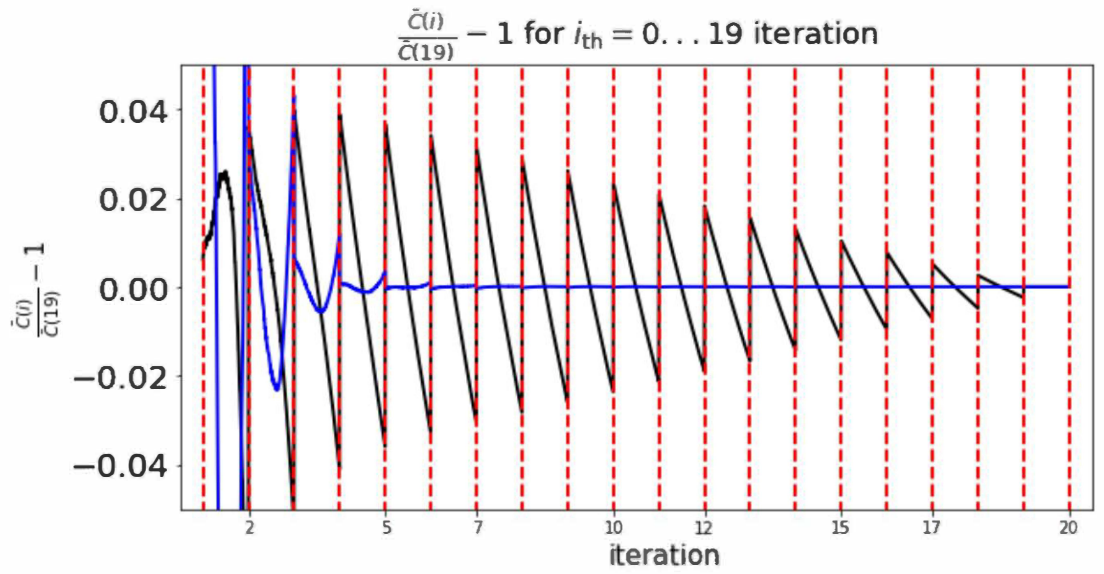

FIGURE 3.2 : The convergence of the parameters $\bar{C}$ for eBOSS DR16 data in 20 iterations. The differences of this parameter for each iteration are shown compared to the one of the last (20_{th}) iteration, which is assumed to be the best converged. All iterations are separated by the red dashed lines. In each iteration, $\bar{C}$ is plotted as a function of the wavelength λ ($\lambda \in [3600, 5500]$Å). A comparison of the new convergence method (blue line) is compared to the original model (black line), as described in Equation 3.5.

measurement) can be seen as a linear combination of all the other true transmission fields $\delta_q(\lambda_j)$ along the same line-of-sight. Therefore, we can model this effect analytically by considering a projection function :

$$\hat{\delta}_q(\lambda_i) = \mathcal{G}(\delta_q(\lambda_{\min}), ..., \delta_q(\lambda_j), ..., \delta_q(\lambda_{\max})). \tag{3.6}$$

Here the transforming function $\mathcal{G}$ can be modeled by applying a distortion matrix η_{ij} to each δ_q by Des Bourboux, Rich et al. 2020 :

$$\delta_q(\lambda_i) \rightarrow \hat{\delta}_q(\lambda_i) = \sum_{j \text{ of all pixels}} \eta_{ij}^q \delta_q(\lambda_j), \tag{3.7}$$

with

$$\eta_{ij}^q = \delta_{ij}^K - \frac{w_j}{\sum\limits_k w_k} - \frac{w_j(\Lambda_i - \bar{\Lambda}_q)(\Lambda_j - \bar{\Lambda}_q)}{\sum\limits_k w_k}. \tag{3.8}$$

Here δ_{ij}^K is the Kronecker delta, k sums over all the pixels, $\Lambda_i = \log \lambda_i$ in log scale, $\bar{\Lambda}_q$ is the mean of $\Lambda_q = \log \lambda$ for the q_{th} quasar spectrum. The weights w_i are estimated by :

$$w_i = \sigma_q^{-2}(\lambda_i)(\frac{1+z_i}{1+2.25})^{\gamma_{\text{Ly}\alpha}-1}, \tag{3.9}$$

taking into account the redshift evolution of the Lyα bias ($\gamma_{\text{Ly}\alpha} = 2.9$ (McDonald, Seljak, Burles et al. 2006)) at an effective redshift $z_{\text{eff}} = 2.25$ (based on SDSS data (York et al. 2000)). The effective redshift is a value near the mean redshift of a quasar sample, determined by minimizing the errors of fitted parameters, see details in Appendix of de Sainte Agathe et al. 2019a. Here $\sigma_q^{-2}(\lambda)$ refers to the pixel variance due to instrumental noise and large-scale structure (LSS) (Des Bourboux, Rich et al. 2020). These pixel variances are also discussed for mocks in Section 4.3.1.

The Lyα auto-correlation

In this section, I describe the method to measure the two-point correlation function using Lyα forests. Given the measured Lyα fluctuation delta field $\hat{\delta}(\vec{x})$ in real space at the position $\vec{x}$ (after applying the distortion matrix), the Lyα two-point auto-correlation function of these delta fields at a fixed separation $\vec{r}$ is given by

$$\xi(\vec{r}) = \langle \hat{\delta}_1(\vec{x}) \hat{\delta}_2(\vec{x} + \vec{r}) \rangle, \tag{3.10}$$

where $\langle\rangle$ denotes an average over $\vec{x}$. The right plot of Figure 3.3 shows a representation of these two delta fields at $\vec{x}$ and $\vec{x} + \vec{r}$. Practically, all of these Lyα delta fields are distributed in the angular redshift space, by their angular position θ and redshift z. Therefore, for two delta fields positioned at (θ, z) and $(\theta + \Delta\theta, z + \Delta z)$ with a separation $(\Delta\theta, \Delta z)$, we perform a coordinate transformation from (θ, z) to $(r_{||}, r_\perp)$:

$$\begin{aligned} r_{||} &= [D_c(z) - D_c(z + \Delta z)] \cos(\frac{\Delta\theta}{2}), \\ r_\perp &= [D_{\text{M}}(z) + D_{\text{M}}(z + \Delta z)] \sin(\frac{\Delta\theta}{2}), \end{aligned} \tag{3.11}$$

where $(r_{||}, r_\perp)$ refer to the directions along and across the mid-point line-of-sight, which is determined by $\mu = \frac{r_{||}}{|\vec{r}|}$. This coordinate transformation is visualized in the left plot of Figure 3.3. Here

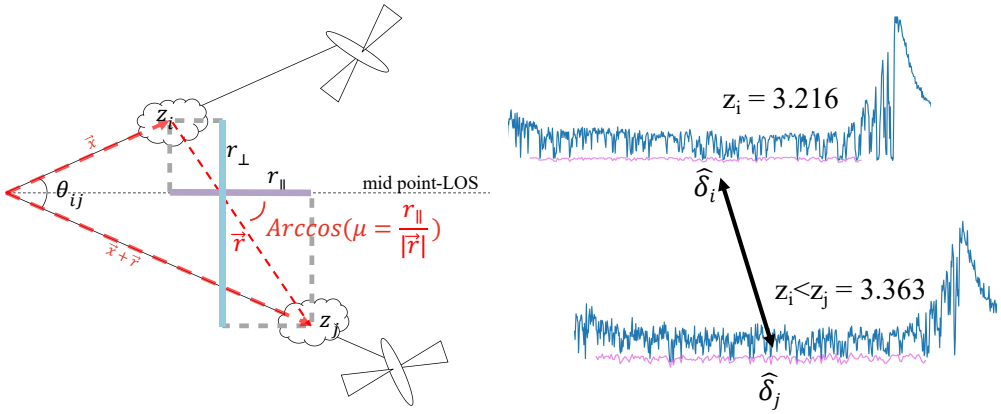

FIGURE 3.3 : The observation of two different Lyα tracers from the same observer (left plot). The direction between the two observed quasars is defined by $\mu = \frac{r_\parallel}{|\vec{r}|}$. The right plot shows Lyα forests from two different lines-of-sight. The two-point correlation function takes all the delta pairs of these delta fields at a certain pixel separation.

$D_c(z) = \int_0^z \frac{dz}{H(z)}$ is the comoving distance, and $D_M(z)$ is the comoving angular diameter distance.

To estimate the 3D Lyα auto-correlation function, we then use an estimator to sum up all the pairs of these delta fields :

$$\hat{\xi}_A = \frac{\sum\limits_{(i,j)\in A} w_i w_j \hat{\delta}_i \hat{\delta}_j}{\sum\limits_{(i,j)\in A} w_i w_j}. \tag{3.12}$$

Here A defines the bins in $(r_{||}, r_\perp)$ space for the correlation function. $\hat{\delta}_i$ and $\hat{\delta}_j$ are two different tracers, i and j refer to wavelength indices, and w_i refers to Lyα weights as defined in Equation 3.9.

We compute the correlations for all pixel pairs within $[0, 200]h^{-1}$Mpc and for separations of $4h^{-1}$Mpc in both directions, which means that binsize is $4h^{-1}$Mpc for both $r_{||}, r_\perp$ and $r_\perp$. The measured correlation function has $N_{\rm bin} = 50 \times 50 = 2500$ bins. For eBOSS DR16 mocks with $\sim 200,000$ quasars, this amounts to $\sim 8.2 \times 10^{12}$ pairs of correlations.

The covariance matrix

For all separation bins in $(r_{||}, r_\perp)$ space, the covariance matrix element between two different bins A and B is defined as

$$C_{\rm AB} = \langle \hat{\xi}_{\rm A} \hat{\xi}_{\rm B} \rangle - \langle \hat{\xi}_{\rm A} \rangle \langle \hat{\xi}_{\rm B} \rangle, \tag{3.13}$$

where $\hat{\xi}_{\rm A}$ and $\hat{\xi}_{\rm B}$ are values of the correlation function in two bins. The whole covariance matrix thus contains $N^2_{\rm bin} = 2500^2$ elements. The correlation matrix is then defined by normalizing the covariance matrix by its diagonal terms :

$$Corr_{\rm AB} = \frac{C_{\rm AB}}{\sqrt{C_{\rm AA} C_{\rm BB}}}. \tag{3.14}$$

In practice, we use the sub-sampling estimation (Delubac, J. E. Bautista et al. 2015) of the covariance matrix by dividing the entire survey into ~ 880 sub-samples, with $n_{\rm side} = 16$ (number of `HEALPIX` pixels per side). In this way, each sub-sample covers $3.7^2 = 13.4$ deg^2 of the sky.

We then estimate the covariance matrix in Equation 3.13 as a weighted average of all these sub-samples (details can be found in Des Bourboux, Rich et al. 2020) :

$$C_{\rm AB} = \frac{1}{W_{\rm A} W_{\rm B}} \sum_s W^s_{\rm A} W^s_{\rm B} [\hat{\xi}^s_{\rm A} \hat{\xi}^s_{\rm B} - \hat{\xi}_{\rm A} \hat{\xi}_{\rm B}], \tag{3.15}$$

where $W_{\rm A}$ is a summed weight over sub-samples s, $W_{\rm A} = \sum_s W^s_{\rm A}$, and $W^s_{\rm A}$ is the sum of the weights of pairs in sky pixels s contributed to bin Az. We use a total number of 880 sub-samples to estimate the covariance matrix, which proved to give a good relative statistical precision of ~ 0.02 (Delubac, J. E. Bautista et al. 2015) for each element of the correlation matrix (and percent level constraint on BAO). This sub-sample method gives an estimation compared to the true covariance matrix, and the number of sub-samples is chosen to match at the same magnitude the number of plates on which the quasars are observed. Moreover, the correlations between different sub-samples are neglected.

The diagonal terms of the covariance matrix are the variances of the correlation function $C_{\rm AA} = {\rm Var}_{\rm A}$, and are inversely proportional to the number of pair counts (Des Bourboux, Rich et al. 2020) :

$${\rm Var}_{\rm A} \approx \frac{\langle \hat{\delta}^2 \rangle^2}{f N^{\rm pair}_{\rm A}}. \tag{3.16}$$

Here $N^{\rm pair}_{\rm A}$ indicates the number of correlation pairs, $\langle \hat{\delta}^2 \rangle^2$ gives the variance of the transmission fields, and f is a factor showing the effective decrease of the number of pairs due to the correlations between neighboring pixels. This variance is further discussed in Section 5.2 for the data quality comparison of eBOSS DR16 and DESI data. The off-diagonal terms of the covariance matrix, however, are much noisier and have a dependence on the separation $(\Delta r_{||}, \Delta r_{\perp})$. We smooth these terms by applying the following approach to the correlation matrix :

$${\rm Corr}^s_{\rm AB} = \frac{\sum_{\rm A',B'} {\rm Corr}_{\rm A',B'}}{N_{\rm AB}}, \tag{3.17}$$

where A′, B′ refer to all the correlation bins that satisfy $\Delta r_{||} = r^{\rm A'}_{||} - r^{\rm B'}_{||}$ and $\Delta r_{\perp} = r^{\rm A'}_{\perp} - r^{\rm B'}_{\perp}$, and $N_{\rm AB}$ is the number of correlation pairs for this normalization.

3.1.2 The Lyα-quasar cross-correlation function

The estimator of the Lyα-quasar cross-correlation function is defined similarly as Equation 3.12 :

$$\xi_A = \frac{\sum\limits_{(i,j)\in A} w_i w_j \hat{\delta}_i}{\sum\limits_{(i,j)\in A} w_i w_j}, \tag{3.18}$$

where quasars are considered as point-like objects with $\hat{\delta}_j = 1$. w_i are the Lyα weights defined in Equation 3.9. However, quasars evolved with different redshift dependence compared to Lyα

forests. We therefore use different weights for quasars :

$$w_j = (\frac{1+z_j}{1+2.25})^{\gamma_{\rm QSO}-1}, \tag{3.19}$$

with $\gamma_{\rm QSO} = 1.44 \pm 0.08$ (Des Bourboux, Rich et al. 2020) at an effective redshift $z_{\rm eff} = 2.25$. The cross-correlation function takes all the pair counts for the same range of $r_\perp$ as before, $r_\perp \in [0, 200]h^{-1}$Mpc with bin width of $4h^{-1}$Mpc. However, since we can take into account quasars in front of the forests with $z_{\rm QSO} < z_{{\rm Ly}\alpha}$, we are capable to look at a wider range for $r_{||}$, namely $r_{||} \in [-200, 200]h^{-1}$Mpc ($r_{||}$ is negative when $\Delta z > 0$, Δz is defined in Equation 3.11 as the redshift difference : $z_{\rm QSO} - z_{{\rm Ly}\alpha}$). In this case, the total number of correlation bins $N_{\rm bin}$ changes from $2,500$ to $5,000$.

3.2 Modeling of the Lyα correlation function

The modeling of the Lyα correlation function is developed in a way such that the position of the BAO peak is independent of the smooth part of the correlation function (which does not contain the BAO peak). The position of the BAO peak is detected in the angular-redshift space, as a function of the angular and redshift separations $(\Delta\theta, \Delta z)$, and after a coordinate transformation (see Equation 3.10), as a function of $(r_{||}, r_{\perp})$. This can be further derived as a function of the Alcock-Paczynski parameters $\alpha_{||}$ and $\alpha_{\perp}$ (Alcock et Paczyński 1979), which give the normalized proportionalities of $D_{\rm H}$ and $D_{\rm M}$ to r_d (introduced in Section 1.1) :

$$\begin{aligned} \alpha_{||} &= \frac{D_{\rm H}(z_{\rm eff}/r_d)}{D_{\rm H}(z_{\rm eff}/r_d)_{\rm fid}} \\ \alpha_{\perp} &= \frac{D_{\rm M}(z_{\rm eff}/r_d)}{D_{\rm M}(z_{\rm eff}/r_d)_{\rm fid}}. \end{aligned} \tag{3.20}$$

Here the subscript 'fid' refers to the fiducial cosmology for which $\alpha_{||} = \alpha_{\perp} = 1$. With this parametrization, the Lyα correlation function can be separated into two parts :

$$\xi(r_{||}, r_{\perp}, \alpha_{||}, \alpha_{\perp}) = \xi_{\rm smooth}(r_{||}, r_{\perp}) + \xi_{\rm peak}(r_{||}\alpha_{||}, r_{\perp}\alpha_{\perp}), \tag{3.21}$$

where $\xi_{\rm smooth}$ is the smooth correlation function underneath the BAO peak, and $\xi_{\rm peak}$ is the peak-only function.

In addition to the Lyα forest auto-correlations, the cross-correlations of Lyα forests with absorbers such as HCDs (see Section 1.2.4) or metals present in the Lyα (or Lyβ) absorption region, or their auto-correlations, could contribute to the total correlation function, and thus bias our measurements. Moreover, the sky subtraction (see Section 2.2.5 Section 3.2.3) in the analysis pipeline also has a non-negligible impact. To model all of these sub-dominant systematics, the total Lyα auto-correlation function is written as :

$$\xi^{{\rm Ly}\alpha\times{\rm Ly}\alpha}_{\rm Total} = \xi^{{\rm Ly}\alpha\times{\rm Ly}\alpha} + \sum_i \xi^{{\rm Ly}\alpha\times m_i} + \sum_{i,j} \xi^{m_i\times m_j} + \xi^{sky}, \tag{3.22}$$

where $\xi^{{\rm Ly}\alpha\times{\rm Ly}\alpha}$ refers to the Lyα absorption auto-correlation function, $m_{i,j}$ refer to different absorbers, and ξ^{sky} is the sky-subtraction correlations. The Lyα-quasar cross-correlation function can be expressed as :

$$\xi^{{\rm Ly}\alpha\times{\rm QSO}}_{\rm Total} = \xi^{{\rm Ly}\alpha\times{\rm QSO}} + \sum_i \xi^{{\rm QSO}\times m_i} + \xi^{TP}, \tag{3.23}$$

with the last term ξ^{TP} modeling the quasar ionizing radiation reducing the Lyα absorption on the surrounding IGM gas. This is the so-called Transverse Proximity (TP) effect (Font-Ribera, Arnau et al. 2013), which is mainly along the line-of-sight. We model this effect assuming an isotropic emission from quasars as :

$$\xi^{TP} = \xi^{TP}_0 \left(\frac{1h^{-1}{\rm Mpc}}{r}\right)^2 \exp\left(\frac{-r}{\lambda_{\rm UV}}\right) \tag{3.24}$$

with the amplitude ξ^{TP}_0 as free parameter, and $\lambda_{\rm UV} = 300h^{-1}$Mpc (Rudie, Steidel, Shapley et Pettini 2013).

3.2.1 Modeling of the Lyα power spectrum

The modeling of the Lyα correlation function is usually performed by modeling its Fourier transform, i.e., the Lyα power spectrum. In the following paragraphs, I describe the quasi-linear power spectrum that models the large-scale structure growth, as well as the non-linear effects involved at small scales.

The quasi-linear power spectrum

The two components of the correlation function in Equation 3.21 are compted by using the linear power spectrum from `CAMB` at a fiducial cosmology and an effective redshift. Starting from a linear power spectrum $P_{\rm L}$, it is firstly Fourier transformed into a correlation function $\xi_{\rm L}$, which only contains the large-scale fluctuations. The smoothed component $\xi_{\rm smooth}$ is obtained by fitting $\xi_{\rm L}$ in the outside region of the BAO peak and interpolating $\xi_{\rm L}$ in the peak region ($86-150h^{-1}$Mpc). $P_{\rm smooth}$ is then the Fourier transform of $\xi_{\rm smooth}$, and $P_{\rm peak}$ is derived by subtracting $P_{\rm smooth}$ from $P_{\rm L}$, yielding :

$$P_{\rm QL}(\vec{k},z) - P_{\rm smooth}(\vec{k},z) = \exp\left(-\frac{k_{||}^2\Sigma_{||}^2 + k_{\perp}^2\Sigma_{\perp}^2}{2}\right)P_{\rm peak}(\vec{k},z), \tag{3.25}$$

where the subscript QL denotes the quasi-linear power spectrum since this has not included the non-linear small-scale effect of Lyα forests. The non-linear correction for BAO broadening (velocity and nonlinear collapse of matter move the position of BAO peak) is modeled as a Gaussian with two parameters ($\Sigma_{||}, \Sigma_{\perp}$), with $\Sigma_{||} = 6.41h^{-1}$Mpc ($\Sigma_{||} = \Sigma_{\perp}(1+f)$, $f \sim \Omega_{\rm m}^{0.55}(z)$ is the linear growth rate of structure) and $\Sigma_{\perp} = 3.26h^{-1}$Mpc (D. J. Eisenstein, H.-J. Seo et White 2007). This effect mainly happens at low redshift, and its redshift evolution is a negligible second-order effect.

Non-linear effects at small scales

At small scales, the non-linear effects of Lyα forests and quasars need to be taken into account and are modeled as a non-linear function for Lyα forests $D^{\rm auto}_{\rm NL,Ly\alpha}(\vec{k})$, and for quasars $D^{\rm auto}_{\rm NL,QSO}(\vec{k})$. Regarding the Ly$\alpha$ auto-correlation function, the non-linear growth of structure will enhance the power spectrum, while pressure due to thermal broadening and peculiar velocities will suppress the power spectrum along the line-of-sight (the finger-of-god effect (Jing et Börner 2001)). A fitting function for $D^{\rm auto}_{\rm NL,Ly\alpha}(\vec{k})$ was proposed in McDonald 2003 and revised in Arinyo-i-Prats, Miralda-Escudé, Viel et Cen 2015 with higher resolution simulations :

$$D^{\rm auto}_{\rm NL,Ly\alpha}(\vec{k}) = \exp\left\{\left[q_1\Delta^2(\vec{k}) + q_2\Delta^4(\vec{k})\right]\left[1 - \left(\frac{\vec{k}}{k_v}\right)^{a_v}\mu^{b_v}\right] - \left(\frac{\vec{k}}{k_p}\right)^2\right\}, \tag{3.26}$$

with

$$\Delta^2(\vec{k}) = \frac{1}{2\pi^2}\vec{k}^3 P_{\rm QL}(\vec{k}). \tag{3.27}$$

Here $\Delta^2(\vec{k})$ refers to the linear matter density fluctuations and $\Delta^4(\vec{k})$ refers to higher order fluctuations. This fitting function has 6 free parameters, where q_1 and q_2 characterize the power spectrum amplitude, k_p describes the suppression below the Jeans scale (Rorai, Hennawi et White 2013) due to gas pressure, and $\{k_v, a_v, b_v\}$ are used to characterize the quasar non-linear peculiar velocities and the thermal broadening effect. This fitting function is further discussed for Lyα forests and HCDs in Figure 6.22 of Section 6.5.

For the Lyα-quasar cross-correlation, the quasar non-linear peculiar velocities have the biggest impact, and $D^{\rm cross}_{\rm NL,Ly\alpha}(\vec{k})$ is given by W. J. Percival et White 2009 :

$$D^{\rm cross}_{\rm NL,Ly\alpha}(\vec{k}) = \frac{1}{1+(k_{||}\sigma_v)^2}, \tag{3.28}$$

where σ_v is a free parameter that characterizes the quasar velocity dispersion. Note that this effect suppresses the power spectrum at a comparable scale as the HCD damping effect, thus is hard to be constrained (see Section 6.5).

In practice, the quasar redshifts are measured using the quasar classification pipeline described in Section 2.2.2, thus resulting in a systematic effect along the line-of-sight on $r_{||}$:

$$r_{||} = r_{||,\rm measure} + \Delta r_{||,\rm QSO}. \tag{3.29}$$

This effect is discussed in detail in Youles et al. 2022.

The binning effect

Since the correlation functions are measured on separation grids with bins of a given width, as described in Section 3.1, the effect of this binning needs to be modeled. The Fourier transform of the rectangle bins can be modeled as a product of sinc functions :

$$G(\vec{k}) = {\rm sinc}(\frac{k_{||}R_{||}}{2}){\rm sinc}(\frac{k_{\perp}R_{\perp}}{2}), \tag{3.30}$$

where $R_{||}$ and $R_{\perp}$ refer to the widths of bins along the radial and transverse directions (here $R_{||} = R_{\perp} = 4h^{-1}{\rm Mpc}$).

The Lyα power spectrum

Gathering all the effects above and taking into account the fact that the tracers are biased, I hereby present the entire expression of the Lyα power spectrum model :

$$P_{\rm F}(\vec{k}) = b_i b_j (1+\beta_i \mu_k^2)(1+\beta_j \mu_k^2) P_{\rm QL}(\vec{k}) D_{\rm NL,Ly\alpha}(\vec{k}) G(\vec{k}), \tag{3.31}$$

where the indices i and j refer to different tracers : $i = j =$ Lyα for the Lyα auto-correlation and $i =$ Lyα, $j =$ QSO for the Lyα-quasar cross-correlation. b is the bias parameter and β is the RSD parameter (see Section 1.1). Since the correlation functions are measured at an effective redshift, we assume a redshift dependence of $b_{\rm Ly\alpha} \propto (1+z)^{\gamma_{\rm Ly\alpha}-1}$ with $\gamma_{\rm Ly\alpha} = 2.9$ (McDonald, Seljak, Burles et al. 2006), and an approximation that $\beta_{\rm Ly\alpha}$ does not have redshift evolution. In this case, the effective redshift could be calculated. For the cross-correlation, we use a quasar redshift dependence as

$$b_{\rm QSO}(z) = 3.77(\frac{1+z}{1+2.334})^{1.44}, \tag{3.32}$$

by setting $b_{\rm QSO}(z_{\rm eff}) = 3.77$ at the effective redshift $z_{\rm eff} = 2.334$ (for eBOSS DR16, we use $z_{\rm eff} = 2.376$ for DESI EDR data). The quasar RSD effect is also assumed to be redshift independent and $\beta_{\rm QSO}$ is expressed as :

$$\beta_{\rm QSO} = \frac{f}{b_{\rm QSO}}, \tag{3.33}$$

where $f = 0.9704$ and $\beta_{\rm QSO} = 0.269$ for the fiducial cosmology of eBOSS DR16 analysis (Ade et al. 2016). For DESI the fiducial cosmology uses Aghanim et al. 2020.

3.2.2 Astrophysical contaminants

Astrophysical contaminants that are observed in quasar spectra such as HCDs and metals, contribute to the Lyα correlation function. They can be modeled as biased discrete matter tracers with their unique biases and redshift dependencies.

High column density systems

High Column Density systems (HCDs, see Section 1.2.4) are seen as strong absorptions with damping wings in the Lyα forests and are usually parametrized by a Voigt profile. The Voigt profile fitting is further described in Section 6.1.1. According to these absorption profiles, HCDs are classified into Damped Lyman-alpha systems (DLAs) with $N_{\rm HI} > 10^{20.3}{\rm cm}^{-2}$ and Lyman limit systems (LLS), with $10^{20.3}{\rm cm}^{-2} > N_{\rm HI} > 10^{17.2}{\rm cm}^{-2}$. DLAs are detectable using visual inspection or machine learning algorithms (see Section 6.1). Therefore, we usually mask the DLAs and smooth the Lyα fluctuations in the delta field. An example is shown in Figure 3.4.

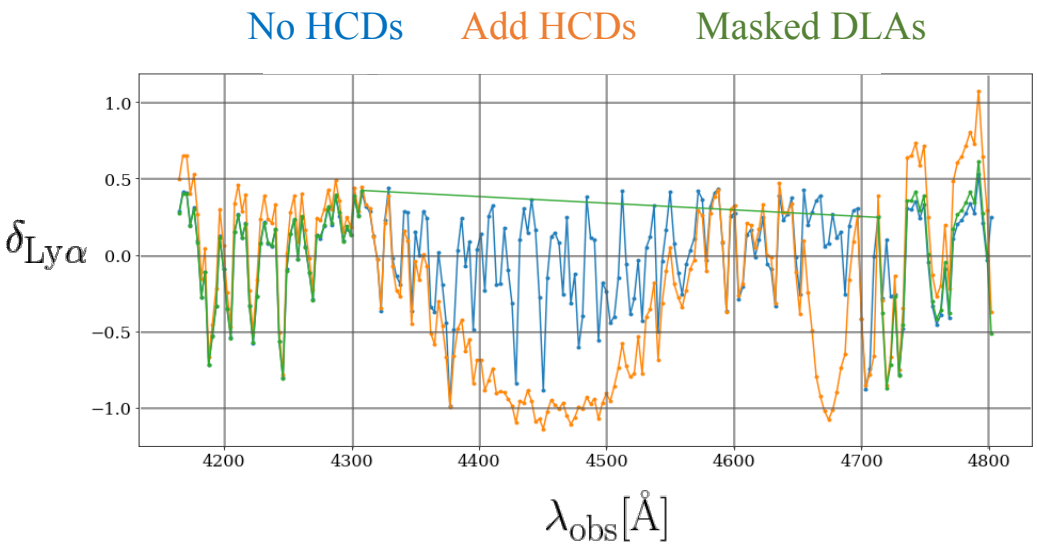

Figure 3.4 : The Lyα transmission (see definition in Section 3.1.1) fluctuation fields, obtained from different Saclay mocks : mocks without HCDs (Blue), mocks with HCDs (Orange), and mocks with DLAs masked (Green). The green curve shows our approach of masking the DLAs and smoothing the fluctuation fields.

Although most large DLAs can be detected and masked, the HCDs undetected because of the limitation of machine learning algorithms will bias our measurement of the Lyα correlation function, since they are falsely treated as Lyα absorption lines. Moreover, the damping wings of HCDs will result in a suppression of the Lyα power spectrum in Fourier space. The modeling of this effect on the Lyα correlation function is detailed in Section 6.2, and I will simply describe

the model here. Similarly to the modeling of Lyα forests that are treated as biased tracers of the underlying dark matter density field with its specific bias and RSD effects (Equation 3.31), HCDs can also be modeled with their characteristic bias and RSD parameters. The contribution of HCDs and Lyα forests can be combined to get the effective Lyα bias and RSD parameters :

$$\begin{aligned} b'_{\mathrm{Ly}\alpha} &= b_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}} F_{\mathrm{HCD}}(k_{||}) \\ b'_{\mathrm{Ly}\alpha}\beta'_{\mathrm{Ly}\alpha} &= b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}}\beta_{\mathrm{HCD}} F_{\mathrm{HCD}}(k_{||}), \end{aligned} \tag{3.34}$$

where b_{HCD} and β_{HCD} refer to the bias and RSD effect of HCDs. $F_{\mathrm{HCD}}(k_{||})$ is a non-linear function to model the HCD damping effect, which is an exponential function $F_{\mathrm{HCD}}(k_{||}) = \exp(-L_{\mathrm{HCD}} k_{||})$ (Equation 6.22) in the eBOSS DR16 analysis (Des Bourboux, Rich et al. 2020).

Broad absorption line quasars

Broad absorption line (BAL) QSOs are a subclass of QSOs with blue-shifted broad absorption lines with velocities larger than 2000 kms^{-1} (R. J. Weymann, S. L. Morris, C. B. Foltz et P. C. Hewett 1991). Depending on the spectral lines that show broad absorptions, BAL QSOs are divided into different classes, such as HiBALs (C IV λ1549), LoBAL (Mg II), FeLoBALs (Fe II), etc. For Lyα forests with $z > 2$ and wavelength range $\lambda_{\mathrm{rf}} \in [1040, 1200]$Å, broad absorption features from N V λ1239, λ1243, O VI λ1032, λ1038, P V λ1118, λ1128 (Ak et al. 2014 ; Hamann, Herbst, Paris et Capellupo 2019) could have non-negligible impacts, and these QSOs are discarded in our analysis. The fraction of these BAL QSOs over the whole QSO sample strongly depends on the selection method, i.e., $10 - 30\%$ using UV and optical wavelengths (C. Foltz, Chaffee, P. Hewett, R. Weymann et S. Morris 1990 ; J. R. Trump et al. 2006), and 40% using IR wavelengths (Dai, Shankar et Sivakoff 2008). In our analysis, a Convolutional Neural Network (CNN) (Guo et Martini 2019) is used to classify BAL QSOs and generate the associated catalog, which was used in the eBOSS DR16 analysis (Des Bourboux, Rich et al. 2020). Two characteristic indices, the Balnicity Index (BI) (R. J. Weymann, S. L. Morris, C. B. Foltz et P. C. Hewett 1991) and the Intrinsic Absorption index (AI) (P. B. Hall et al. 2002), are used to characterize BALs using C IV and Si IV absorption regions. BI is defined as an integration of the quasar flux over the blueshift velocity range from 25,000 to 3,000 kms^{-1} for C IV or Si IV :

$$\mathrm{BI} = -\int_{25000}^{3000} (1 - \frac{f(v)}{0.9}) C(v) dv, \tag{3.35}$$

where $f(v)$ is the quasar flux at the shifted velocity v relative to the considered emission line, $C(v)$ is a binary function defined as :

$$C(v) = \begin{cases} 1, \text{ when } (1 - \frac{f(v)}{0.9}) \text{ continuously positive over a } 2,000 \text{ kms}^{-1} \text{ wide range of velocities,} \\ 0, \text{ on the other hand.} \end{cases} \tag{3.36}$$

AI is defined as :

$$\mathrm{AI} = -\int_{25000}^{0} (1 - \frac{f(v)}{0.9}) C'(v) dv, \tag{3.37}$$

Catalog	QSOs	BI > 0	BI%	AI > 0	AI%	BI > 0 and AI > 0
DESI EDR	199398	6674	3.3%	27716	13.9%	6662
DESI EDR ($z > 2$)	116523	4314	3.7%	17920	15.4%	4305

TABLEAU 3.1 : Catalog of BAL QSOs for DESI EDR data. The definition of different columns : QSOs : Total number of QSOs in the catalog ; BI> 0 : number of QSOs with BI> 0 ; BI% : percentage of QSOs with BI> 0 ; AI> 0 : number of QSOs with AI> 0 ; AI% : percentage of QSOs with AI> 0 ; BI> 0 and AI> 0 : number of QSOs with both BI> 0 and AI> 0.

where $C'(v)$ is

$$C'(v) = \begin{cases} 1, \text{ when } (1 - \dfrac{f(v)}{0.9}) \text{ continously positive over 450 kms}^{-1} \text{ wide range of velocities,} \\ 0, \text{ on the other hand.} \end{cases} \tag{3.38}$$

Essentially, BI characterizes absorption troughs larger than $2,000$ kms^{-1} while AI determines smaller troughs.

In the DESI EDR data, we found 6662 BAL QSOs out of 199398 QSOs with BI > 0 and AI > 0, as shown in Table 3.1. For Lyα QSOs with $z > 2$, we found 4305 BAL QSOs out of 116523 QSOs with the fraction of ~ 3.7%. The distribution of BI and AI are shown in Figure 4.3, for both DESI EDR data and DESI EDR mocks.

Metal absorption lines

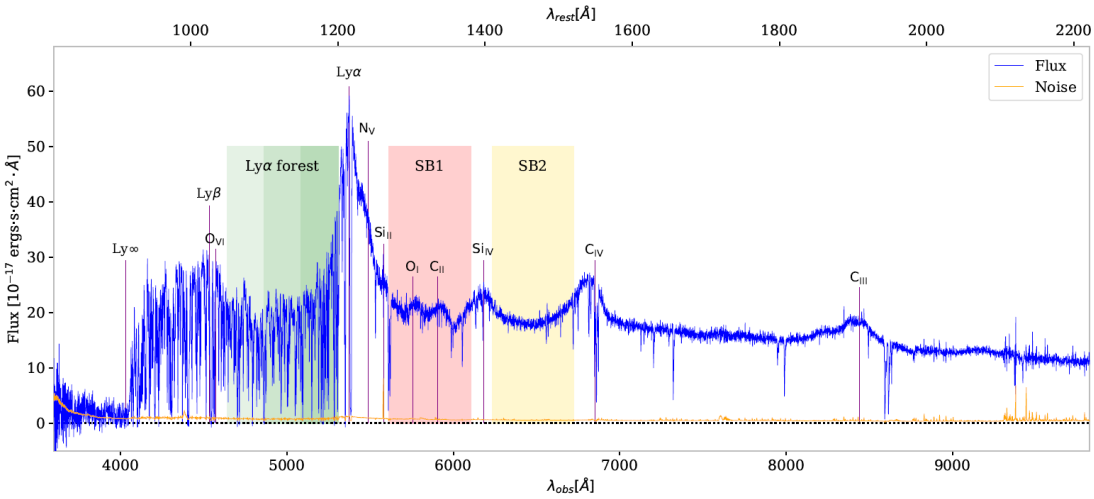

FIGURE 3.5 : A quasar spectrum observed from DESI at the redshift $z_{\rm QSO} = 3.42$, showing different metal absorption lines. Credits : RAVOUX 2022.

Elements with atomic numbers above 2 are referred to as metals, and have absorption and emission spectra different from Hydrogen. Figure 3.5 shows a list of metal emissions observed in DESI quasar spectra, i.e., N V, Si II, O I, C II, Si IV, C IV, C III, etc. Among these metal transitions, redshifted absorption lines of Si III λ1207, Si II λ1190, Si II λ1193 and Si II λ1260

Metal line	$\lambda_{\rm Metal}$[Å]	$r_{\parallel}[h^{-1}{\rm Mpc}]$
Si III	1207	-21
Si II	1190	-64
Si II	1193	-56
Si II	1260	+111

TABLEAU 3.2 : The presence of different metal lines present in quasar spectra affects the measurement of the Lyα correlation function. This table shows the affected positions of different metal lines on the correlation function along the line-of-sight $r_{\parallel}$ at an effective redshift $z_{\rm eff} = 2.334$.

could overlap with Lyα absorptions and be falsely treated as Lyα absorptions (the metal redshift is measured as $\lambda_{\rm obs}/\lambda_{{\rm Ly}\alpha} - 1$ instead of $\lambda_{\rm obs}/\lambda_{\rm Metal} - 1$). This effect will bias the measurement of the Lyα correlation function and need to be added separately, each metal line with its own bias and RSD parameters, as shown in Equation 3.31.

To solve this systematic problem, a reconstruction matrix is applied to each bin of the Lyα correlation function (BLOMQVIST, PIERI et al. 2018). The metals mainly affect the correlation function along the line-of-sight at $r_{\perp} = 0$ and $r_{\parallel} \approx (1+z)D_{\rm H}(z)(\lambda_{\rm Metal} - \lambda_{{\rm Ly}\alpha})/\lambda_{{\rm Ly}\alpha}$. For different metals, the affected position of $r_{\parallel}$ at an effective redshift $z_{\rm eff} = 2.334$ is summarized in Table 3.2 (DES BOURBOUX, RICH et al. 2020).

3.2.3 Sky subtraction

For SDSS (J. E. BAUTISTA et al. 2017 ; DES BOURBOUX, RICH et al. 2020), the sky subtraction is performed for spectra obtained with each spectrograph in the data reduction pipeline. The Poisson fluctuations that are in the sky spectra will induce an extra correlation for data taken from the same spectrograph, and bias the measurement of the auto-correlation function along the line-of-sight. Although this correlation could be removed by subtracting correlation pairs from the same spectrograph, the continuum fitting (described in Section 3.1) generates a smooth distortion along $r_{\parallel}$, that can be modeled using a Gaussian function (DES BOURBOUX, K. S. DAWSON et al. 2019) :

$$\xi^{\rm sky}(r_{\parallel}, r_{\perp}) = \begin{cases} \frac{A_{\rm sky}}{\sigma_{\rm sky}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{r_{\perp}^2}{\sigma_{\rm sky}^2}\right) & \text{, if } r_{\parallel} = 0 \\ 0 \quad \text{, if } r_{\parallel} \neq 0, \end{cases} \tag{3.39}$$

where $A_{\rm sky}$ and $\sigma_{\rm sky}$ are two free parameters referring to the scale and the width of this correlation. The contribution of this sky subtraction correlation to the total correlation function is shown in Equation 3.22.

For DESI, the calibration of spectra is performed for each of the 10 petals (each petal holding one spectrograph). The sky subtraction is modeled with an empirical function (GORDON et al. 2023) :

$$\xi^{\rm sky}(r_{\parallel}, r_{\perp}) = \begin{cases} A_{\rm inst}(\frac{r_{\perp}}{80} - 1)^2 & \text{, if } r_{\perp} < 80h^{-1}{\rm Mpc} \text{ and } r_{\parallel} = 0, \\ 0 \quad \text{, if } r_{\perp} > 80h^{-1}{\rm Mpc} \text{ and } r_{\parallel} \neq 0. \end{cases} \tag{3.40}$$

Here $A_{\rm inst}$ is a free parameter, and the limit of $80h^{-1}$Mpc corresponds to the angular size of a petal on the sky at the effective redshift.

3.3 Summary and prospects

In this chapter, I presented the analysis pipeline for the measurement and the modeling of the Lyα correlation function. To validate this pipeline, different types of Lyα mocks have been developed by several groups within eBOSS/DESI. I will describe the construction of these mocks in the next chapter, and the analysis results using these mocks in Section 5.1 of Chapter 5. Furthermore, this pipeline has been used for previous Lyα analyses (J. E. Bautista et al. 2017 ; de Sainte Agathe et al. 2019a ; Des Bourboux, Rich et al. 2020), and will be used for the DESI Lyα analysis. I will describe in Chapter 5 a preliminary comparison of the Lyα analysis using eBOSS DR16 data and DESI EDR data.

It is essential to have a clear understanding of the various systematic effects mentioned in this chapter. During my thesis, I mainly contributed to the `Continuum Fitting` process (described in this chapter), the binning effect (analysis of the Lyα correlation functions with different binsize), and the modeling of HCDs on the Lyα correlation function. I will describe in detail a new model, the `Voigt` model, in Chapter 6, which is the most important contribution of this thesis to the DESI collaboration.

Bibliographie du présent chapitre

Alcock, C. et B. Paczyński (1979). "An evolution free test for non-zero cosmological constant". In : *Nature* 281.5730, p. 358-359.

Foltz, C., F. Chaffee, P. Hewett, R. Weymann et S. Morris (1990). "On the Fraction of Optically-Selected QSOs with Broad Absorption Lines in Their Spectra". In : *Bulletin of the American Astronomical Society*. T. 22, p. 806.

Weymann, R. J., S. L. Morris, C. B. Foltz et P. C. Hewett (1991). "Comparisons of the emission-line and continuum properties of broad absorption line and normal quasi-stellar objects". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 373, May 20, 1991, p. 23-53.* 373, p. 23-53.

York, D. G. et al. (2000). "The sloan digital sky survey : Technical summary". In : *The Astronomical Journal* 120.3, p. 1579.

Jing, Y. et G. Börner (2001). "Scaling properties of the redshift power spectrum : theoretical models". In : *The Astrophysical Journal* 547.2, p. 545.

Hall, P. B. et al. (2002). "Unusual broad absorption line quasars from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 141.2, p. 267.

McDonald, P. (2003). "Toward a Measurement of the Cosmological Geometry at z˜ 2 : Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets". In : *The Astrophysical Journal* 585.1, p. 34.

McDonald, P., U. Seljak, S. Burles et al. (2006). "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 163.1, p. 80.

Trump, J. R. et al. (2006). "A catalog of broad absorption line quasars from the sloan digital sky survey third data release". In : *The Astrophysical Journal Supplement Series* 165.1, p. 1.

Eisenstein, D. J., H.-J. Seo et M. White (2007). "On the robustness of the acoustic scale in the low-redshift clustering of matter". In : *The Astrophysical Journal* 664.2, p. 660.

Dai, X., F. Shankar et G. R. Sivakoff (2008). "2MASS reveals a large intrinsic fraction of BALQSOs". In : *The Astrophysical Journal* 672.1, p. 108.

Percival, W. J. et M. White (2009). "Testing cosmological structure formation using redshift-space distortions". In : *Monthly Notices of the Royal Astronomical Society* 393.1, p. 297-308.

Delubac, T., J. Rich et al. (2013). "Baryon acoustic oscillations in the Lyα forest of BOSS quasars". In : *Astronomy & Astrophysics* 552, A96.

Font-Ribera, A., E. Arnau et al. (2013). "The large-scale quasar-Lyman α forest cross-correlation from BOSS". In : *Journal of Cosmology and Astroparticle Physics* 2013.05, p. 018.

Rorai, A., J. F. Hennawi et M. White (2013). "A new method to directly measure the Jeans scale of the intergalactic medium using close quasar pairs". In : *The Astrophysical Journal* 775.2, p. 81.

Rudie, G. C., C. C. Steidel, A. E. Shapley et M. Pettini (2013). "The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift< z>= 2.4". In : *The Astrophysical Journal* 769.2, p. 146.

Ak, N. F. et al. (2014). "The dependence of C IV broad absorption line properties on accompanying Si IV and Al III absorption : Relating quasar-wind ionization levels, kinematics, and column densities". In : *The Astrophysical Journal* 791.2, p. 88.

Arinyo-i-Prats, A., J. Miralda-Escudé, M. Viel et R. Cen (2015). "The non-linear power spectrum of the Lyman alpha forest". In : *Journal of Cosmology and Astroparticle Physics* 2015.12, p. 017.

Delubac, T., J. E. Bautista et al. (2015). "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars". In : *Astronomy & Astrophysics* 574, A59.

Ade, P. A. et al. (2016). "Planck 2015 results-xiii. cosmological parameters". In : *Astronomy & Astrophysics* 594, A13.

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z= 2.3 with SDSS DR12 Lyα-Forests". In : *Astronomy & Astrophysics* 603, A12.

Blomqvist, M., M. M. Pieri et al. (2018). "The triply-ionized carbon forest from eBOSS : cosmological correlations with quasars in SDSS-IV DR14". In : *Journal of Cosmology and Astroparticle Physics* 2018.05, p. 029.

de Sainte Agathe, V. et al. (sept. 2019a). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : 629, A85, A85. arXiv : 1904.03400 [astro-ph.CO].

Des Bourboux, H. D. M., K. S. Dawson et al. (2019). "The extended baryon oscillation spectroscopic survey : measuring the cross-correlation between the Mg ii flux transmission field and quasars and galaxies at z= 0.59". In : *The Astrophysical Journal* 878.1, p. 47.

Guo, Z. et P. Martini (2019). "Classification of Broad Absorption Line Quasars with a Convolutional Neural Network". In : *The Astrophysical Journal* 879.2, p. 72.

Hamann, F., H. Herbst, I. Paris et D. Capellupo (2019). "On the structure and energetics of quasar broad absorption-line outflows". In : *Monthly Notices of the Royal Astronomical Society* 483.2, p. 1808-1828.

Aghanim, N. et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In : *Astronomy & Astrophysics* 641, A6.

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Ravoux, C. (2022). "One-and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman-α forest samples". Thèse de doct. Université Paris-Saclay.

Youles, S. et al. (2022). "The effect of quasar redshift errors on Lyman-α forest correlation functions". In : *Monthly Notices of the Royal Astronomical Society* 516.1, p. 421-433.

Gordon, C. et al. (2023). "3D Correlations in the Lyman-$alpha$ Forest from Early DESI Data". In : *arXiv e-prints*, arXiv-2308.

Chapitre 4

The Production of Mocks

Hydrodynamical or N-body cosmological simulations are commonly used in cosmology. They are essential for Lyα analyses to test the analysis pipeline, analytical or instrumental systematics, and estimate the covariance matrix of the correlation function. However, large simulation volumes are needed to cover tens of Gpc^3, while high resolution is also required to simulate the intergalactic medium at Jean's scale (Jeans 1902), $\sim 100h^{-1}$Kpc for the Lyα absorption. This requires large computational resources as well as accurate physical modeling. Therefore, synthetic Lyα data, the so-called Lyα mocks, are created based on Gaussian random fields (GRF). To produce quasar and Lyα forest distributions that match observations, a log-normal approximation (Coles et Jones 1991 ; Angulo et O. Hahn 2022) is applied for the quasar density field and the fluctuating Gunn-Peterson approximation (FGPA) (Gunn et B. A. Peterson 1965) is used to generate the Lyα forest optical depth (see Section 1.2.4). Synthetic quasar spectra are then simulated by applying quasar continuum and instrumental noise with the help of the `quickquasars` package (Herrera-Alcantar et al. in preparation). These mocks are not as realistic as simulations but are useful to test the Lyα analysis pipeline, the Lyα and quasar biases, the implementation of astrophysical contaminants (HCDs, BALs, and metals), the instrumental effects, the validation of the covariance matrix, as well as the BAO parameter constraints.

Early-stage mocks proposed by Font-Ribera, McDonald et Miralda-Escudé 2012 have been used for the BOSS DR11 analysis (Delubac, J. E. Bautista et al. 2015) of the Lyα auto-correlation function, where the Lyα transmission fields were generated along the lines-of-sight of quasars. However, no cross-correlation between quasars and Lyα forests was included, which was later found useful for the detection of the BAO peak (Font-Ribera, Kirkby et al. 2014). An updated approach was then proposed by J. Le Goff et al. 2011, where quasars were assigned to GRF with a log-normal probability, and thus were correlated with the associated Lyα transmissions. This approach was adapted for the auto- and cross-correlation analysis for the eBOSS DR14 analysis (Blomqvist, Des Bourboux et al. 2019 ; Sainte Agathe et al. 2019b). Furthermore, the increase in the statistical power of the eBOSS DR16 analysis (Des Bourboux, K. S. Dawson et al. 2019) required more realistic mocks and more accurate modeling of systematic effects. Based on two different approaches of adding the RSD effect into the mocks, two groups within the eBOSS collaboration have developed two types of mocks : the Saclay mocks (Etourneau et al. in preparation), using a modified FGPA with a velocity-gradient tensor and the LyaCoLoRe mocks (Farr, Font-Ribera, Des Bourboux et al. 2020), using the FGPA with gravitational linear velocities. These two versions of mocks are further used for the DESI analysis, with updated models of the instrumental effects (see Section 4.2).

During my PhD, I contributed to the generation of DESI mocks as a main co-author (Herrera-

Alcantar et al. in preparation), responsible for the implementation of astrophysical contaminants, i.e., HCDs, BALs, etc. These mocks are produced for the DESI EDR data (see Section 2.2.4), and the forecast for the DESI Y5 data. I will further present the Lyα analysis using these mocks in Section 5.1.

In this chapter, I will describe the main steps to produce mock Lyα transmissions in Section 4.1, synthetic quasar spectra in Section 4.2, and different type of mocks for eBOSS and DESI analysis in Section 4.3 and Section 4.4.

4.1 The Lyα raw mocks

I present in this Section the first step for the generation of Lyα mocks : the raw mocks, which only contain Lyα transmitted flux field (i.e., transmission fields), and no quasar continuum, instrumental noise, or astrophysical contaminants. In these mocks, quasars are inserted following a log-normal probability such that the quasar density field recovers the observed quasar bias while cross-correlating with the Lyα forests. Lyα transmissions are generated by applying the FGPA to the quasar density field, with tuned FGPA parameters to recover the measured observed Lyα bias. In order to model the redshift evolution of IGM, two different approaches are carried out in two different versions of mocks : the LyaCoLoRe mocks use simple gravitational linear velocities, and the Saclay mocks use a modified FGPA with a velocity-gradient tensor.

4.1.1 The Saclay mocks

Saclay mocks were developed for the eBOSS DR16 analysis and will be further used in the DESI analysis. They are produced with a dedicated code (publicly available code on GitHub [1]), which generates the matter density field using GRF, and constructs Lyα transmissions using log-normal approximation and the FGPA. A visualization of all these steps is presented in Figure 4.1. I will describe these steps in the following sections.

During my thesis, I mostly performed my analysis on Saclay mocks, including the continuum fitting pipeline (see Section 3.1 and Figure 3.1), the binning effect (see Section **??**), Lyα correlation functions (see Section 5.1), comparison with data (see section 4.2.2), and HCDs (see Section 6.2). I also create a series of specific mocks for HCDs with or without Lyα forests, that are particularly useful for the analysis of HCDs (see Section 4.4).

The matter density field

We first generate a simulation box with a volume of $2560 \times 2560 \times 1536$ cubic voxels [2], with each voxel side length of $L_{\rm voxel} = 2.19h^{-1}$Mpc at $z = 0$. Then Gaussian random fields (GRF) $\delta_{\rm GRF}$ are generated in each voxel, to produce the matter density field, which, in Fourier space, is defined as

$$\hat{\delta}_{\rm L}(\mathbf{k}) = \sqrt{\frac{P_m(\mathbf{k})}{V}}\hat{\delta}_{\rm GRF}(\mathbf{k}). \tag{4.1}$$

Here V is the volume of the simulation box $V = L^3_{\rm voxel}$, $P_m(\mathbf{k})$ is an input 3D matter power spectrum, obtained using the Code for Anisotropies in the Microwave Background (CAMB (Lewis, Challinor et Lasenby 2000)) with a set of fiducial parameters (Ade et al. 2016). Each Fourier mode of $\hat{\delta}_{\rm GRF}(\mathbf{k})$ is computed by taking a 3D Fourier transform of the matter density field, using the fast Fourier transform algorithm (FFT). Then a quantity $\sqrt{\frac{P_m(\mathbf{k})}{V}}$ is applied for each $\hat{\delta}_{\rm GRF}(\mathbf{k})$, to generate large-scale matter density fluctuations. The matter density field in real space $\delta_{\rm L}(\mathbf{r})$ is then the inverse Fourier transform of $\hat{\delta}_{\rm L}(\mathbf{k})$. Note that $\delta_{\rm L}(\mathbf{r})$ only includes the large-scale fluctuations (hence the subscript L) with scales large than $L_{\rm voxel}$, while small-scale astrophysical (baryonic) effects at the galaxy or sub-galaxy levels are not taken into account.

[1] `https://github.com/igmhub/SaclayMocks`
[2] Regular cubes of 3D pixels.

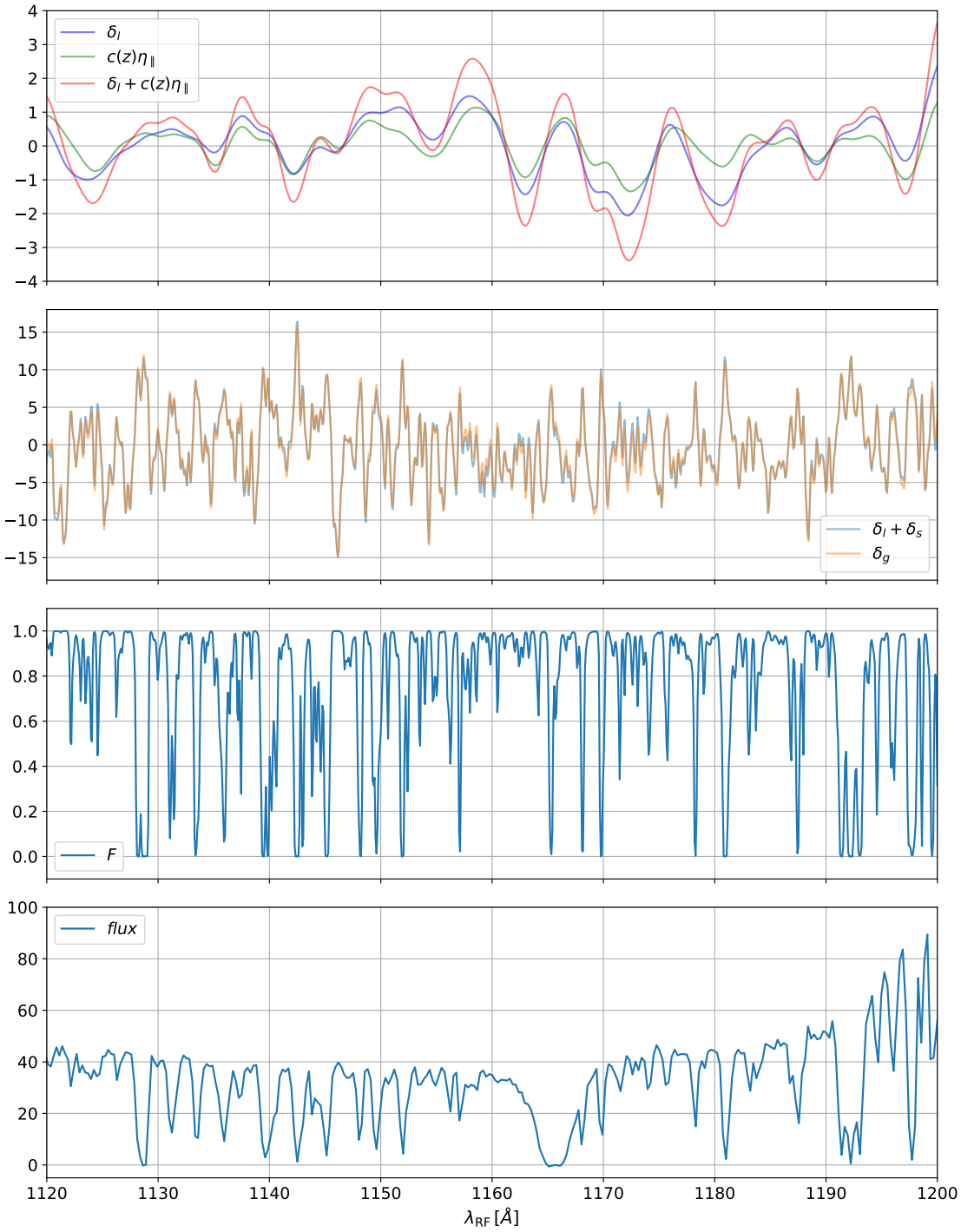

FIGURE 4.1 : Steps of raw mock production for Saclay mocks : matter density field and velocity-gradient tensor field that accounts for the RSD effect (first plot) ; the combination of the large-scale and small-scale matter density field (second plot, δ_g represents $\delta_l + \delta_s + c(z)\eta_{||}$) ; Ly$\alpha$ transmitted flux fields (third plot) and synthetic quasar spectrum (fourth plot) simulated using `quickquasars` (see Section 4.2). Credits : Saclay mock (ETOURNEAU et al. in preparation).

The quasar density field

The next step of the mock production is generating a desired quasar catalog in redshift space. Each quasar is assigned properly to the matter density field, so that mocks recover the observed quasar bias and redshift evolution, while also correlating with Lyα transmissions. For this purpose, the quasar density field is considered as a discrete biased tracer of the matter density field, and is generated by :

$$\hat{\delta}_q(\mathbf{k}) = \sqrt{\frac{P_q(\mathbf{k})}{V}}\hat{\delta}_{\mathrm{GRF}}(\mathbf{k}). \tag{4.2}$$

Here $P_q(\mathbf{k})$ is the Fourier transform of $\xi_q(\mathbf{r}, z_0)$, which is the quasar density field derived from the matter density field. In order to ensure that the quasar correlation function reproduces the observed biases, quasars are assigned to voxels following a probability proportional to the lognormal field, such that

$$\xi_q(\mathbf{r}, z) = \log\left(1 + \xi_{\mathrm{exp},q}\right)(\mathbf{r}, z), \tag{4.3}$$

where

$$\xi_{\mathrm{exp},q}(\mathbf{r}, z) = b_q^2(z)\xi_m(\mathbf{r}, z). \tag{4.4}$$

Here $\xi_{\mathrm{exp},q}(\mathbf{r}, z)$ the correlation function of the lognormal field $\exp\left(\delta_q(\mathbf{r})\right)$, since the correlation function of $\exp\left(\delta_q(\mathbf{r})\right)$ is $\xi_{\mathrm{exp},q}(\mathbf{r}, z) = \exp\left(\xi_q(\mathbf{r}, z)\right) - 1$ (Coles et Jones 1991). $\xi_m(\mathbf{r}, z)$ is the two-point correlation function of the matter density field, which is the Fourier transform of $P_m(\mathbf{k})$ at a given redshift z.

The quasar redshift evolution is implemented by applying the redshift-dependant quasar bias : $b_q(z) = 3.7(\frac{1+z}{1+2.33})^{1.7}$ (Laurent et al. 2017), with $b_{\mathrm{eff}} = 3.7$ an effective quasar bias at an effective redshift $z_{\mathrm{eff}} = 2.33$. In practice, due to computation limitation, the redshift-dependent quasar correlation function is obtained by interpolating over only three redshifted quasar density fields at $z = 1.9, 2.75$ and 3.6 (enough to produce a smooth redshift evolution).

A number $N(z)$ quasars are assigned to each voxel, to reach the target density of a desired survey. Therefore, the log-normal probability of assigning a quasar is proportional to $N(z)\exp\left(\delta_q(\mathbf{r})\right)$. Inside each voxel, the positions of quasars are uniformly distributed.

The quasar redshift-space distortions

As introduced in Section 1.2.3, the measurement of galaxy and quasar positions will suffer from systematic uncertainties due to their peculiar velocities along the lines of sight. In order to take into account this redshift-space distortion (RSD) effect of quasars in our mocks, redshift-dependant velocity fields need to be applied to the quasar density fields. The positions of quasars in real space are defined as $\vec{X}(z)$ at redshift z, the shifted measured distance $\Delta X = ||\Delta\vec{X}||$ from the observer along the line of sight is

$$\Delta X = \frac{1+z}{H(z)}v_{||}(z) = \frac{f(z)}{f_0 H_0}v_{||}(z_0), \tag{4.5}$$

where $H(z)$ is the Hubble parameter (introduced in Section 1.1), $f(z) = \frac{dlnG}{dlna}$ is the linear growth rate with $a = \frac{1}{1+z}$ the universe scale factor and G the growth factor. $v_{||}$ is the peculiar velocity of quasars along each line-of-sight. Subscripts 0 denote quantities at $z = 0$. The Fourier transform of the velocity fields can be expressed in terms of the matter density field (Dodelson et F. Schmidt 2020) :

$$\hat{v}_j(\vec{k}) = \frac{if(z)H(z)}{1+z}\frac{k_j}{k^2}\hat{\delta}_L(\vec{k}). \tag{4.6}$$

The peculiar velocities can thus be computed by projecting the velocity field along the line-of-sight :

$$v_{||} = \vec{u} \cdot \vec{v}(\mathbf{r}), \tag{4.7}$$

with $\vec{u} = \frac{\vec{X}}{||\vec{X}||}$ being the unit vector along a given line-of-sight.

The associated velocity-gradient fluctuation within linear approximation in $\vec{k}$-space is

$$\hat{\eta}_{pq}(\vec{k}) = f \frac{k_p k_q}{k^2} \hat{\delta}_L(\vec{k}), \tag{4.8}$$

The line-of-sight velocity gradient is then $\eta_{||} = u_p u_q \eta^{pq}$.

The Lyα transmitted flux fields

As mentioned in previous sections, the quasar density fields are generated using GRF at each simulation voxel. In order to produce lines-of-sight in the optical wavelength range (for DESI $\lambda \in [3476.11, 5591.566]$Å), the density fields are interpolated at each pixel (6524 pixels of $0.2h^{-1}$Mpc). Gaussian smoothing is applied at each pixel to avoid discontinuities with neighboring pixels. This smoothing removes the small-scale effects with $k > k_s = \frac{\pi}{L_{\text{voxel}}}$, and adds additional correlations to the Lyα correlation function. However, this smoothing effect can be modeled analytically, and its uncertainty is much smaller than other observational uncertainties (see Etourneau et al. in preparation for more details). Since the small-scale fluctuations are smoothed, an additional small-scale field needs to be introduced. It is obtained from an input 1D flux power spectrum, which is an integration of the 3D flux power spectrum :

$$P^{1D}(k_{||}) = \frac{1}{2\pi} \int_0^\infty P^F(k_{||}, k_\perp) k_\perp dk_\perp. \tag{4.9}$$

The small-scale matter density field is thus determined as :

$$\hat{\delta}_S(z, \vec{k}) = \sqrt{\frac{P^{1D}(z, k_{||})}{L^3_{\text{pixel}}}} \hat{\delta}_{\text{GRF}}(\vec{k}). \tag{4.10}$$

Here $P^{1D}(z, k_{||})$ is an input 1D flux power spectrum at a certain redshift z, which is tuned to reproduce the 1D flux power spectrum from observations (Chabanier, Palanque-Delabrouille et al. 2019), L_{pixel} refers to the pixel size in the measurement. The total matter density field is then composed of both the large-scale and the small-scale fields :

$$\delta_{\text{m}}(z) = \delta_{\text{L}}(z) + \delta_{\text{S}}(z). \tag{4.11}$$

The Lyα absorption fields F (the so-called Lyα transmitted flux fields), are constructed by taking an exponential of the optical depth :

$$F(z) = \exp\left(-\tau(z)\right), \tag{4.12}$$

where the Lyα optical depth $\tau(z)$ is determined by modifying the Fluctuating Gunn-Peterson Approximation (FGPA, an approximation neglecting the thermal broadening and peculiar velocity fluctuations of neutral hydrogen atoms) (Gunn et B. A. Peterson 1965), from $\tau(z) =$

$a_{\rm GP}(z)\exp\left(b_{\rm GP}(z)G(z)\delta_{\rm m}\right)$ to :

$$\tau(z) = a_{\rm GP}(z)\exp\left(b_{\rm GP}(z)G(z)(\delta_{\rm m} + c_{\rm GP}(z)\eta_{||})\right). \tag{4.13}$$

This model has three free paramaters : $a_{\rm GP}$ is related to the Lyα bias $b_{{\rm Ly}\alpha}$, $c_{\rm GP}$ controls the RSD parameter $\beta_{{\rm Ly}\alpha}$, and $b_{\rm GP}$ accounts for the redshift dependence of Lyα forests. $\eta_{||}$ is the velocity gradient along the line-of-sight related to the RSD effect. It is computed as the projection of the velocity-gradient tensor field along a given line-of-sight [3] :

$$\eta_{||} = u_i u_j \eta^{ij}, \tag{4.14}$$

where in Fourier space :

$$\hat{\eta}^{ij}(\mathbf{k}) = \frac{k^i k^j}{k^2} f \hat{\delta}_L(\mathbf{k}). \tag{4.15}$$

Given Equation 4.12, we can then derive the expression of the Lyα transmitted flux field as

$$F = \exp\left(-a_{\rm GP}(z)\exp\left(b_{\rm GP}(z)G(z)(\delta_{\rm L}(z) + \delta_{\rm S}(z) + c_{\rm GP}(z)\eta_{||})\right)\right). \tag{4.16}$$

$a_{\rm GP}$, $b_{{\rm Ly}\alpha}$, and $c_{\rm GP}$ are tuned to recover their observed values (Palanque-Delabrouille et al. 2013) : $b_{\rm GP}$ can be predicted by considering the equilibrium of photo-ionization of HI and the recombination of electrons and protons (Etourneau et al. in preparation). In this scenario, the redshift evolution of IGM with an equation of state $(1+\delta)^{\gamma(z)-1}$ gives $b_{\rm GP} = 2-0.7(\gamma(z)-1) = 1.58$ at $z \sim 3$ (with $\gamma(z=3) = 1.6$ (Hui et Gnedin 1997)).

4.1.2 The LyaCoLoRe mocks

The LyaCoLoRe mocks are generated following a similar procedure as the Saclay mocks. Firstly, the CoLoRe [4] (Ramirez-Pérez, Sanchez, Alonso et Font-Ribera 2022) package is used to create the quasar catalog and Gaussian field skewers, where a set of Gaussian random fields are produced and quasars are sampled into these fields following a lognormal transformation with an input number density and bias. The Gaussian field skewers at different redshifts are then generated by taking an interpolation of the Gaussian field and an associated radial velocity field. The LyaCoLoRe [5] package takes these Gaussian field skewers as input and adds small-scale fluctuations by considering a desired 1D flux power spectrum (McDonald, Seljak, Burles et al. 2006), defined as

$$P_{\rm 1D}(k) \propto (1 + (\frac{k}{k_0})^n)^{-1}, \tag{4.17}$$

where k_0 and n are two free parameters tuned to achieve the observed 1D power spectrum (McDonald, Seljak, Burles et al. 2006). The final skewers are derived as

$$\delta_F(z) = \delta_C + \sigma_{\rm S}(z)\delta_{\rm S}, \tag{4.18}$$

where δ_C is the Gaussian skewers taken from CoLoRe package, $\delta_{\rm S}$ stands for the small-scale fluctuations, and $\sigma_{\rm S}(z)$ is another free parameter to account for redshift evolution.

The Lyα optical depth is further generated by using the FGPA approximation and considering the RSD effect due to gravitational linear velocities in the IGM at different redshifts. The flux transmission skewers are converted from optical depth by taking an exponential : $F = \exp(-\tau)$.

[3] Here we use the Einstein notation to represent the projection along each dimension (Einstein 1916).
[4] `https://github.com/damonge/CoLoRe`
[5] `https://github.com/igmhub/LyaCoLoRe`

The main difference between the Saclay mocks and the LyaCoLoRe mocks is at the step of generating the optical depth : the Saclay mocks use a modified FGPA with a velocity-gradient tensor to account for the RSD effect, while the LyaCoLoRe mocks use the FGPA with gravitational linear velocities, which is simpler but less realistic.

4.2 The Lyα synthetic spectra

The raw mocks generated in the previous section contain only Lyα transmitted flux fields. They are useful to test the Lyα analysis pipeline by studying their correlation function. However, in real observation, we measure quasar spectra, and estimate Lyα transmissions by taking their absorption fluctuations with respect to the quasar unabsorbed continuum, fitted using the pipeline introduced in Section 3.2. In this continuum fitting procedure, various systematic effects will bias the measurement and need to be understood, such as the distortion matrix, astrophysical contaminants, and instrumental effects (see Section 3.2 for all these effects). Therefore, synthetic quasar spectra need to be simulated in our mocks, as precisely as real observed spectra, to validate our analysis pipeline.

For this purpose, a simulation package named `quickquasars` [6] (Herrera-Alcantar et al. in preparation) in the `desisim` [7] simulation package is used to generate synthetic quasar spectra based on the Lyα transmission skewers in raw mocks. It includes several steps to simulate a quasar spectrum, namely : a quasar continuum, astrophysical contaminants, and experimental effects such as instrument response and observing conditions. I describe these steps in the following sub-sections.

4.2.1 The quasar continuum

The quasar continuum is generated by using one of the two possible templates : `Simqso` (default) or `PCA-qso`.

Simqso

The `Simqso` software in the `desisim` package produces the quasar continuum from a broken power-law, and adds each Gaussian emission line with defined observed wavelength, width, and dispersion (defined using spectra from Harris et al. 2016). Quasars are produced in the redshift range $z \in [2.1, 3.5]$ and with rest-frame wavelength range $\lambda_{\rm rf} \in [800, 3300]$ Å, while BAL or DLA quasars are excluded. Note that the emission line parameters are adjusted differently for eBOSS DR16 mocks (Des Bourboux, Rich et al. 2020) and DESI mocks (Herrera-Alcantar et al. in preparation) to match their desired mean continuum. Figure 4.2 shows the good agreement of emission features between quasar spectra from DESI mocks and DESI EDR data (described in Section 2.2.4).

PCA-qso

These templates are generated from a principal component analysis (PCA) decomposition of quasar spectra collected from SDSS DR7 ($z \in [0.4, 2]$) and BOSS DR10 ($z \in [2, 4]$). A random sample of eigenvalues and eigenvectors of PCA templates are first generated, then a parametrization procedure is performed to match the desired DESI spectra. A detailed description of this process can be found in Herrera-Alcantar et al. in preparation.

4.2.2 Astrophysical contaminants

Different astrophysical contaminants, e.g., BALs, HCDs, and metals are inserted into the quasar synthetic spectra using the following approaches :

[6] `https://github.com/desihub/desisim/blob/main/py/desisim/scripts/quickquasars.py`
[7] `https://github.com/desihub/desisim`

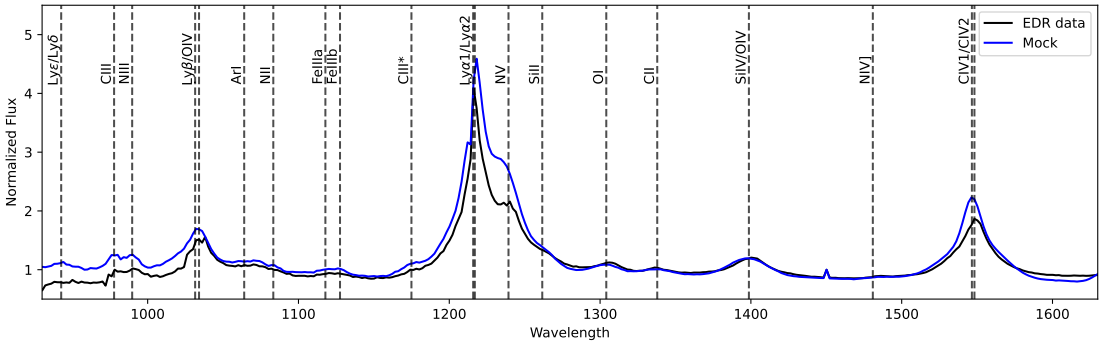

FIGURE 4.2 : Comparison of an observed quasar spectrum from DESI EDR data and a synthetic spectrum from DESI mocks produced using `quickquasars`. The mock spectrum reproduces well the emission features seen in DESI data. Credits : DESI EDR mock (HERRERA-ALCANTAR et al. in preparation).

Broad Absorption Lines systems

Within the DESI collaboration, I am responsible for the insertion of BALs (Section 3.2.2) into our mocks. BALs are simulated using 1500 templates created from the SDSS DR16 BAL catalog (GUO et MARTINI 2019) with 53, 760 BALs from 320, 821 quasars. We randomly select simulated quasars in mocks with the same BAL quasar ratio as in the data ($\sim 13\%$, see Section 3.2.2), and apply BAL corrections by multiplying the quasar continuum F_{CONT} with the BAL continuum F_{BAL} (fitted using DR14 quasars with Principal Component Analysis, see GUO et MARTINI 2019). A comparison of BAL features (used to determine BALs using C IV and Si IV absorption regions, see Section 3.2.2), the Balnicity Index (BI ,R. J. WEYMANN, S. L. MORRIS, C. B. FOLTZ et P. C. HEWETT 1991) and the Intrinsic Absorption index (AI, P. B. HALL et al. 2002), is shown in Figure 4.3. It compares BALs from DESI EDR mocks with those detected from DESI EDR data using a CNN algorithm (GUO et MARTINI 2019) with $\mathrm{BI_{CIV}} > 0$ and $\mathrm{AI_{CIV}} > 0$ (the subscript C IV denotes the metal line C IV). This comparison shows that our mocks reproduce well the distribution of BALs of the DESI data.

High Columns Density systems

I am also responsible for assigning HCDs (see Section 3.2.2) into our mocks. HCDs are usually modeled by Voigt profiles (see Section 6.2.3) at the Lyα optical depth level. This means that we should add HCDs in the Lyα transmissions of mocks, rather than add them in the quasar continuum. Moreover, this operation ensures the production of the correlations of HCDs and Lyα forests. HCDs are inserted into mocks following a given redshift distribution and a probability density distribution of HI column densities N_{HI}, hereafter denoted as n, computed using the IGM physics package `pyigm` [8] (J. PROCHASKA, TEJOS, WOTTA et al. 2017 ; J. X. PROCHASKA, MADAU, O'MEARA et FUMAGALLI 2014), which was calibrated by fitting a set of detected HCDs in the literature (J. X. PROCHASKA, MADAU, O'MEARA et FUMAGALLI 2014). The probability distribution of n is defined as (R. F. CARSWELL, MORTON, M. G. SMITH, STOCKTON, TURNSHEK

[8] https ://github.com/pyigm/pyigm

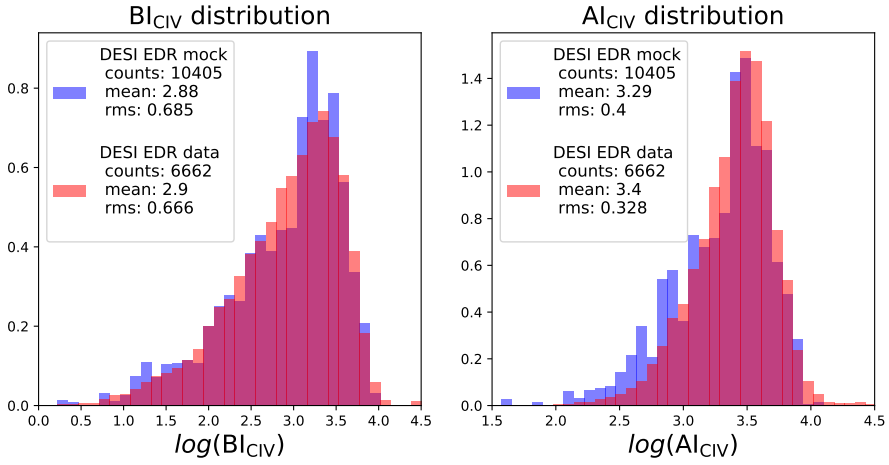

FIGURE 4.3 : Normalized distribution of BI$_{CIV}$ and AI$_{CIV}$ from DESI EDR data using CNN (red) and DESI EDR mock (blue). Credits : DESI EDR mock (HERRERA-ALCANTAR et al. in preparation).

et R. J. WEYMANN 1984; RUDIE, STEIDEL, SHAPLEY et PETTINI 2013)

$$f(n,X)dndX = \frac{m}{\Delta n \Delta X} dndX, \tag{4.19}$$

where m is the number of HCDs with column densities in the range Δn, and with the comoving path lengths (J. N. BAHCALL et PEEBLES 1969) along the line-of-sight in the range ΔX. The comoving path length is obtained from (in a flat universe)

$$\Delta X = \int_{z_{\rm min}}^{z_{\rm max}} \frac{H_0}{H(z)}(1+z)^2 dz = \int_{z_{\rm min}}^{z_{\rm max}} \frac{(1+z)^2}{\sqrt{\Omega_\Lambda + \Omega_{\rm m}(1+z)^3}} dz. \tag{4.20}$$

The locations of HCDs are determined by choosing the peaks of the Gaussian density fields above a given threshold. This threshold is tuned to get the desired bias for HCDs, $b_{\rm HCD}(z)$. Usually, this value is chosen to be $b_{\rm HCD}(z) = 2$, which is measured by the cross-correlations between DLAs and Lyα forests (PÉREZ-RÀFOLS, MIRALDA-ESCUDÉ, ARINYO-I-PRATS, FONT-RIBERA et MAS-RIBAS 2018). Figure 4.4 shows the comparison of the eBOSS DR16 mocks and DESI EDR mocks for the distribution of $N_{\rm HI}$ and $z_{\rm DLA}$. There is no difference in the input distribution of $N_{\rm HI}$. However, DESI EDR mocks are produced with a higher range and means for $z_{\rm DLA}$ and $z_{\rm QSO}$, because of the observation of higher redshift quasars in DESI data.

Metals

Different redshifted metal absorption lines (Section 3.2.2), such as Si II λ1260, Si III λ1207, Si II λ1193, and Si II λ1190, will overlap with Lyα forests and thus bias the measured Lyα correlation function. These metal contaminants are inserted into our mocks by re-scaling the Lyα optical depth τ with a set of characteristic coefficients A_m for each metal m individually, which gives the metal optical depth $\tau_m = A_m \tau$. These coefficients represent the strength of these metal correlations compared to the Lyα correlations, and can be adjusted by tuning the metal biases in mocks to what is measured in observational data. Figure 4.5 shows the tuned metal biases b_η

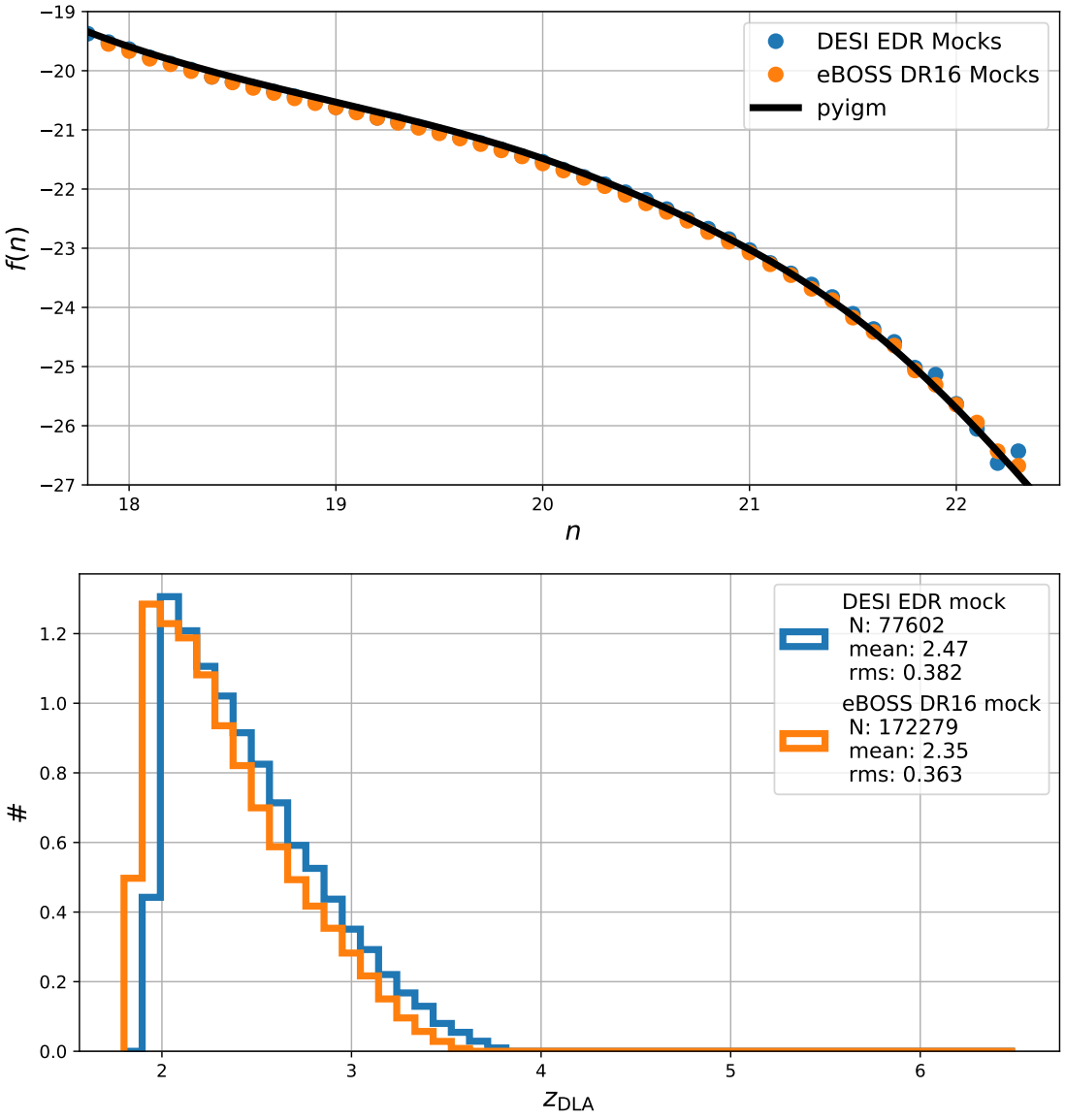

FIGURE 4.4 : The upper plot shows the probability distribution of $f(n)$, and the lower plot shows the histogram of z_{DLA}. A comparison is made for the eBOSS DR16 mocks and DESI EDR mocks. The black curve gives the input distribution into the mocks given by the `pyigm` package.

for both eBOSS mocks and DESI EDR mocks (these values were tuned to match the values from eBOSS DR14 data (DE SAINTE AGATHE et al. 2019a)), compared with the measurements from observational data, i.e., eBOSS DR14 data, eBOSS DR16 data, and DESI EDR data.

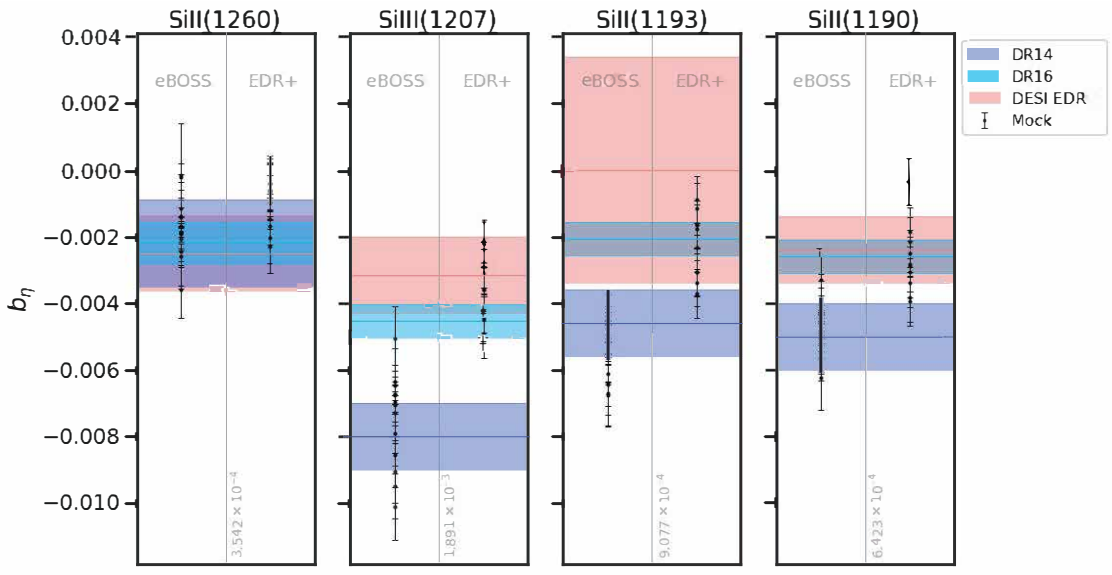

FIGURE 4.5 : The measurement of metal biases b_η in the Lyα auto-correlation function. Colored bands show the results from eBOSS DR14 data, eBOSS DR16 data, and DESI EDR data. The black dots show the results of a stack of 10 mocks for eBOSS (left panels) and DESI EDR (right panels). The values of the characteristic metal coefficients A_m used to generate the mocks are presented at the bottom, and are tuned to match the values from eBOSS DR14 data. Credits : DESI EDR mock (HERRERA-ALCANTAR et al. in preparation).

4.3 Mocks for eBOSS/DESI surveys

In this section, I describe the experimental conditions, and the survey settings introduced in the previous mocks to simulate a cosmology survey. For the eBOSS analysis, I work on the generation of mocks built for the eBOSS DR16 data. For DESI, two types of mocks are made : the DESI EDR mocks are used for the analysis of the DESI EDR data (only include the survey validation data and two months of main survey data, see Section 2.2.4) ; the DESI Y5 (the entire DESI five years data) mocks are made to cover the entire DESI footprint and quasar density.

4.3.1 Experimental effects

We use the `specsim` package implemented in `quickquasars` to simulate synthetic quasar spectra with good agreement on the real instrumental effects. In this section I will describe these different effects in our mock production.

Spectrograph

As described in Section 2.1.2 and Section 2.2.3, each of the ten DESI spectrographs has three arms (g, r, and z) to cover blue, red, and near-infrared bands, while the instrument for BOSS/eBOSS only used two bands (B and R). Nevertheless, for most Lyα quasars and forests, spectrograph B is enough to cover the useful wavelength range. The pixel binning of wavelengths used for eBOSS and DESI is different : a logarithm grid of pixels was used in eBOSS, whereas DESI uses a linear grid of pixels. Small discrepancies appear when changing the pixel grid from logarithm to linear, which is detailed in Ignasi et al. in preparation.

Exposure time

The typical exposure time of each observation is fixed as 2,000s for eBOSS mocks, to be consistent with data. For DESI, all objects are expected to have 4,000s of effective exposure time, which is twice longer than that of eBOSS. We therefore assign each DESI quasar with 4,000s exposures. However, if the simulated quasars are not expected to have the same exposure time, then for each of the HEALPIX pixels of the footprint, we assign a multiple of 1000s for each quasar with respect to the distribution exposure time distribution.

Magnitude distribution

We randomly sample the transmission skewers of the raw mocks, and assign magnitudes for each band following a given magnitude distribution. For DESI EDR mocks as an example, the desired magnitude distribution is described in Chaussidon et al. 2023, and is shown in Figure 4.6. This comparison shows a good agreement between the magnitude distributions of our mocks and DESI data.

Source

The source flux is characterized by its spectral energy distribution, modeled as a point source profile with a quasar continuum. We use the `specsim` package to convert an input spectral energy distribution (SED) into arrays of expected mean detected flux with variances for each arm of the spectrograph. Then the source templates are created by setting a very large exposure time to get rid of noise (typically 2,000,000s).

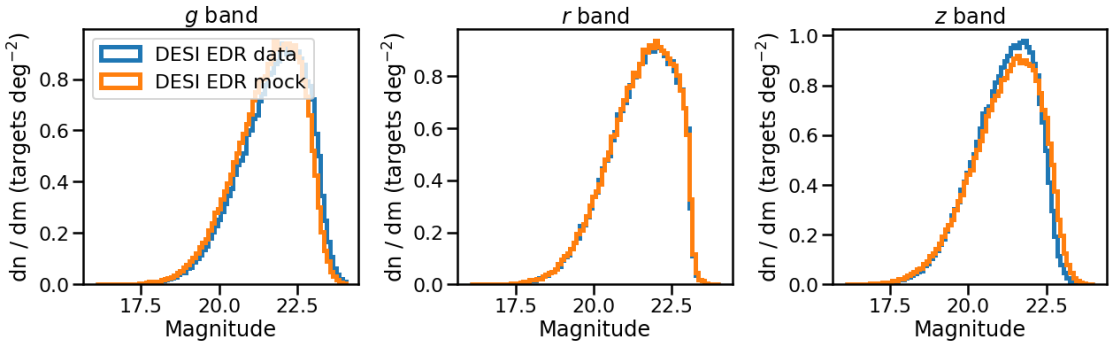

FIGURE 4.6 : The comparison of magnitude distributions of three bands g, r, z, for DESI EDR mocks (orange) and DESI EDR data (blue).

Atmosphere

For DESI, we use a point spread function (PSF) to model the observed spectral flux of a point source at the telescope, which is convolved with the source profile. The effect of the atmosphere, is modeled by a fixed Moffat model (MOFFAT 1969) :

$$I(r) = I_0(1 + \frac{r}{\alpha})^{\beta}. \tag{4.21}$$

Here $\beta = 3.5$ and $\alpha = \frac{\text{FWHM}}{2\sqrt{2^{1/\beta}-1}}$ (ABARESHI et al. 2022), where FWHM is computed as twice the radius at which $I(r) = 0.5I_0$.

The instrument model

A set of instrumental parameters, listed in Table 4.1, is used as an input for the instrumental model, implemented in the `specsim` package of `quickquasars`. These parameters are used to characterize different instrumental effects such as the readout noise, and the gain of CCDs for each of the spectrograph cameras, the collected flux, fiber loss, observing conditions, etc. As a result, an example of a simulated quasar spectrum is shown in Figure 4.7, where one can see separately the quasar continuum without any Lyα absorption fluctuations, the noiseless spectrum, and the spectrum with all the instrumental effects taken into account.

4.3.2 The survey settings

Based on the instrumental models described in the previous section, I will now provide a description of the survey settings in order to simulate the eBOSS/DESI surveys.

Redshift distribution

As described in Section 4.1.1, we assign quasars into the mocks following a lognormal probability field, while ensuring the redshift distribution from the IGM physics package `pyigm` [9] (J. PROCHASKA, TEJOS, WOTTA et al. 2017 ; J. X. PROCHASKA, MADAU, O'MEARA et FUMAGALLI 2014). Figure 4.8 shows the good agreements between the redshift distributions used in our

[9] https ://github.com/pyigm/pyigm

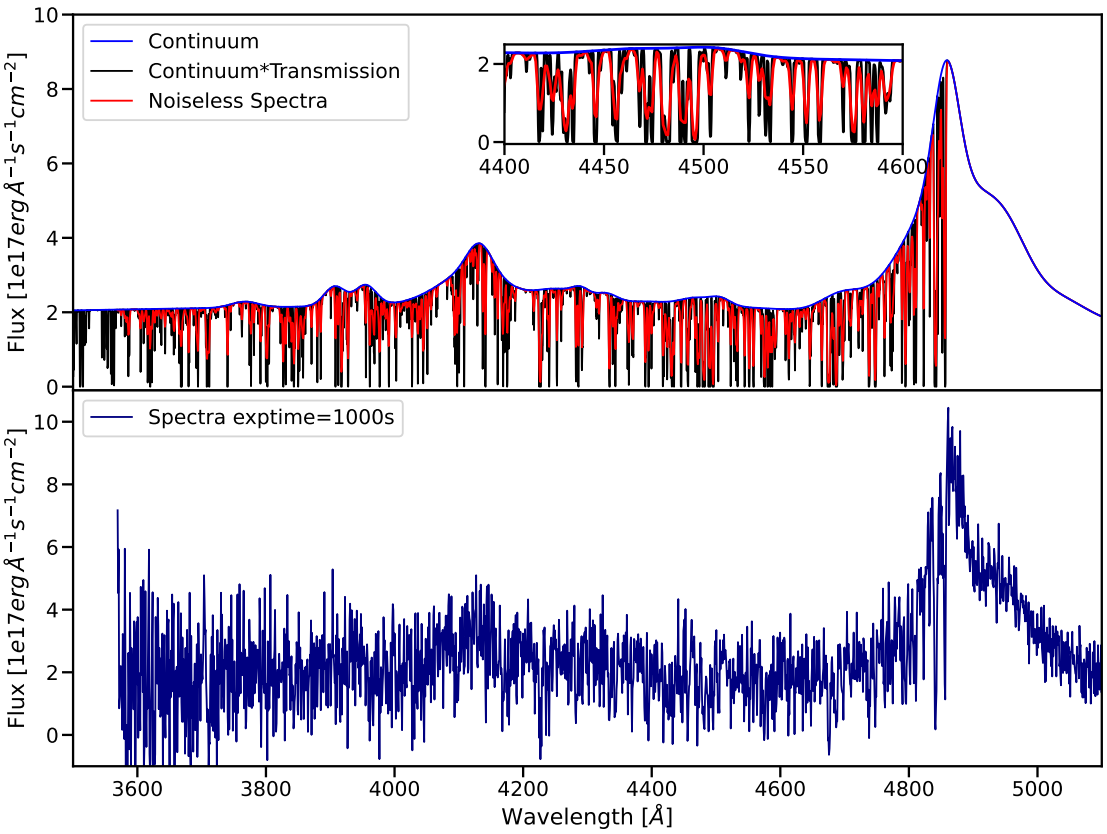

FIGURE 4.7 : Simulated quasar spectrum using `quickquasars`. The upper plot shows the quasar continuum (blue), the continuum with Lyα absorptions (black), quasar spectrum with spectrograph resolution added (red, noiseless), while the lower plot shows the spectrum with noises added and an effective exposure time of 1000s. Credits : DESI EDR mock (HERRERA-ALCANTAR et al. in preparation).

Parameter	Value
Instrument : DESI	
Primary mirror Diameter	3.797 m
Fiber diameter	107.0 um
Field radius	414.0 mm
Fiber loss method	fastsim
Cameras : b,r,z	
Read noise	$3.29, 3.69, 3.69$ electron/pixel^2
Dark current	$1.89, 1.14, 1.14$ electron/(hour pixel^2)
Gain	$1.0, 1.0, 1.0, 1.0$ electron/adu
Observatory : KPNO	
Nominal exposure time	1000s
Temperature	15 ^{o}C
Relative humidity	0

TABLEAU 4.1 : Input parameters for `specsim` used to model the instrumental effects of the telescope. Credits : DESI EDR mock (HERRERA-ALCANTAR et al. in preparation).

mocks compared with their relevant data. One can also tell from the plots that we use a redshift distribution in DESI slightly different ($z \sim 2.25$) from what was used in eBOSS.

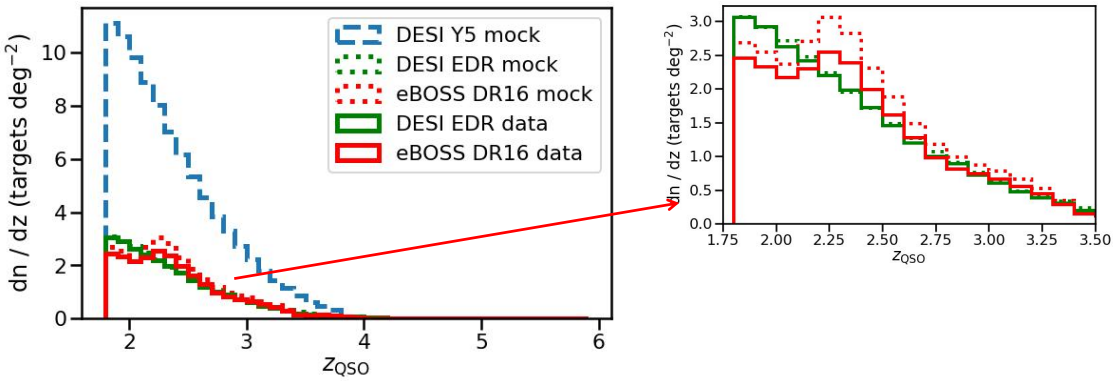

FIGURE 4.8 : Comparison of the redshift distributions of quasars, for DESI EDR mocks (dashed green), DESI EDR data (solid green), eBOSS DR16 mocks (dashed red), eBOSS DR16 data (solid red). The forecasted distribution for DESI Y5 mocks (dashed blue) is also shown as a comparison.

Footprint for eBOSS mocks

The eBOSS DR16 mocks that I work on are produced to cover the 10,000 deg^2 of the eBOSS survey footprint, and an average quasar density of around 25 deg^{-2}. Figure 4.9 presents a comparison of the footprints of eBOSS DR16 mocks and DESI Y5 mocks. The DR16 mocks contain a catalog of 261854 quasars, 232544 forests, and 91659 HCDs with $z > 2$. The cosmological parameter values used to generate these mocks are $\{\Omega_m = 0.31457, \Omega_\Lambda = 0.68543, \Omega_k = 0\}$, taken from (ADE et al. 2016).

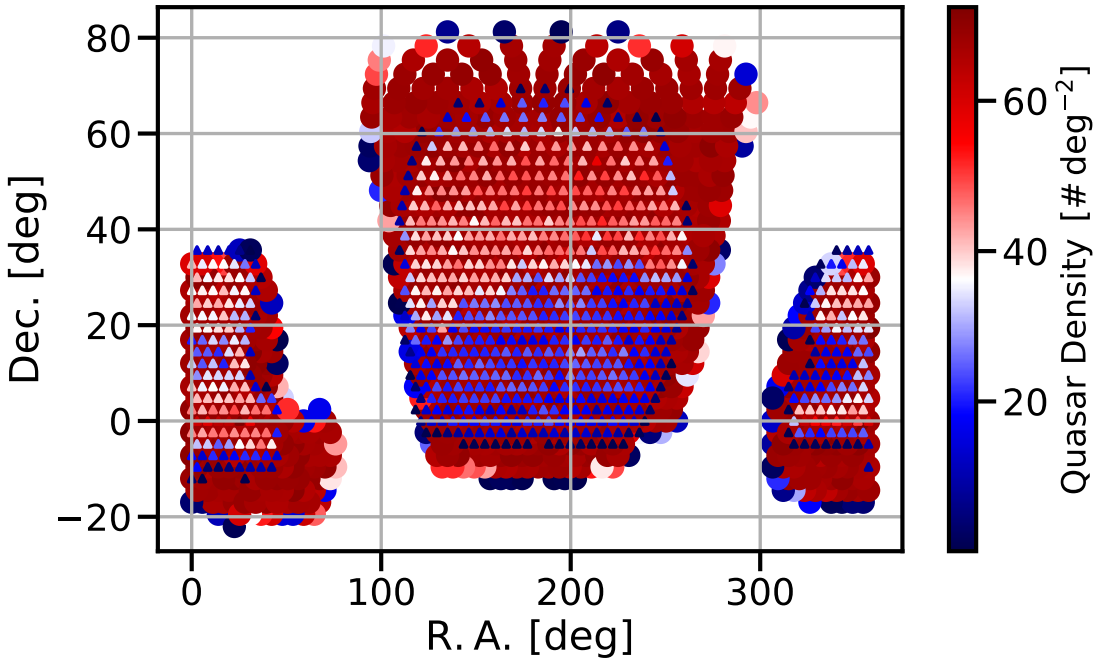

FIGURE 4.9 : Comparison of survey footprint and Lyα quasar density ($z > 1.8$) for eBOSS DR16 mocks (10,000 deg^2, filled triangles) and DESI Y5 mocks (14,000 deg^2, filled circles). The NGC covering 9,900 deg^2 is shown on the left, and the SGC covering 4,400 deg^2 is shown on the right. The plots are created using `HEALpix` pixels with nside=16 (the number of pixels on each side).

Footprint for DESI mocks

Figure 4.10 shows the footprint of DESI EDR mocks, compared with the nominal DESI footprint. We can tell from Figure 4.9 and Figure 4.10 that the average Lyα quasar density ($z > 1.8$) of the DESI EDR mocks is at the same level as the eBOSS DR16 mocks, ~ 25 deg^{-2}, since the DESI EDR data only covers a small amount of DESI data, and without co-adding exposures. The DESI Y5 mocks are constructed to match the footprints of the entire DESI survey, with the forecasted average Lyα quasar density at around 100 deg^{-2}. The fiducial cosmological parameter values used for DESI mocks are taken from Aghanim et al. 2020.

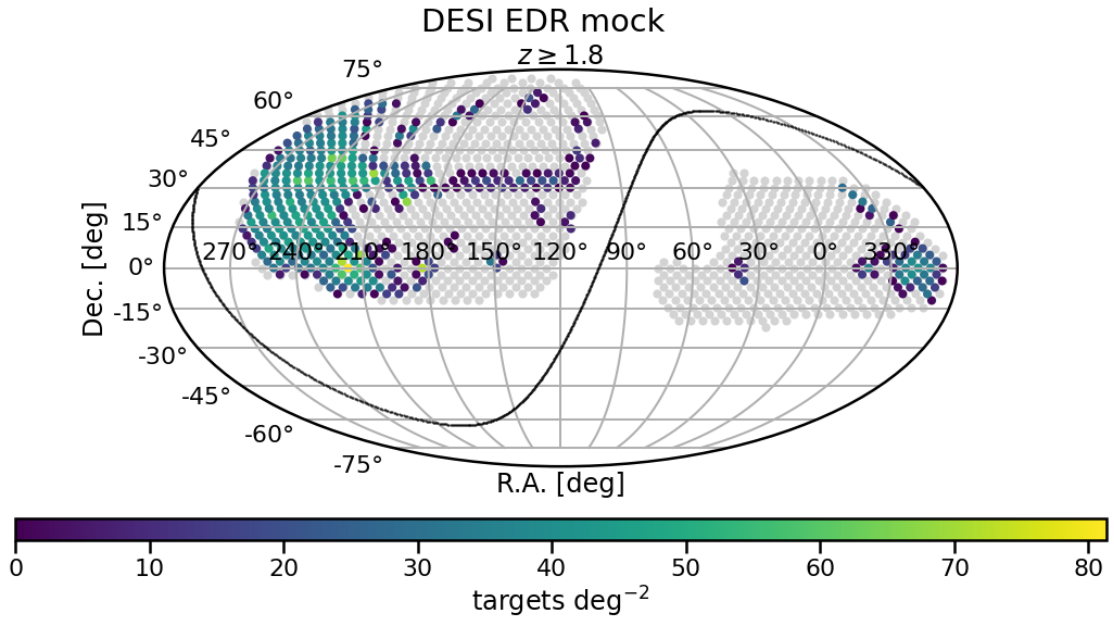

Figure 4.10 : The comparison of survey footprint and Lyα quasar density ($z > 1.8$) for DESI EDR mocks (light blue) and DESI data (14,000 deg^2, shaded gray). The plots are made using `HEALPIX` pixels with nside=16.

4.4 Mocks for eBOSS/DESI analysis

Based on the procedure described in the previous sections, different types of Lyα mocks were produced to validate the analysis pipeline and various systematic effects used to analyse real data, e.g. the continuum fitting, the addition of HCDs, the addition of metals, etc. I hereby define these different mocks :

- $eboss-raw$ mocks : as was described in Section 4.1.1, $eboss-raw$ mocks contain only Lyα transmission skewers, without simulations of quasar spectra. These mocks are particularly helpful for the validation of Lyα biases, continuum fitting, distortion matrix, etc.
- $eboss-0.0$ mocks : these mocks were produced by adding synthetic quasar continuum and noise to the raw mocks.
- $eboss-0.2$ mocks : these mocks were produced by adding HCDs with HI column densities ($N_{\rm HI}$, hereafter n) following a realistic probability distribution $f(n)$, using the method described in Section 4.2. $f(n)$ is obtained from the `pyigm` package, as is shown in Figure 4.4.
- $eboss-0.2+$ mocks : in order to study the systematic effect of HCDs in the Lyα analysis and better understand the modeling of this effect, I create several stacks of 10 Saclay mocks with or without HCDs. Specifically, all the HCDs hold the same n, where $log(n) = 19.5, 20.0, 20.5, 21 {\rm cm}^{-2}$.
- $eboss-0.3$ mocks : add metals following the method described in Section 4.2 on top of $eboss-0.2$ mocks.
- $eboss-0.4$ mocks : based on $eboss-0.2+$ mocks, I also generate mocks without Lyα forests, with only HCDs.

Figure 4.11 shows a comparison of the simulated quasar spectra for these different scenarios. The mocks with no Lyα forests and no HCDs are produced with only the continuum and noise, while those with HCDs have the same $log(n) = 21$ cm^{-2}.

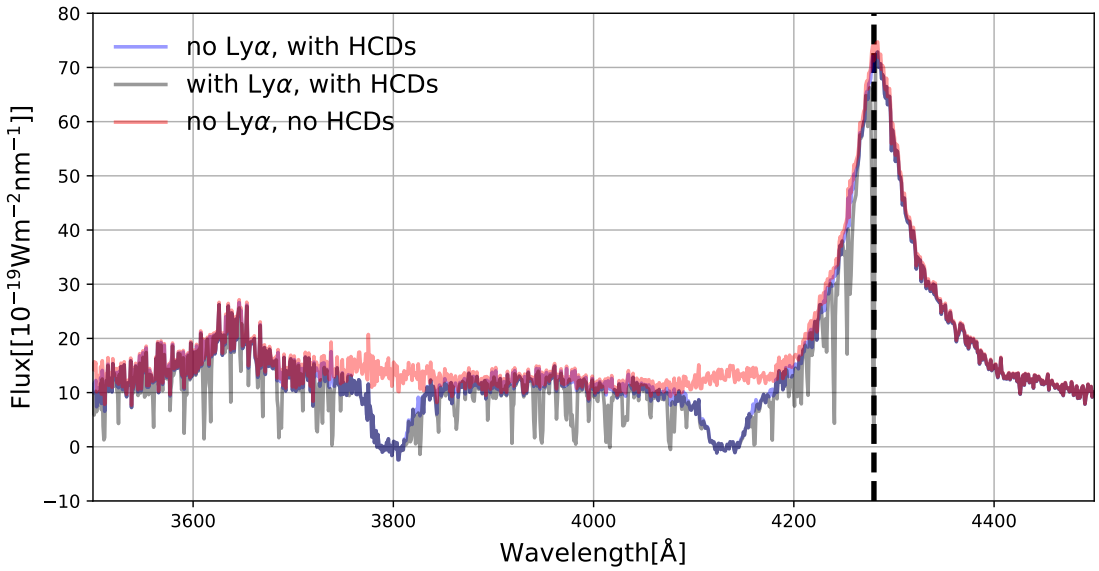

FIGURE 4.11 : A quasar spectrum at the redshift $z = 2.52$, taken from eBOSS DR16 mocks for different scenarios : no Lyα forests and no HCDs (red), no Lyα forests and with HCDs (blue), with Lyα forests and with HCDs (black). The Lyα peak is shown as the black dashed line.

4.4.1 Summary and prospects

In this chapter, I described the main steps to produce simulated Lyα transmissions, and synthetic quasar spectra for raw mocks. Then I described the method to implement astrophysical contaminants (HCDs, BALs, etc) and instrumental effects to make raw mocks more realistic. Moreover, I also presented different versions of mocks (different astrophysical contaminants) for the eBOSS and DESI surveys.

In the next chapter, I will describe the Lyα analyses based on these mocks, and compare with those for real data (eBOSS and DESI data).

Bibliographie du présent chapitre

JEANS, J. H. (1902). "I. The stability of a spherical nebula". In : *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 199.312-320, p. 1-53.

EINSTEIN, A. (1916). "The foundation of the general theory of relativity." In : *Annalen Phys.* 49.7. Sous la dir. de J.-P. HSU et D. FINE, p. 769-822.

GUNN, J. E. et B. A. PETERSON (nov. 1965). "On the Density of Neutral Hydrogen in Intergalactic Space." In : 142, p. 1633-1636.

BAHCALL, J. N. et P. PEEBLES (1969). "Statistical tests for the origin of absorption lines observed in quasi-stellar sources". In : *The Astrophysical Journal* 156, p. L7.

MOFFAT, A. (1969). "A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry". In : *Astronomy and Astrophysics, Vol. 3, p. 455 (1969)* 3, p. 455.

Carswell, R. F., D. C. Morton, M. G. Smith, A. N. Stockton, D. A. Turnshek et R. J. Weymann (1984). "The absorption line profiles in Q1101-264". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 278, March 15, 1984, p. 486-498. Research supported by the Science and Engineering Research Council and Radcliffe Trust.* 278, p. 486-498.

Coles, P. et B. Jones (1991). "A lognormal model for the cosmological mass distribution". In : *Monthly Notices of the Royal Astronomical Society* 248.1, p. 1-13.

Weymann, R. J., S. L. Morris, C. B. Foltz et P. C. Hewett (1991). "Comparisons of the emission-line and continuum properties of broad absorption line and normal quasi-stellar objects". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 373, May 20, 1991, p. 23-53.* 373, p. 23-53.

Hui, L. et N. Y. Gnedin (1997). "Equation of state of the photoionized intergalactic medium". In : *Monthly Notices of the Royal Astronomical Society* 292.1, p. 27-42.

Lewis, A., A. Challinor et A. Lasenby (2000). "Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models". In : *The Astrophysical Journal* 538.2, p. 473.

Hall, P. B. et al. (2002). "Unusual broad absorption line quasars from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 141.2, p. 267.

McDonald, P., U. Seljak, S. Burles et al. (2006). "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 163.1, p. 80.

Le Goff, J. et al. (2011). "Simulations of BAO reconstruction with a quasar Ly-α survey". In : *Astronomy & Astrophysics* 534, A135.

Font-Ribera, A., P. McDonald et J. Miralda-Escudé (2012). "Generating mock data sets for large-scale Lyman-α forest correlation measurements". In : *Journal of Cosmology and Astroparticle Physics* 2012.01, p. 001.

Palanque-Delabrouille, N. et al. (2013). "The one-dimensional Lyα forest power spectrum from BOSS". In : *Astronomy & Astrophysics* 559, A85.

Rudie, G. C., C. C. Steidel, A. E. Shapley et M. Pettini (2013). "The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift< z>= 2.4". In : *The Astrophysical Journal* 769.2, p. 146.

Font-Ribera, A., D. Kirkby et al. (2014). "Quasar-Lyman α forest cross-correlation from BOSS DR11 : Baryon Acoustic Oscillations". In : *Journal of Cosmology and Astroparticle Physics* 2014.05, p. 027.

Prochaska, J. X., P. Madau, J. M. O'Meara et M. Fumagalli (2014). "Towards a unified description of the intergalactic medium at redshift z 2.5". In : *Monthly Notices of the Royal Astronomical Society* 438.1, p. 476-486.

Delubac, T., J. E. Bautista et al. (2015). "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars". In : *Astronomy & Astrophysics* 574, A59.

Ade, P. A. et al. (2016). "Planck 2015 results-xiii. cosmological parameters". In : *Astronomy & Astrophysics* 594, A13.

Harris, D. W. et al. (2016). "The composite spectrum of BOSS quasars selected for studies of the Lyα forest". In : *The Astronomical Journal* 151.6, p. 155.

Laurent, P. et al. (2017). "Clustering of quasars in SDSS-IV eBOSS : study of potential systematics and bias determination". In : *Journal of Cosmology and Astroparticle Physics* 2017.07, p. 017.

Prochaska, J., N. Tejos, C. Wotta et al. (2017). "pyigm/pyigm : Initial Release for Publications". In : *UCSC, Zenodo, doi* 10.

Pérez-Ràfols, I., J. Miralda-Escudé, A. Arinyo-i-Prats, A. Font-Ribera et L. Mas-Ribas (2018). "The cosmological bias factor of damped Lyman alpha systems : dependence

on metal line strength". In : *Monthly Notices of the Royal Astronomical Society* 480.4, p. 4702-4709.

Blomqvist, M., H. D. M. Des Bourboux et al. (2019). "Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14". In : *Astronomy & Astrophysics* 629, A86.

Chabanier, S., N. Palanque-Delabrouille et al. (2019). "The one-dimensional power spectrum from the SDSS DR14 Lyα forests". In : *Journal of Cosmology and Astroparticle Physics* 2019.07, p. 017.

de Sainte Agathe, V. et al. (sept. 2019a). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : 629, A85, A85. arXiv : 1904.03400 [astro-ph.CO].

Des Bourboux, H. D. M., K. S. Dawson et al. (2019). "The extended baryon oscillation spectroscopic survey : measuring the cross-correlation between the Mg ii flux transmission field and quasars and galaxies at z= 0.59". In : *The Astrophysical Journal* 878.1, p. 47.

Guo, Z. et P. Martini (2019). "Classification of Broad Absorption Line Quasars with a Convolutional Neural Network". In : *The Astrophysical Journal* 879.2, p. 72.

Sainte Agathe, V. de et al. (2019b). "Baryon acoustic oscillations at z= 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : *Astronomy & Astrophysics* 629, A85.

Aghanim, N. et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In : *Astronomy & Astrophysics* 641, A6.

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Dodelson, S. et F. Schmidt (2020). *Modern cosmology*. Academic press.

Farr, J., A. Font-Ribera, H. D. M. Des Bourboux et al. (2020). "LyaCoLoRe : synthetic datasets for current and future Lyman-α forest BAO surveys". In : *Journal of Cosmology and Astroparticle Physics* 2020.03, p. 068.

Abareshi, B. et al. (2022). "Overview of the instrumentation for the Dark Energy Spectroscopic Instrument". In : *The Astronomical Journal* 164.5, p. 207.

Angulo, R. E. et O. Hahn (2022). "Large-scale dark matter simulations". In : *Living Reviews in Computational Astrophysics* 8.1, p. 1.

Ramirez-Pérez, C., J. Sanchez, D. Alonso et A. Font-Ribera (2022). "CoLoRe : fast cosmological realisations over large volumes with multiple tracers". In : *Journal of Cosmology and Astroparticle Physics* 2022.05, p. 002.

Chaussidon, E. et al. (2023). "Target Selection and Validation of DESI Quasars". In : *The Astrophysical Journal* 944.1, p. 107.

Etourneau, T. et al. (in preparation).

Herrera-Alcantar, H. K. et al. (in preparation). *DESI Lyman-alpha synthetic spectra*.

Ignasi et al. (in preparation). *Lya catalog paper*.

Chapitre 5

Results of the Lyα analysis on mocks and data

The two-point correlation function of discrete matter tracers of the quasi-linear matter density field, e.g., galaxies (W. J. Percival, Cole et al. 2007 ; W. J. Percival, Reid et al. 2010) and quasars (Ata et al. 2018), is a powerful probe to constrain the BAO peak position. The BAO peak position can also be measured using continuous matter tracers, such as Lyman-α forests (J. E. Bautista et al. 2017 ; de Sainte Agathe et al. 2019a ; Des Bourboux, Rich et al. 2020) (see Section 1). The continuous fields of these forests are used as biased matter tracers of the underlying dark matter field, in which the BAO scale is imprinted. Since Lyα forests are seen in high-redshift ($z > 2$) quasar spectra, they provide most highest redshift constraint of the BAO scale (see Figure 2.4), thus are helpful for the study of dark energy models and structure growth in the universe (combining RSD analyses).

In this chapter, I describe the measurement and analysis of the Lyα auto- and cross-correlation functions, following the pipeline and model described in Chapter 3. I perform the analysis on both DESI EDR data, eBOSS DR16 data, and different types of mocks (introduced in Chapter 4). Then I compare their correlation functions, fitting results, and correlations between model parameters. The analysis on Lyα mocks show that the Lyα analysis pipeline performs well and motivates further development of the model for HCDs and metals. A good consistency is found between the correlation functions and parameter constraints of DESI data and DR16 data. However, further investigation should be carried out to compare these two datasets using future DESI data.

In this chapter, I will also give a detailed analysis of the masking of large HCDs (i.e., DLAs) in Lyα pixels, in order to minimize the systematic effect of HCDs. My results show that HCDs mainly affect the correlation function along the line-of-sight beneath the BAO peak, and suggest an investigation of improving the HCD model.

In the DESI collaboration, I have contributed to most of these analyses during my PhD study, and the results are summarized in Gordon et al. 2023 ; Etourneau et al. in preparation ; Ting et al. in preparation.

5.1 Results of mock analyses

In the DESI collaboration, I am one of the main contributors of mock analyses, focusing on Saclay mocks (see Section 4.1.1). Part of my mock analysis is included in Etourneau et al. in preparation ; Herrera-Alcantar et al. in preparation and Ting et al. in preparation. In this section, I measure the correlation functions of Lyα mocks following the same pipeline as used for data (see Section 3.1.1). For the fitting of these mocks (see the model in Section 3.2.1), we have 7 free parameters $\{\alpha_{||}, \alpha_{\perp}, b_{\eta,\mathrm{Ly}\alpha}, \beta_{\mathrm{Ly}\alpha}, b_{\eta,\mathrm{HCD}}, \beta_{\mathrm{HCD}}, L_{\mathrm{HCD}}\}$ for the auto-correlation function, where $b_{\eta,\mathrm{Ly}\alpha} = b_{\mathrm{Ly}\alpha} \times \beta_{\mathrm{Ly}\alpha}$ and $b_{\eta,\mathrm{HCD}} = b_{\mathrm{HCD}} \times \beta_{\mathrm{HCD}}$. We have one more free parameter for the cross-correlations, taking into account the quasar velocity distribution parameter σ_v (see Section 3.2.1). The parameters b_{QSO} and β_{QSO} are fixed because they are not significantly constrained by the cross-only (using only the cross-correlation function) fits (Des Bourboux, Rich et al. 2020).

I will present the analysis on a series of eBOSS mocks (DESI mocks are not ready at the moment) : $eboss-raw$, $eboss-0.0$, $eboss-0.2$, and $eboss-0.3$ mocks (see the definition in Section 4.4). Each of them is produced for a stack of 10 Saclay mocks.

5.1.1 The auto-correlation function

Figure 5.1 shows the measurement of the auto-correlation function ($r \in [0, 200]h^{-1}$Mpc with binsize = $4h^{-1}$Mpc for both $r_{||}$ and $r_{\perp}$, resulting 2500 correlation function bins) for the above four types of mocks to fit these measurements. To fit these measurements, I use the Kaiser model (see Equation 1.42 in Section 1.2.3) for mocks with no HCDs, i.e., $eboss-raw$ (red) and $eboss-0.0$ (black, mocks without HCDs). For $eboss-0.2$ (blue, mocks with HCDs) and $eboss-0.3$ mocks (yellow, mocks with HCDs and with metals), the `Exp` model (see Equation 6.22) is used to fit HCDs. Note that instead of fixing L_{HCD} to $10h^{-1}$Mpc (designed to determine the characteristic suppression scale of HCDs, which relates to the HCD size, see Equation 6.22), it is a free parameter. For $eboss-0.3$ mocks, the Kaiser model is used for each metal line, as described in Section 3.2.2. One can tell from Figure 5.1 and Table 5.1 that :

- HCDs and metals mainly affect the Lyα auto-correlation function at smaller scales ($r \in [20, 80]h^{-1}$Mpc) beneath the BAO scale, and along the line-of-sight, when $\mu > 0.8$. The BAO peak position is not affected by these effects, as seen from the constraints on $\alpha_{||}$ and $\alpha_{\perp}$ where all the scenarios give $\alpha_{||} \sim 1$ and $\alpha_{\perp} \sim 1$ within 1σ.
- The existence of HCDs and metals contributes to additional biases in the correlation function. Extra correlation features due to Si III λ1207, Si II λ1190 and Si II λ1193 appear significantly in auto- and cross-correlation functions with $0.95 < \mu < 1.0$, at $r_{||} \sim 20h^{-1}$Mpc and $r_{||} \sim 60h^{-1}$Mpc, as predicted in Table 3.2.
- The discrepancy between the $eboss-raw$ mocks (red) and the $eboss-0.0$ mocks (black) is due to the distortion effect of the continuum fitting (see Section 3.1.1). This distortion effect can be corrected with the help of the distortion matrix as described in Section 3.1.1.
- Compared to the fits of the $eboss-0.0$ mocks, we obtain worse χ^2 for mocks with HCDs (one more free parameter) or metals (five more free parameters), which motivates the search for better models.
- In all scenarios, we get $b_{\eta,\mathrm{Ly}\alpha} \sim -0.2$ and $\beta_{\mathrm{Ly}\alpha} \sim 1.7$, which agrees with measurements from observations (see the fits of eBOSS DR16 data and DESI data in Section 5.2). $b_{\mathrm{HCD}} \sim -0.02$ which is around 5 times smaller than $b_{\mathrm{Ly}\alpha} = b_{\eta,\mathrm{Ly}\alpha}/\beta_{\mathrm{Ly}\alpha} \sim -0.1$. This verifies our

Mocks	eboss-raw	eboss-0.0	eboss-0.2	eboss-0.3
Correlations	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$LYα
HCD Model	Kaiser model	Kaiser model	Exp model	Exp model
χ^2	1667.2	1463.39	1533.46	1601.4
$N_{\rm data}$	1574	1574	1574	1574
$N_{\rm par}$	4	4	7	12
P	0.04	0.97	0.72	0.24
$\alpha_{\parallel}$	0.994±0.023	0.983±0.012	0.984±0.015	1.008±0.017
$\alpha_{\perp}$	0.985±0.014	1.010±0.008	1.003±0.009	0.990±0.010
$b_{\eta,{\rm LY}\alpha}$	-0.194±0.002	-0.207±0.001	-0.208±0.002	-0.202±0.003
$\beta_{{\rm LY}\alpha}$	1.800±0.040	1.680±0.010	1.580±0.050	1.980±0.160
$b^F_{\rm HCD}$			-0.026±0.004	-0.045±0.006
$\beta_{\rm HCD}$			0.480±0.090	0.550±0.090
$b_{\eta,{\rm SiII}\lambda1260}$				-0.00231±0.00020
$b_{\eta,{\rm SiIII}\lambda1207}$				-0.00614±0.00037
$b_{\eta,{\rm SiII}\lambda1193}$				-0.00183±0.00016
$b_{\eta,{\rm SiII}\lambda1190}$				-0.00299±0.00020
$b_{\eta,{\rm CIVeff}}$				-0.01943±0.00288
$L_{\rm HCD}$			9.480±2.570	8.990±1.790

TABLEAU 5.1 : Best fit parameters of the Lyα auto-correlation function, for different eBOSS Saclay mocks : $eboss-raw$, $eboss-0.0$, $eboss-0.2$, and $eboss-0.3$. P is the p-value of the fit. Here $b^F_{\rm HCD}$ is the flux bias of HCDs, since we use the `Exp` model for HCDs (see definition in Equation 6.22).

assumption that HCDs are a small correction to the Lyα forest power spectrum. The best-fitting value of $L_{\rm HCD}$ is around $10h^{-1}$Mpc. Since the BAO parameters are hardly affected by this parameter, it is reasonable to fix it to $10h^{-1}$Mpc, as was used in the eBOSS DR16 analysis (DES BOURBOUX, RICH et al. 2020).

5.1.2 The cross-correlation function

In addition to the Lyα auto-correlation function, I also measure the Lyα-quasar cross-correlation function for the same four types of mocks. The correlations, as well as their best-fitting models, are shown in the four bottom panels of Figure 5.1, and their associated numerical fits are summarized in Table 5.2. Note that $r_{\rm min} = 20h^{-1}$Mpc is used for both auto- and cross-correlation functions. One can tell from Figure 5.1 and Table 5.2 that :

- For all scenarios, the cross-correlation function gives worse χ^2 and p-value (P in the table) compared to the auto-correlation function. This might be due to various reasons such as the small-scale fluctuations are not realistic enough in our mocks since they are produced by using an input 1D power spectrum (see Section 4.1.1), or the modeling of HCDs and metals need to be improved. I will present the analysis result of the first effect later in this section, which proves the small-scale issues and suggests setting $r_{\rm min} = 40h^{-1}$Mpc for our future analysis. As for the second reason, I will present the improved result using a new HCD model, the `Voigt` model, in Section 6.2. Note that the fit for $eboss-raw$ mocks is very discrepant but the χ^2 value is not that bigger than the other fits. The reason may be that the points in the poorly fitted region have larger errors, while the fits are dominated by small scales where errors are relatively small.

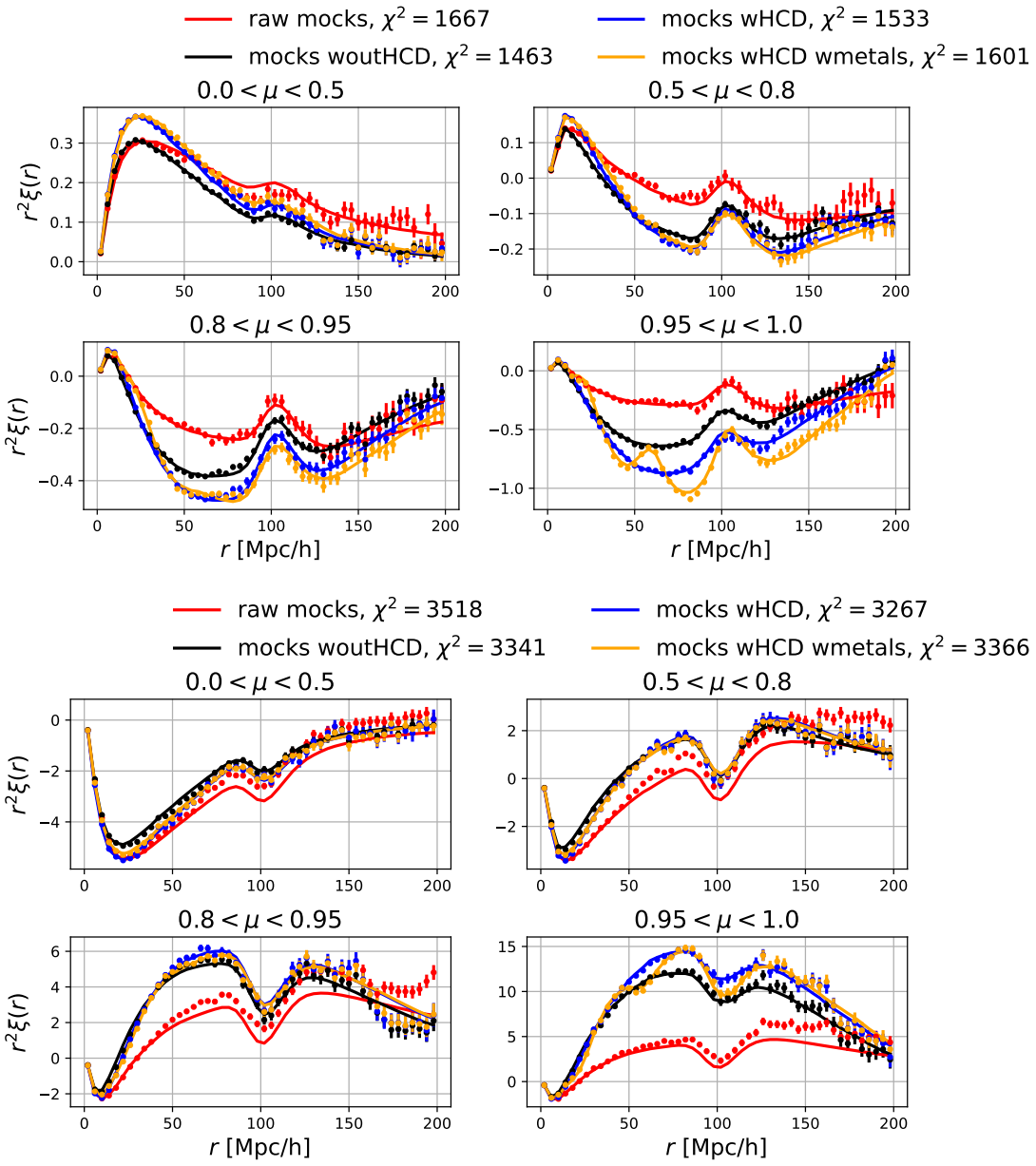

FIGURE 5.1 : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for pixels in the Lyα region (see definition in Table 1.2), and for different eBOSS Saclay mocks : $eboss-raw$ (red), $eboss-0.0$ (black), $eboss-0.2$ (blue), and $eboss-0.3$ (yellow). Each correlation function is computed from a stack of ten mocks. The correlations are multiplied by r^2 to better see the BAO scale. Best-fitting models are shown as solid curves.

Mocks	eboss-raw	eboss-0.0	eboss-0.2	eboss-0.3
Correlations	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$QSO
HCD Model	Kaiser model	Kaiser model	Exp model	Exp model
χ^2	3518.04	3341.68	3267.73	3366.23
$N_{\rm data}$	3148	3148	3148	3180
$N_{\rm par}$	4	4	8	13
P	0.0	0.01	0.05	0.01
$\alpha_{\parallel}$	1.004±0.007	0.999±0.010	0.997±0.012	1.002±0.011
$\alpha_{\perp}$	1.011±0.007	0.999±0.008	1.003±0.010	0.997±0.009
$b_{\eta,{\rm LY}\alpha}$	-0.179±0.001	-0.187±0.002	-0.192±0.003	-0.183±0.005
$\beta_{{\rm LY}\alpha}$	1.560±0.020	1.580±0.020	1.660±0.040	1.960±0.120
$b^F_{\rm HCD}$			-0.043±0.003	-0.047±0.006
$\beta_{\rm HCD}$			0.630±0.080	0.670±0.080
$b_{\eta,{\rm SiII}\lambda1260}$				-0.00229±0.00017
$b_{\eta,{\rm SiIII}\lambda1207}$				-0.00577±0.00021
$b_{\eta,{\rm SiII}\lambda1193}$				-0.00135±0.00020
$b_{\eta,{\rm SiII}\lambda1190}$				-0.00251±0.00019
$b_{\eta,{\rm CIVeff}}$				-0.00500±0.00260
$L_{\rm HCD}$			13.020±2.180	5.790±1.430

TABLEAU 5.2 : Best fit parameters of the Lyα-quasar cross-correlation function, for different eBOSS Saclay mocks : $eboss-raw$, $eboss-0.0$, $eboss-0.2$, and $eboss-0.3$.

- For all scenarios, we do not recover the same $b_{\eta,{\rm Ly}\alpha}$ as what we obtained from the auto-correlation function, even for mocks with only Lyα forests. This problem may be due to the same reason as above : the small-scale fluctuations are not realistic enough in our mocks. This will be discussed further below.
- The constraints on all the metal biases are $1-2\sigma$ away from the constraints obtained from the auto-correlation function, which motivates a further investigation of the metal model.
- In our study, we fix the quasar velocity dispersion parameter as $\sigma_v = 5h^{-1}$Mpc, which is poorly constrained using mocks.

In order to investigate the reason for the mismatch of Lyα biases constrained from the auto- and cross-correlation function, I fit the cross-correlation function of $eboss-0.0$ mocks with different $r_{\rm min} = 20, 30, 40, 50h^{-1}$Mpc ($b_{\rm QSO}$, $\beta_{\rm QSO}$, and σ_v are fixed since they are not sensitive to our mocks). A summary of these fits is shown in Table 5.3. It shows that a better p-value is obtained with an increased $r_{\rm min}$. Moreover, we obtain a smaller discrepancy between the auto and cross biases from $\sim 7\sigma$ to $\sim 2.5\sigma$ for $b_{\eta,{\rm Ly}\alpha}$ and $\beta_{{\rm Ly}\alpha}$ by changing $r_{\rm min} = 20h^{-1}$Mpc to $r_{\rm min} = 50h^{-1}$Mpc. This suggests that the current Saclay mocks do not well generate the small-scale Lyα fluctuations $< 30h^{-1}$Mpc (a further study of this issue is ongoing in the DESI collaboration). This result also suggests we fix $r_{\rm min} = 40$ or $50h^{-1}$Mpc for future Lyα analysis.

Mocks	eboss-0.0	eboss-0.0	eboss-0.0	eboss-0.0
Correlations	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$QSO
HCD Model	Kaiser model	Kaiser model	Kaiser model	Kaiser model
$r_{\rm min}$	20	30	40	50
χ^2	3341.68	3219.45	3109.32	3013.6
$N_{\rm data}$	3148	3102	3030	2946
$N_{\rm par}$	4	4	4	4
P	0.01	0.06	0.14	0.17
$\alpha_{\|\|}$	0.999±0.010	0.999±0.010	0.998±0.010	0.998±0.010
$\alpha_{\perp}$	0.999±0.008	0.999±0.008	0.999±0.008	0.999±0.008
$b_{\eta,{\rm LY}\alpha}$	-0.187±0.002	-0.193±0.002	-0.194±0.002	-0.196±0.003
$\beta_{{\rm LY}\alpha}$	1.580±0.020	1.640±0.030	1.630±0.030	1.640±0.030

TABLEAU 5.3 : Best fit parameters of the Lyα-quasar cross-correlation function, for $eboss-0.0$ mocks, with different $r_{\rm min} = 20, 30, 40, 50 h^{-1}$Mpc.

5.2 Results of the Lyα data analysis

In this section, I describe the Lyα analysis on DESI EDR and eBOSS DR16 data, and compare these results with the results obtained on mocks. In the DESI collaboration, I am one of the main contributors of this analysis. The results for DESI EDR data are summarized in GORDON et al. 2023.

5.2.1 Quasar catalogs

I present in this section a brief description of the quasar catalogs for both eBOSS DR16 and DESI EDR data :

- The eBOSS DR16 quasar catalog contains a total number of 341,468 quasars with $z > 1.77$ and 210,005 quasars with $z > 2.1$.
- The DESI EDR catalog is obtained by running the target selection and post-observation classification using the method introduced in section 2.2.2. It contains 147,899 quasars with $z > 1.77$ and ~ 70000 quasars with $z > 2.1$.

Figure 5.2 shows the redshift distributions of these two datasets. One can find that their shapes are similar if the distributions are normalized.

5.2.2 Measurement of correlation functions

I measure the auto- and cross-correlation function for both DESI EDR and eBOSS DR16 data, following a similar pipeline as the one described in Section 3.1[1]. The difference in the auto-correlation function of these two datasets (measured by the quantity $\frac{(\xi_{\rm DESI}-\xi_{\rm DR16})}{\sqrt{\sigma^2_{\rm DR16}+\sigma^2_{\rm DESI}}}$) is shown in the first plot of Figure 5.3, a two-dimensional plot along $r_{||}$ and $r_{\perp}$. The green lines give the direction of $\mu = \frac{r_{||}}{|\vec{r}|} = 0.5, 0.8, 0.95$.One can tell from this plot that the difference between the correlation functions at each point is homogeneous. The histogram of these differences is shown in the upper right plot of Figure 5.3, indicating a small difference between these two correlation functions, with a standard deviation $\sim 1\sigma$.

[1]For eBOSS DR16 data we measured the correlation function on a log wavelength bins, while for DESI we use a linear bin.

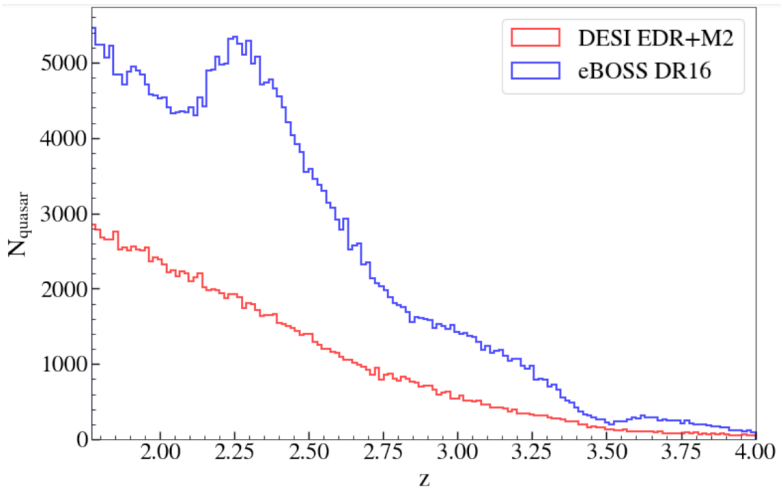

FIGURE 5.2 : Redshift distributions of quasar catalogs for the eBOSS DR16 data (blue) and the DESI EDR data (red). The minimum redshift of quasars is 1.77. Credits : GORDON et al. 2023.

We can further quantify the data quality by using Equation 3.16, where we have Variance $\approx \frac{\langle\delta^2\rangle^2}{fN^{\text{pairs}}}$. Since we have twice the number of Lyα forests in eBOSS DR16 data, we obtain finally three times the correlation pairs than for DESI EDR data (middle-left plot of Figure 5.3). If we further take into account the variance of each correlation (middle-right plot of Figure 5.3), the total weighted fluctuations of all the forests can be estimated by the product of Variance$*\mathrm{N}_{\text{pairs}}$, as shown in the bottom plot of Figure 5.3. This indicates that DESI EDR data is ~ 1.9 times noisier than eBOSS DR16 data ($1.9 = \sqrt{3.5}$). However, since DESI EDR data only contain one observation for each target and the entire DESI observation will re-observe each target four times, the final DESI correlation function should be less noisy than eBOSS data.

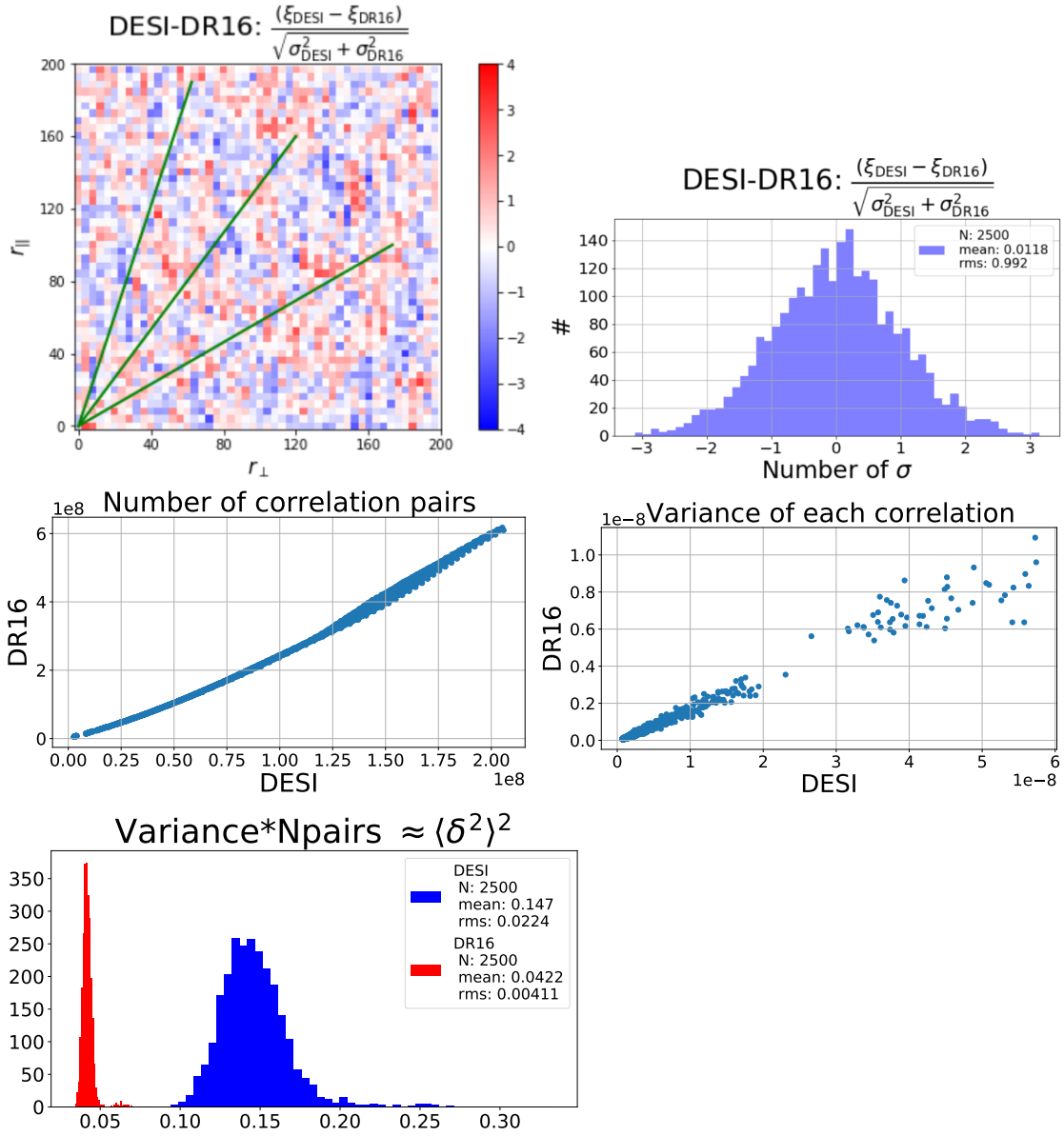

FIGURE 5.3 : Upper left plot : difference of the Lyα auto-correlation function between DESI EDR and eBOSS DR16 data. The quantity $\frac{(\xi_{\rm DESI}-\xi_{\rm DR16})}{\sigma_{\rm DR16}}$ is used to characterize this difference. All the correlation pairs are shown in two dimensions along $r_{||}$ and $r_\perp$. The green lines give the direction of $\mu = \frac{r_{||}}{|\vec{r}|} = 0.5, 0.8, 0.95$. Upper right plot : the histogram of the difference of correlation functions. Mid left plot : number of all the correlation pairs of Lyα forests for these two datasets. Mid-right plot : variances of each correlation pair. Bottom plot : product of variance and the number of pairs, which approximately gives the weighted mean of fluctuations of all the forests.

Data	eBOSS DR16 auto	eBOSS DR16 cross	eBOSS DR16 combined
Correlations	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$LYα + LY$\alpha \times$QSO
DLAs	masking	masking	masking
HCD Model	Exp model	Exp model	Exp model
$N_{\rm data}$	1590	3180	4770
χ^2	1576.36	3219.08	4846.48
$N_{\rm par}$	14	14	17
$\alpha_{\parallel}$	0.982±0.042	0.934±0.039	0.959±0.030
$\alpha_{\perp}$	1.051±0.032	1.060±0.032	1.052±0.023
$b_{\eta,{\rm LY}\alpha}$	-0.173±0.012	-0.234±0.036	-0.193±0.006
$\beta_{{\rm LY}\alpha}$	3.174±1.011	1.936±0.772	1.971±0.174
$b^F_{\rm HCD}$	-0.104±0.019	-0.030±0.056	-0.064±0.008
$\beta_{\rm HCD}$	0.524±0.083	0.500±0.090	0.606±0.083
$b_{\eta,{\rm SiII}\lambda1260}$	-0.00315±0.00110	-0.00159±0.00081	-0.00197±0.00051
$b_{\eta,{\rm SiIII}\lambda1207}$	-0.00919±0.00178	-0.00159±0.00104	-0.00494±0.00055
$b_{\eta,{\rm SiII}\lambda1193}$	-0.00311±0.00087	0.00218±0.00128	-0.00111±0.00048
$b_{\eta,{\rm SiII}\lambda1190}$	-0.00435±0.00103	-0.00483±0.00128	-0.00293±0.00049
$b_{\eta,{\rm CIVeff}}$	-0.00500±0.00256	-0.00500±0.00256	-0.00500±0.00256
$L_{\rm HCD}$	2.264±0.567	0.000±2.703	4.390±0.829
$A_{\rm sky}$	0.010±0.001		0.009±0.001
$\sigma_{\rm sky}$	30.307±1.680		30.898±1.708
$\Delta r_{\parallel,{\rm QSO}}(h^{-1}{\rm Mpc})$		0.195±0.128	0.007±0.117
$\sigma_v(h^{-1}{\rm Mpc})$		9.457±0.473	6.823±0.292
$\xi_0^{\rm TP}$			0.721±0.065

TABLEAU 5.4 : Best fit parameters of the Lyα auto-correlation function, the Lyα-quasar cross-correlation function, and the combined fits, for eBOSS DR16 data.

5.2.3 Results of the correlation functions

For eBOSS DR16 data, in addition to the 7 free parameters $\{\alpha_{\parallel}, \alpha_{\perp}, |b_{\eta,{\rm Ly}\alpha}|, \beta_{{\rm Ly}\alpha}, |b^F_{\rm HCD}|, \beta_{\rm HCD}, L_{\rm HCD}\}$ used in the fits of mocks (see Section 5.1), we use 5 metal bias parameters $\{b_{\eta,{\rm SiII}\lambda1260}, b_{\eta,{\rm SiIII}\lambda1207}, b_{\eta,{\rm SiII}\lambda1193}, b_{\eta,{\rm SiII}\lambda1190}, b_{\eta,{\rm CIVeff}}\}$ and two sky subtraction parameters (see Section 3.2.1) $\{A_{\rm sky}, \sigma_{\rm sky}\}$ for the Lyα auto correlation function, a quasar velocity parameter σ_v and a quasar redshift error parameter $\Delta r_{\parallel,{\rm QSO}}$ for the cross correlation (see Section 3.2.1 and Equation 3.29), and one more free parameter $\xi_0^{\rm TP}$ (see Equation 3.24) for the combined fits.

For DESI EDR data, the BAO parameters $\{\alpha_{\parallel}, \alpha_{\perp}\}$ and the $b_{\eta,{\rm C\ IV}_{eff}}$ metal bias are blinded (before the analyses are finalized, we use strategies like adding random shifts to data to hide the critical values of parameters from potential influence by priors). These parameters are therefore fixed in my analysis. For their fits, $A_{\rm inst}$ is used instead of $A_{\rm sky}$ and $\sigma_{\rm sky}$ to characterize the sky subtraction effect (see Equation 3.40). A free parameter $A_{\rm BAO}$ is used to characterize the amplitude of the BAO peak in the correlation function, using a modified formula of Equation 3.21 :

$$\xi(r_{\parallel}, r_{\perp}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\rm smooth}(r_{\parallel}, r_{\perp}) + A_{\rm BAO}\xi_{\rm peak}(r_{\parallel}\alpha_{\parallel}, r_{\perp}\alpha_{\perp}). \tag{5.1}$$

I present in Figure 5.4 the measurement and best-fit models of the Lyα auto- and cross-correlation functions, for the eBOSS DR16 and DESI EDR data. One can tell from the plots that the errors of DESI EDR correlations are larger than those of the eBOSS DR16 since it only contains about half the number of quasars. The BAO peak can be seen at $\sim 100h^{-1}$Mpc in both datasets, for both the auto- and cross-correlation functions.

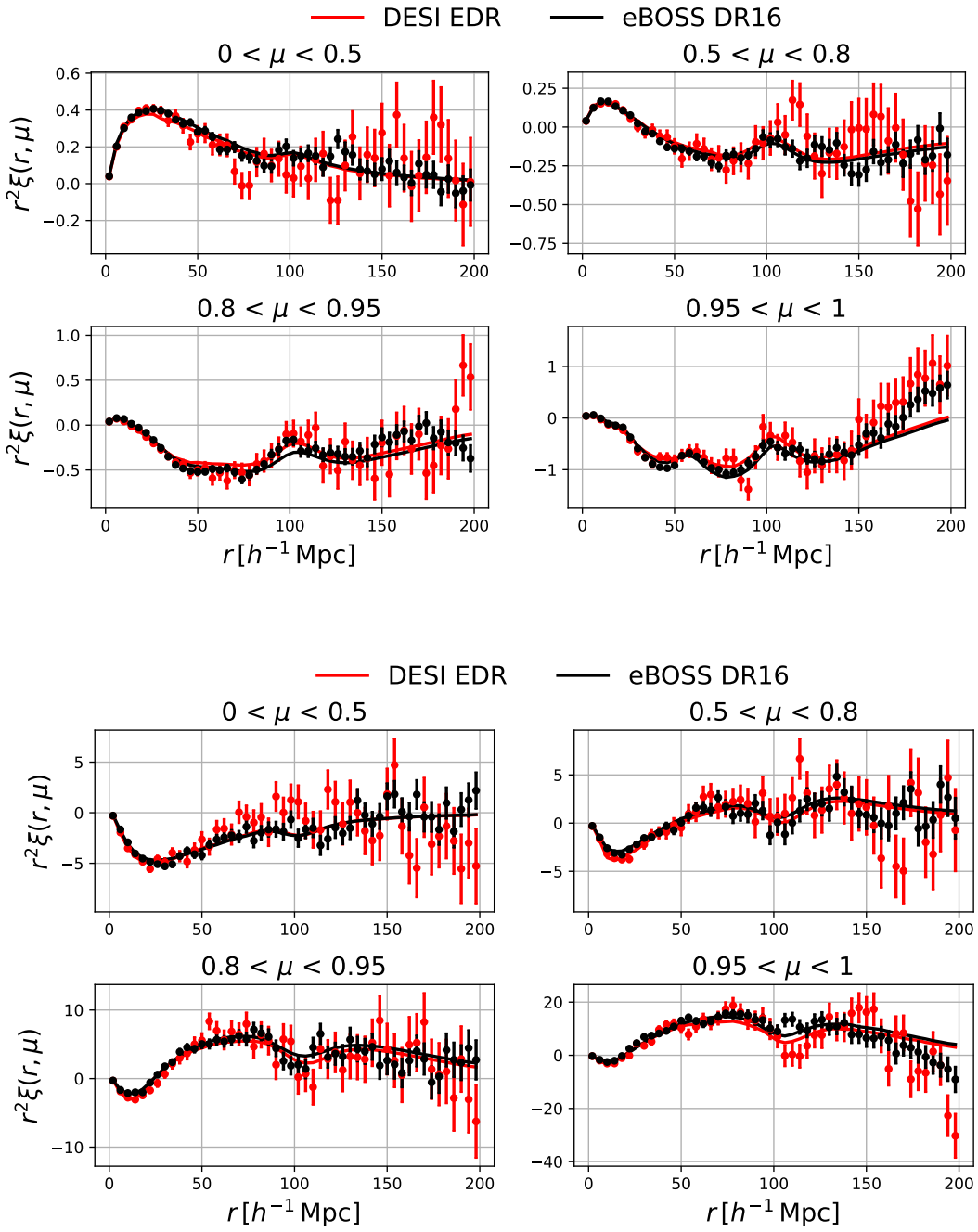

FIGURE 5.4 : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for eBOSS DR16 (black) and DESI EDR data (red), with pixels in the Lyα region. The correlations are multiplied by r^2 to better see the BAO scale. The solid curves show the best-fit models (using the `Exp` model for HCDs), in four wedges of $|\mu| = |\frac{r_\parallel}{r}|$. The fitted range is chosen as $r \in [10, 180]h^{-1}$Mpc (note that the strategy of using a higher value of $r_{\rm min}$ is not applied yet).

Data	DESI EDR auto	DESI EDR cross	DESI EDR combine
Correlations	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$LYα + LY$\alpha \times$QSO
DLAs	masking	masking	masking
HCD Model	Exp model	Exp model	Exp model
$N_{\rm data}$	1590	3180	4770
χ^2	1659.16	3205.02	4877.08
$N_{\rm par}$	11	13	14
$\alpha_{\parallel}$	1.0	1.0	1.0
$\alpha_{\perp}$	1.0	1.0	1.0
$b_{\eta,{\rm LY}\alpha}$	-0.207±0.013	-0.171±0.015	-0.200±0.007
$\beta_{{\rm LY}\alpha}$	1.968±0.472	1.329±0.176	1.540±0.165
$b^F_{\rm HCD}$	-0.045±0.022	-0.030±0.008	-0.025±0.009
$\beta_{\rm HCD}$	0.496±0.085	0.500±0.089	0.496±0.088
$b_{\eta,{\rm SiII}\lambda1260}$	-0.00236±0.00139	-0.00253±0.00114	-0.00225±0.00082
$b_{\eta,{\rm SiIII}\lambda1207}$	-0.00489±0.00159	-0.00298±0.00134	-0.00359±0.00090
$b_{\eta,{\rm SiII}\lambda1193}$	-0.00087±0.00098	-0.00000±0.00085	-0.00000±0.00094
$b_{\eta,{\rm SiII}\lambda1190}$	-0.00349±0.00106	-0.00150±0.00114	-0.00272±0.00073
$L_{\rm HCD}$	4.447±3.553	6.656±4.529	5.997±3.326
$\Delta r_{\parallel,{\rm QSO}}(h^{-1}{\rm Mpc})$		-2.278±0.162	-2.303±0.186
$\sigma_v(h^{-1}{\rm Mpc})$		4.848±0.581	5.849±0.504
$\xi_0^{\rm TP}$			0.993±0.111
$A_{\rm BAO}$	1.672±0.477	0.924±0.402	1.194±0.306
$A_{\rm inst}$	0.000±0.002	0.000±0.002	0.000±0.002

TABLEAU 5.5 : Best fit parameters of the Lyα auto-correlation function, the Lyα-quasar cross-correlation function, and the combined fits, for DESI EDR data.

The fits for the auto- and cross-correlations, as well as their combined correlation (fit of the auto- and cross-correlation at the same time, using a combined likelihood), are shown in Table 5.4 for eBOSS DR16 data, and Table 5.5 for DESI EDR data. One can tell from these results that :

- The combined fit gives tighter constraints on the parameters that are used in both the auto- and cross-correlation. It helps especially in constraining all the bias parameters, except the C IV$_{\rm eff}$ bias. Because of this, we prefer to fix the value of $b_{\eta,\mathrm{C\,IV(eff)}}$ for the fit of DESI EDR data.
- Both the auto- and cross-correlation can be used to constrain the BAO parameters, $\alpha_{||}$ and $\alpha_{\perp}$. The combination of these two correlations gives tighter constraints.
- The ratio of $b^F_{\rm HCD}/b_{\eta,\mathrm{Ly}\alpha}$ is $\sim 70\%$ for the auto-correlation and $\sim 30\%$ for the combined correlation, which is much larger than what we detect in mocks $\sim 10\%$ (see Table 6.4). This might be due to various reasons : HCDs are not inserted into the mocks in the way that they distribute in the universe ; the modeling of the HCD bias is not accurate enough, and it captures something else than HCDs in real data. Moreover, for both datasets, the cross-correlation hardly constrain $\beta_{\rm HCD}$ and $L_{\rm HCD}$. In all scenarios, $\beta_{\rm HCD}$ is not well constrained (dominated by priors for both datasets). The best value for $L_{\rm HCD}$ of eBOSS data is $\sim 2-4h^{-1}$Mpc, which is several sigmas away from the value in mocks (see Table 6.4). These values of $L_{\rm HCD}$ is smaller than the bin size of the measured correlation function $4h^{-1}$Mpc, meaning that the HCD modeling is modeling something else than HCDs.
- These two datasets give comparable χ^2 and constrained values of $|b_{\eta,\mathrm{Ly}\alpha}|$ and $\beta_{\mathrm{LY}\alpha}$ within several sigmas. Moreover, we do not apply the sky subtraction in DESI data, while a new parameter $A_{\rm BAO}$ is used to characterize the amplitude of the correlation function. A further comparison of the same parameter constraints is investigated in the next Section.

5.2.4 Correlation between parameters

In order to visualize the constraints and correlations on the fitted parameters, I investigate the parameter inference by performing Gaussian likelihood with Gaussian distributed parameters. This approach does not give the true posteriors of these parameters that can be determined by performing Markov chain Monte Carlo (MCMC) simulations, but it is helpful to understand their correlations. The correlation ρ between two parameters A and B is computed by :

$$\rho_{\rm A,B} = \frac{cov(\mathrm{A,B})}{\sigma_{\rm A}\sigma_{\rm B}}, \tag{5.2}$$

where cov(A,B) stands for the covariance between these two parameters, $\sigma_{\rm A}$ and $\sigma_{\rm B}$ are the standard deviations for A and B. cov(A,B) is defined as :

$$\mathbb{E} = [(\mathrm{A}-\mu_{\rm A})(\mathrm{B}-\mu_{\rm B})]. \tag{5.3}$$

Here $\mu_{\rm A}, \mu_{\rm B}$ are the means and $\mathbb{E}$ is the expectation value.

I present triangle plots (which show a series of sub-plots for all the parameters) for eBOSS DR16 data in Figure 5.5 and for DESI EDR data in Figure 5.6. From these plots, we draw the following conclusions :

- The BAO parameters, $\alpha_{||}$ and $\alpha_{\perp}$, are not correlated with other Lyα parameters, seen from Figure 5.5. This is good, since we do not want to have these parameters correlated with other systematic effects.

- $|b_{\eta,\mathrm{Ly}\alpha}|$ is anti-correlated with $|b^F_{\mathrm{HCD}}|$, suggesting that the sum of these biases determines the total bias. However, it is not strongly anti-correlated with other metal biases. It is also anti-correlated with $\beta_{\mathrm{Ly}\alpha}$, which suggests that we need to investigate a better combination of parameters to determine the monopole and quadrupole of the correlation function. This is further discussed in Section 6.2.

- From the auto-correlation, L_{HCD} is strongly correlated with $|b^F_{\mathrm{HCD}}|$, indicating that what matters is the product of these two parameters. A better model needs to be investigated, and I will describe this study in Section 6.2.

- For eBOSS DR16 data, all metal bias parameters, except $b_{\eta,\mathrm{C\,IV(eff)}}$, are strongly correlated with each other. Moreover, they are also correlated with $|b_{\eta,\mathrm{Ly}\alpha}|$. It motivates the search for better models for metals.

- For DESI EDR data, $\beta_{\mathrm{Ly}\alpha}$ is strongly correlated with $|b_{\mathrm{HCD}}|$, and β_{HCD} is not well constrained. This might be due to the limitation of the `Exp` HCD model (see Equation 6.22), and an improved result using the `Voigt` model will be presented in Section 6.2. $b_{\eta,\mathrm{SiII}\lambda 1193}$ is not well constrained by the cross-correlation and thus should be fixed.

- For both eBOSS and DESI fits, contours from the auto- and cross-correlation overlap in most cases, which allow meaningful combined constraints to be derived.

To investigate further the comparison between the two datasets, I fix the BAO parameters $\alpha_{||} = \alpha_{\perp} = 1$, and $b_{\eta,\mathrm{C\,IV(eff)}}$, while using A_{BAO} for both datasets. This comparison is shown in Figure 5.7. One can tell from the plots that the two datasets agree on the correlations and constraints on most parameters. The exception is b_{HCD} and $\beta_{\mathrm{Ly}\alpha}$ where the two datasets are 2σs away using combined fits. These two parameters are also strongly correlated with each other. To break this degeneracy, a new HCD model is developed, and a less correlated parametrization is proposed in the next chapter.

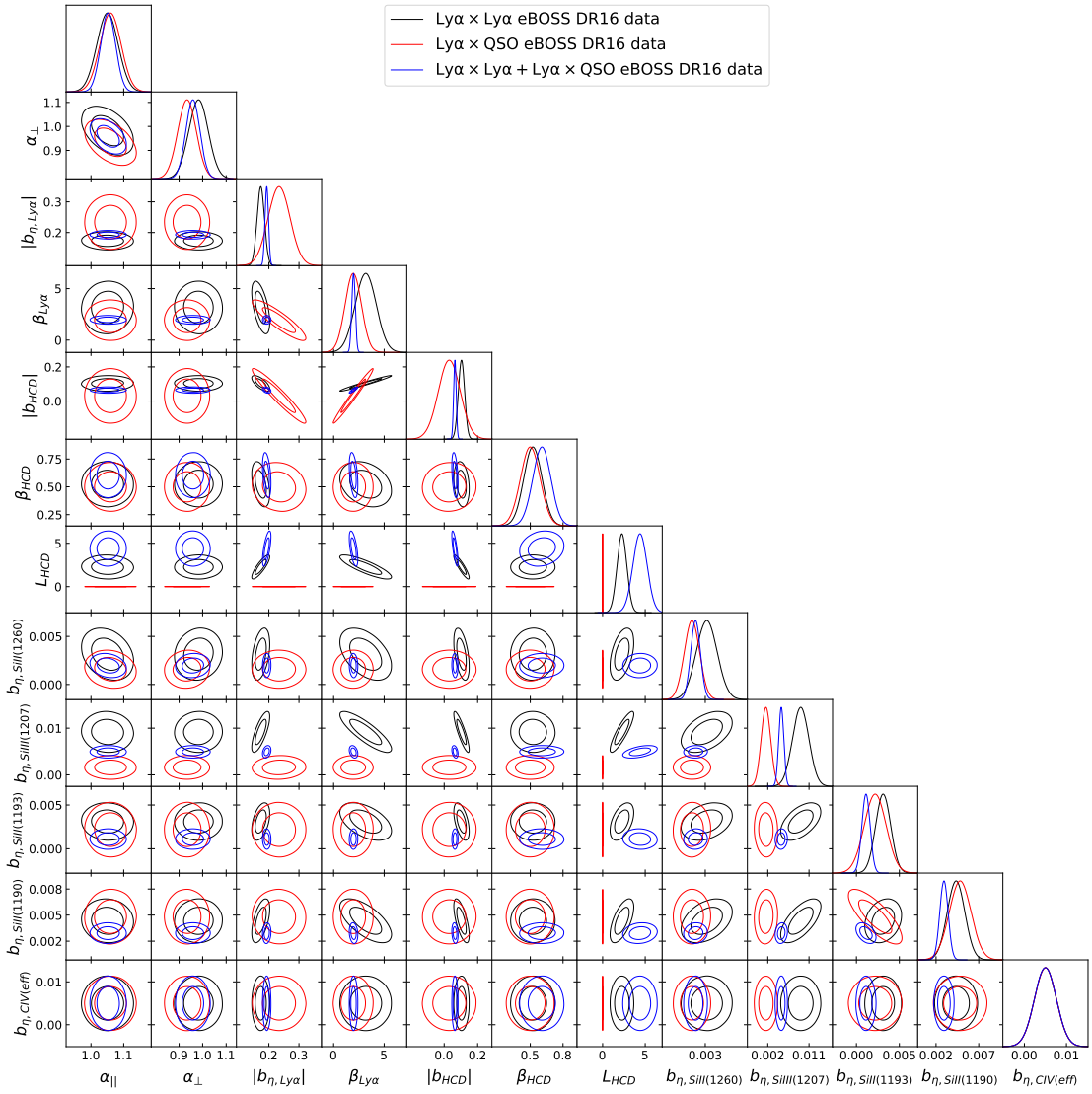

FIGURE 5.5 : Triangle plot for the Lyα parameters constraints $\{\alpha_{||}$, $\alpha_{\perp}$, $|b_{\eta,\mathrm{LY}\alpha}|$, $\beta_{\mathrm{LY}\alpha}$, $|b^{F}_{\mathrm{HCD}}|$, β_{HCD}, L_{HCD}, $b_{\eta,\mathrm{SiII}\lambda1260}$, $b_{\eta,\mathrm{SiIII}\lambda1207}$, $b_{\eta,\mathrm{SiII}\lambda1193}$, $b_{\eta,\mathrm{SiII}\lambda1190}$, $b_{\eta,\mathrm{CIV\,eff}}\}$ using the Lyα auto-correlation function (black), Lyα-quasar cross-correlation (red), and auto + cross combined fits (blue), with eBOSS DR16 data.

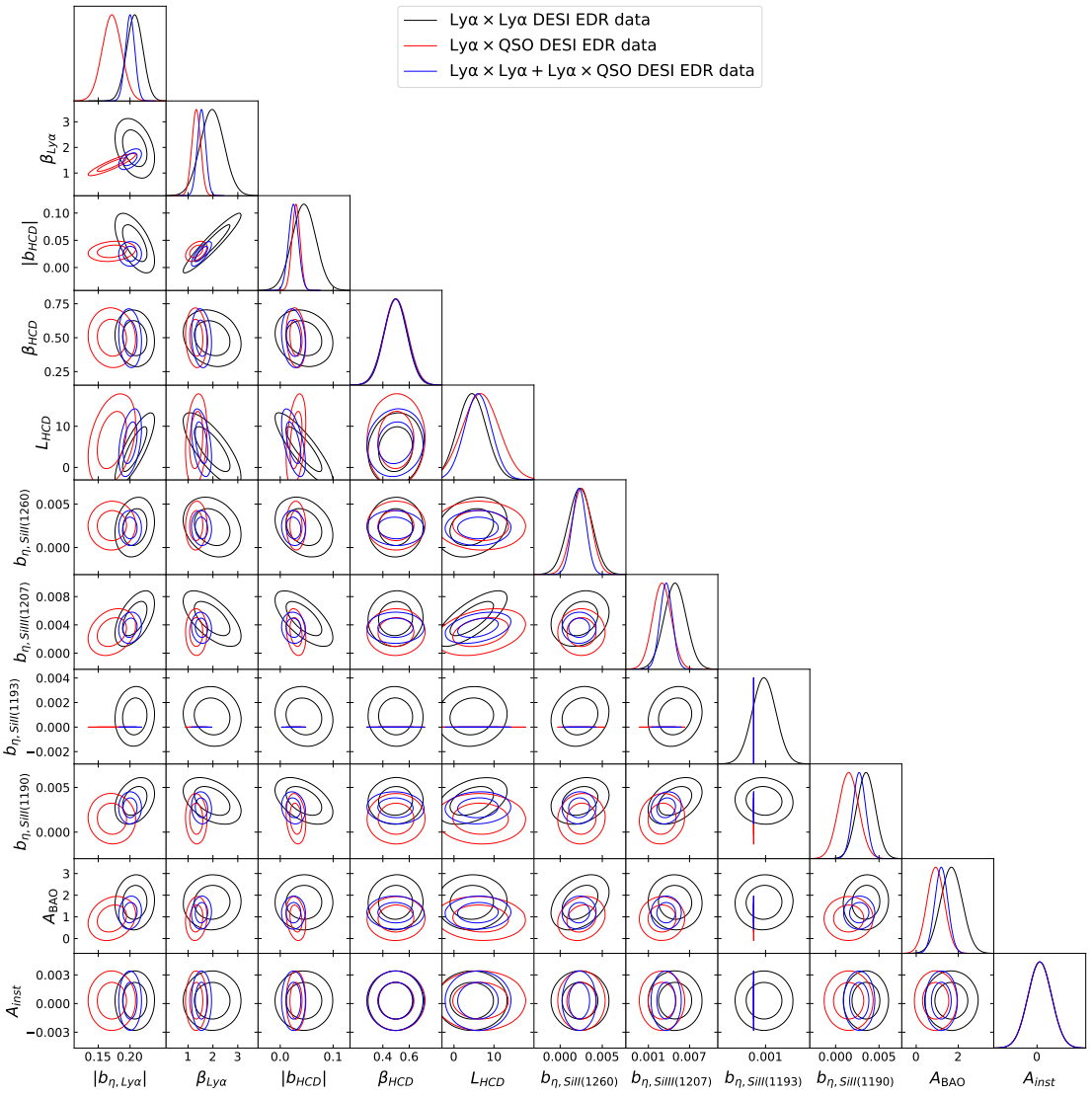

FIGURE 5.6 : Triangle plot for the Lyα parameters constraints $\{|b_{\eta,\mathrm{LY}\alpha}|$, $\beta_{\mathrm{LY}\alpha}$,$|b^F_{\mathrm{HCD}}|$, β_{HCD}, L_{HCD},$b_{\eta,\mathrm{SiII}\lambda1260}$, $b_{\eta,\mathrm{SiIII}\lambda1207}$, $b_{\eta,\mathrm{SiII}\lambda1193}$, $b_{\eta,\mathrm{SiII}\lambda1190}$, $A_{\mathrm{BAO}}\}$ using the Lyα auto-correlation function (black), Lyα-quasar cross-correlation (red), and auto + cross combined fits (blue), with DESI EDR data.

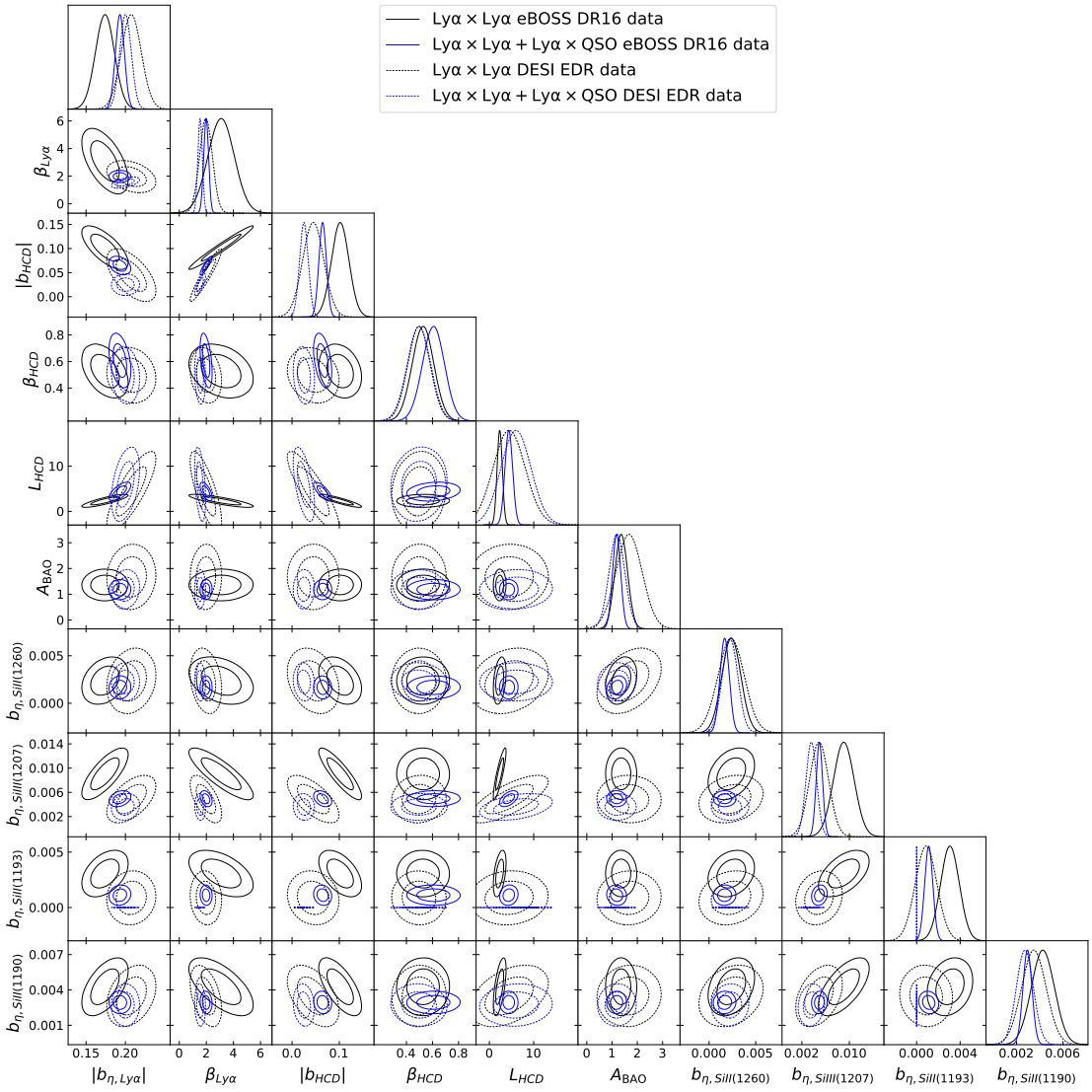

FIGURE 5.7 : Triangle plot for the Lyα parameters constraints $\{|b_{\eta,\mathrm{LY}\alpha}|, \beta_{\mathrm{LY}\alpha}, |b^F_{\mathrm{HCD}}|, b_{\eta,\mathrm{SiII}\lambda1260}, b_{\eta,\mathrm{SiII}\lambda1207}, b_{\eta,\mathrm{SiII}\lambda1193}, b_{\eta,\mathrm{SiII}\lambda1190}\}$ using the Lyα auto-correlation function (black) and auto + cross combined fits (blue), with eBOSS DR16 data (solid) and DESI EDR data (dashed). Here I fixed the BAO parameters.

5.3 Masking DLAs

As will be introduced in Section 6.1, large DLAs with $\log(N_{\rm HI}) > 20.3{\rm cm}^{-2}$ can be detected by machine learning algorithms. Since we can not fully model the HCD impact with analytical formulas, masking the DLA pixels is an efficient way to minimize their effect. In this section, I discuss the impact of masking DLAs on the Lyα auto-correlation function. Three datasets are used for comparison in this study : DESI EDR data, eBOSS DR16 data, and $eboss-0.2$ mocks. Note that for DESI and DR16 data, I am using the DLA catalogs described in Section 6.1.3. For mocks, the true catalog of inserted HCDs is used for masking. This comparison is shown in Figure 5.8, and the associated fits are summarized in Table 5.6. In Figure 5.8, the quantity $(\xi_{\rm wDLAmasking} - \xi_{\rm woutDLAmasking})/\sigma_{\rm woutDLAmasking}$ is used to characterize the difference between correlation functions computed with or without DLA masking. This is shown as two-dimensional plots along $r_{||}$ and $r_{\perp}$. One can tell from the plots that :

- The impact of masking DLAs is mainly along the line-of-sight when $40h^{-1}{\rm Mpc} < r_{||} < 100h^{-1}{\rm Mpc}$, and in transverse direction when $r_{\perp} < 40h^{-1}{\rm Mpc}$. The small-scale impact is weaker since we fit from $r_{\rm min} = 10h^{-1}{\rm Mpc}$.
- The correlation function for DESI EDR data is the least affected by DLA masking, with the smallest standard deviation for the characteristic quantity. Mocks are most affected by this effect, since we are using a true DLA catalog. The insertion of HCDs into mocks could also bias this study since they are uniformly randomly located in the possible peaks of the matter density field (see Section 4.1.1). This effect can be further investigated in a future analysis.
- For these three samples, masking or not masking DLAs do not change significantly the results : for every parameter, the results in the two options are within 1 sigma. However, for eBOSS DR16 data, both cases provide constraints on $L_{\rm HCD}$ smaller than the bin size ($4h^{-1}{\rm Mpc}$) of the measured correlation function, meaning that the HCD model is modeling something else than HCDs. This suggests a better modeling of HCDs, which I will describe in the next chapter.

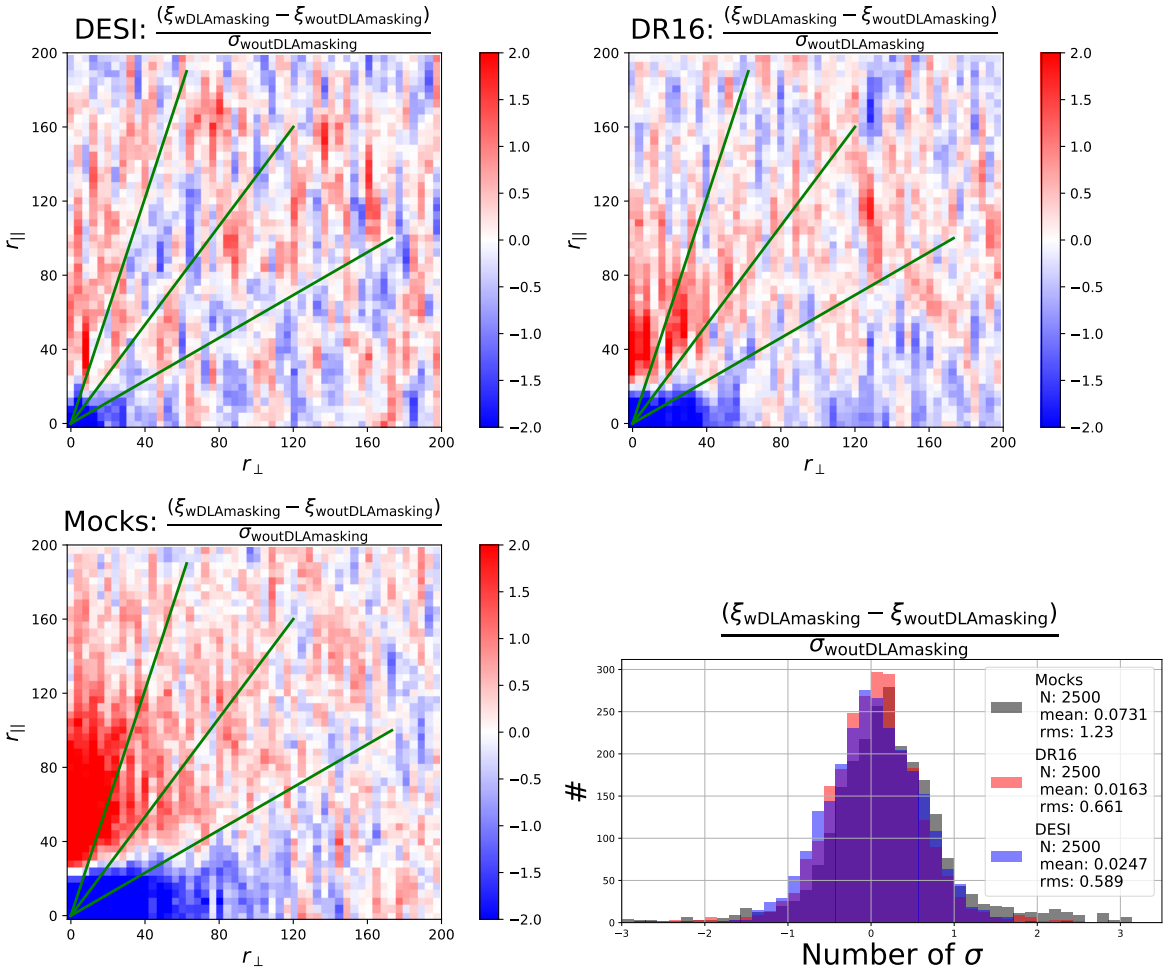

FIGURE 5.8 : Comparison of the Lyα auto-correlation function with or without DLA masking. The quantity $\frac{(\xi_{\text{wDLAmasking}} - \xi_{\text{woutDLAmasking}})}{\sigma_{\text{woutDLAmasking}}}$ is used to characterize this difference. The quantity is shown as 2D plots for all the $(r_{||}, r_\perp)$ pairs. The green curves give the direction of $\mu = \frac{r_{||}}{|\vec{r}|} = 0.5, 0.8, 0.95$. The histogram for these three datasets is presented in the last plot : DESI EDR data (red), eBOSS DR16 data (blue), and $eboss - 0.2$ mocks (black).

Data	eboss-0.2	eboss-0.2	eBOSS DR16	eBOSS DR16	DESI EDR	DESI EDR
Correlations	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα
DLAs	masking	no masking	masking	no masking	masking	no masking
HCD Model	Exp model	Exp model	Exp model	Exp model	Exp model	Exp model
χ^2	1523.56	1533.46	1576.22	1594.96	1659.04	1709.72
$N_{\rm data}$	1574	1574	1590	1590	1590	1590
$N_{\rm par}$	7	7	14	14	12	12
P	0.78	0.72	0.49	0.36	0.08	0.01
$\alpha_{\|\|}$	0.980±0.014	0.984±0.015	0.981±0.042	0.974±0.044	1.000	1.000
$\alpha_{\perp}$	1.006±0.009	1.003±0.009	1.048±0.034	1.042±0.034	1.000	1.000
$b_{\eta,{\rm LY}\alpha}$	-0.204±0.002	-0.208±0.002	-0.175±0.013	-0.173±0.013	-0.183±0.014	-0.194±0.013
$\beta_{{\rm LY}\alpha}$	1.670±0.020	1.580±0.050	3.230±1.260	5.250±3.290	1.590±0.340	1.870±0.410
$b^F_{\rm HCD}$	-0.019±0.001	-0.026±0.004	-0.105±0.022	-0.139±0.020	-0.063±0.020	-0.043±0.021
$\beta_{\rm HCD}$	0.470±0.090	0.480±0.090	0.530±0.080	0.510±0.080	0.510±0.090	0.510±0.090
$b_{\eta,{\rm SiII}\lambda1260}$			-0.00316±0.00116	-0.00342±0.00140	-0.00342±0.00140	-0.00338±0.00144
$b_{\eta,{\rm SiIII}\lambda1207}$			-0.00932±0.00204	-0.01046±0.00249	-0.00501±0.00169	-0.00510±0.00176
$b_{\eta,{\rm SiII}\lambda1193}$			-0.00315±0.00093	-0.00309±0.00100	-0.00215±0.00110	-0.00322±0.00128
$b_{\eta,{\rm SiII}\lambda1190}$			-0.00440±0.00112	-0.00490±0.00127	-0.00358±0.00120	-0.00271±0.00128
$b_{\eta,{\rm CIVeff}}$			-0.00513±0.00262	-0.00503±0.00260	-0.00511±0.00263	-0.00497±0.00259
$L_{\rm HCD}$	2.290±0.750	9.480±2.570	2.280±0.630	2.590±0.520	5.180±3.330	4.350±3.660

TABLEAU 5.6 : Best fit parameters of the Lyα auto-correlation function, for $eboss-0.2$ mocks, eBOSS DR16 data, and DESI EDR data, with or without DLAs masking.

5.4 Summary and prospects

In this chapter, I measured the correlation functions for different mocks and data (eBOSS DR16 and DESI EDR) using the analysis pipeline described in Section 3.1.1 and fitted them with the model introduced in Section 3.2.

The results of the auto-correlation functions of the simplest mocks, i.e., $eboss-raw$ and $eboss-0.0$ mocks, show a successful validation of the Lyα analysis pipeline, with good χ^2 and parameter constraints. However, taking into account astrophysical contaminants (HCDs and metals) yields worse χ^2 and strong correlations between parameters. This motivates the search for better models of HCDs (see the next chapter) and metals.

For the cross-correlation functions of mocks, I found a significant discrepancy comparing the constraints on $b_{\eta,{\rm Ly}\alpha}$ and $\beta_{{\rm Ly}\alpha}$ with the values obtained using the auto-correlation functions. This might be due to the unrealistic construction of small-scale Lyα fluctuations or the quasar non-linear velocities. To minimize this impact, I tested several cuts on $r_{\rm min}$ and suggested to fix $r_{\rm min} = 40h^{-1}$Mpc for future analysis.

I presented a preliminary comparison of DESI EDR and eBOSS DR16 Lyα analysis. Comparable χ^2 and similar parameter correlations were found between their fits. The DESI EDR data show encouraging data quality and the need for more systematic studies to prepare the upcoming enormous DESI dataset.

In the next chapter, I will present an analysis of HCDs, one of the most important systematic effects of Lyα BAO. I will introduce a new model, the so-called `Voigt` model, to better characterize this effect, solve the current puzzles that we met using mocks, and prepare for future DESI analysis.

Bibliographie du présent chapitre

Percival, W. J., S. Cole et al. (2007). "Measuring the baryon acoustic oscillation scale using the sloan digital sky survey and 2dF galaxy redshift survey". In : *Monthly Notices of the Royal Astronomical Society* 381.3, p. 1053-1066.

Percival, W. J., B. A. Reid et al. (2010). "Baryon acoustic oscillations in the Sloan Digital Sky Survey data release 7 galaxy sample". In : *Monthly Notices of the Royal Astronomical Society* 401.4, p. 2148-2168.

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z= 2.3 with SDSS DR12 Lyα-Forests". In : *Astronomy & Astrophysics* 603, A12.

Ata, M. et al. (2018). "The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample : first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2". In : *Monthly Notices of the Royal Astronomical Society* 473.4, p. 4773-4794.

de Sainte Agathe, V. et al. (sept. 2019a). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : 629, A85, A85. arXiv : 1904.03400 [astro-ph.CO].

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Gordon, C. et al. (2023). "3D Correlations in the Lyman-$alpha$ Forest from Early DESI Data". In : *arXiv e-prints*, arXiv-2308.

Etourneau, T. et al. (in preparation).

Herrera-Alcantar, H. K. et al. (in preparation). *DESI Lyman-alpha synthetic spectra.*

Ting, T. et al. (in preparation). *Modeling of the High Column Density systems in The Lyman-α forest.*

Chapitre 6

Detection and modeling of the High Column Density systems

In this chapter, I will describe the most important contribution of this thesis to the DESI Lyα collaboration : the analysis of High Column Density systems (hereafter HCDs), one of the most important systematic effects for Lyα forests BAO (McDonald, Seljak, Cen, Bode et Ostriker 2005 ; Viel, Haehnelt, R. Carswell et T.-S. Kim 2004 ; Font-Ribera et Miralda-Escudé 2012 ; Rogers, Bird, Peiris, Pontzen, Font-Ribera et Leistedt 2018b). As described in Section 1.2.4, HCDs are dense concentrated gas regions in IGM, with Neutral Hydrogen (hereafter HI) column densities $N_{\rm HI} > 10^{17.2}{\rm cm}^{-2}$. Large HCDs with $N_{\rm HI} > 10^{20.3}{\rm cm}^{-2}$ are called Damped Lyα systems (DLAs, see Section 3.2.2), which show strong absorption features in quasar spectra with broad wings. These DLAs are thus detectable by visual inspection, model-driven fitting (e.g., Voigt profile fitting (J. X. Prochaska, Herbert-Fort et Wolfe 2005 ; Noterdaeme, Petitjean, Ledoux et Srianand 2009 ; Noterdaeme, Petitjean, Carithers et al. 2012)), or data-driven machine-learning algorithms (Parks, J. X. Prochaska, Dong et Cai 2018 ; Garnett, S. Ho, Bird et J. Schneider 2017 ; Fumagalli, Fotopoulou et Thomson 2020 ; Chabanier, Etourneau et al. 2022 ; Wang et al. 2022 ; Jiaqi et al. in preparation). I describe two machine learning DLA finders : the CNN DLA finder (Parks, J. X. Prochaska, Dong et Cai 2018) and Gaussian Processes DLA finder (M.-F. Ho, Bird et Garnett 2021) and their comparison in Section 6.1. A combination of these two finders is applied to construct the DESI DLA catalog (Jiaqi et al. in preparation), and I will further compare this catalog with the eBOSS DR16 DLA catalog (Chabanier, Etourneau et al. 2022).

The damping wings of the HCD absorption profile (see details in Section 1.2.4) will result in a suppression on the Lyα forest power spectrum along the line-of-sight at $0.01 < k_{||} < 1h{\rm Mpc}^{-1}$, and a broadband impact on the correlation function. The modeling of this impact is essential, in order to determine the correct Lyα biases and RSD parameters. Different phenomenological models were used in the previous analyses, such as the `Sinc` model for DR14 (de Sainte Agathe et al. 2019a) and the `Exp` model for DR16 (Des Bourboux, Rich et al. 2020) analyses. However, these models failed to give a good fitting for Lyα forests with HCDs, and do not provide a clear physical understanding of the HCD bias. **[During my thesis, I have developed a three-parameter empirical fitting function, the $L\beta\gamma$ model, to characterize the damping effect of HCDs on the Lyα correlation function and power spectrum.]** This model shows no difference with the `Exp` model when applied to eBOSS Saclay mocks with HCDs, while showing encouraging improvement when applied to eBOSS DR16 data in the range of $20h^{-1}{\rm Mpc} < r < 80h^{-1}{\rm Mpc}$. This suggests that the $L\beta\gamma$ model is probably modeling an effect

beyond HCDs, which has a non-negligible impact on the Lyα correlations.

[I further developed a theoretical model, which I call the `Voigt` model, based on the Voigt absorption profile that parametrizes the damping wings of HCDs, and takes into account the HI column density probability distribution of HCDs.] It has no additional free parameters, providing a physical measurement of the bias and RSD parameters of HCDs, as well as a good constraint on the Lyα parameters. The simplified formula of this model was inspired by Font-Ribera et Miralda-Escudé 2012 and was proposed in Rogers, Bird, Peiris, Pontzen, Font-Ribera et Leistedt 2018a without analytical derivation and normalization information of HCD halo bias. My contribution to this model allows the physical measurement of both Lyα and HCD parameters and further consideration of higher-order correlations. The good performance of this model on both mocks and data suggests its further implementation for future DESI analyses. Some extended studies could also be carried out with this model, such as the HCD impact on the Alcock-Paczyński effect (see Section 1).

6.1 Detection of DLAs

With the enormous number of Lyα forests observed by current or future large cosmology surveys (e.g., eBOSS/DESI), it is not possible to visually inspect all quasar spectra and construct DLA catalogs artificially. Moreover, absorption profiles of small HCDs can hardly be distinguished from Lyα absorptions. It is therefore essential to develop automatic and accurate algorithms to construct DLA catalogs. In this section, I describe explicitly a traditional fitting algorithm (the `Voigt` profile fitting), and two machine learning algorithms (CNN and Gaussian Processes), that were used in the eBOSS DR16 analysis (Chabanier, Etourneau et al. 2022), and will be used for future DESI data (Wang et al. 2022 ; Jiaqi et al. in preparation).

During my thesis, I contributed to the comparison of the eBOSS DR16 and DESI EDR DLA catalogs (built using a combination of the CNN and Gaussian Processes algorithms, see description below), and also worked on the development of a Bayesian-based CNN algorithm. This new algorithm is still under study and a preliminary result was presented at the IAP Colloquium 2021 (Ting Tan et Balland 2021). The progress of this study is not included in this thesis.

6.1.1 Voigt profile fitting

In astrophysics and absorption spectroscopy, fitting observed spectral lines using Voigt profiles was used as a common tool Sundius 1973, which determines the absorption lines as a product of Gaussian profiles (for Doppler shift) and Lorentzian profiles (for collisions of atoms). The good resolution of the SDSS/DESI spectra has made it possible to classify DLA systems and measure their $N_{\rm HI}$ using automatic algorithms such as the voigt profile fitting. The voigt profile fitting algorithm for DLAs was developed in (J. X. Prochaska et Herbert-Fort 2004), and further applied to SDSS data (J. X. Prochaska, Herbert-Fort et Wolfe 2005 ; Noterdaeme, Petitjean, Ledoux et Srianand 2009 ; Noterdaeme, Petitjean, Carithers et al. 2012). To find a DLA, it estimates the median signal-to-noise ratio (SNR) of the quasar spectrum ($\mathrm{SNR_{QSO}}$) using a characteristic window of 150 pixels starting from 51-200 pixels blueward (to get rid of the high SNR peak) of the Lyα peak. For quasars with lower redshifts where the Lyα peak lies at less than 200 pixels from the start of the spectrum, $\mathrm{SNR_{QSO}}$ is calculated as the median SNR of the 150 pixels starting from 50 pixels redward of the Lyα peak. A characteristic quantity $\mathrm{SNR_{DLA}} = \mathrm{SNR_{QSO}}/2.5$ is then defined to set the threshold of absorption to find a DLA (2.5 is an empirical choice). For each pixel j in the quasar spectrum, a sliding window of $6(1+z_j)$

pixels is used to measure the relevant $\mathrm{SNR}_{\mathrm{QSO},j}$ in this window range, where $z_j = \lambda_j / z_{\mathrm{Ly}\alpha} - 1$. This redshift-dependant sliding window is chosen to match the width of the DLA central region, where $\mathrm{SNR}_{\mathrm{QSO},j} < \mathrm{SNR}_{\mathrm{DLA}}$. This algorithm was trained and tested on synthetic quasar spectra with resolution and SNR similar to SDSS data, and reached a completeness of 100% for DLAs with $N_{\mathrm{HI}} > 10^{20.4}\mathrm{cm}^{-2}$. However, the purity of this method is limited by many false positive classifications, with a lot of BAL quasars or HCDs blended with absorption lines from Lyα clouds.

To estimate the N_{HI} of the found DLAs, a Voigt Profile (see Section 6.2.3) is used to fit the absorptions. Figure 1.10 shows an example of the Voigt profile fitting for a DLA with an estimated $logN_{\mathrm{HI}} = 22.0 \pm 0.1$ (J. X. Prochaska et Herbert-Fort 2004), where the uncertainty is obtained as the 95% confidence level interval.

This Voigt-profile fitting algorithm gives an accurate estimation of DLA redshift and column densities. However, it is strongly affected and limited by the SNR of quasar spectra. Its completeness and purity drop quickly for low SNR quasar spectra, thus not applicable for large datasets such as eBOSS data and DESI data (that has a large number of low SNR quasar spectra).

6.1.2 Machine learning approaches

In this section I will describe the machine learning methods that have been applied as an alternative method to Voigt profile fitting for the classification of DLAs, as well as the investigation of further potential methods.

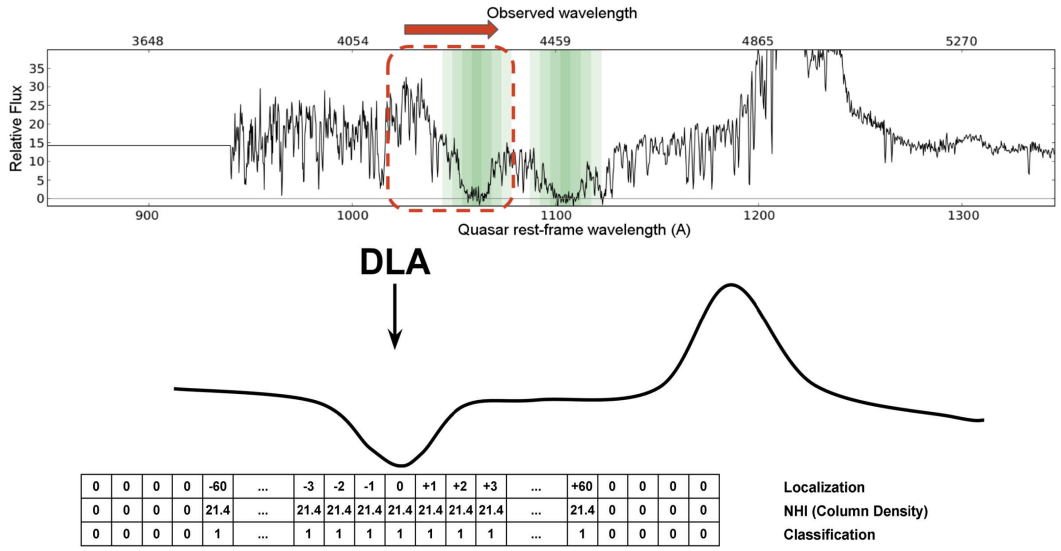

Figure 6.1 : The upper plot shows the sliding window of the CNN that slashes the quasar spectrum into pieces with equal length. The lower plot shows the multi-class labels, which define the location of redshifts (location of each pixel in the range [-60,60] compared to the DLA center), N_{HI} (0 or the value of $\log \mathrm{N_{HI}}$), and the existence of DLAs (0 or 1 for classification), respectively (Parks, J. X. Prochaska, Dong et Cai 2018).

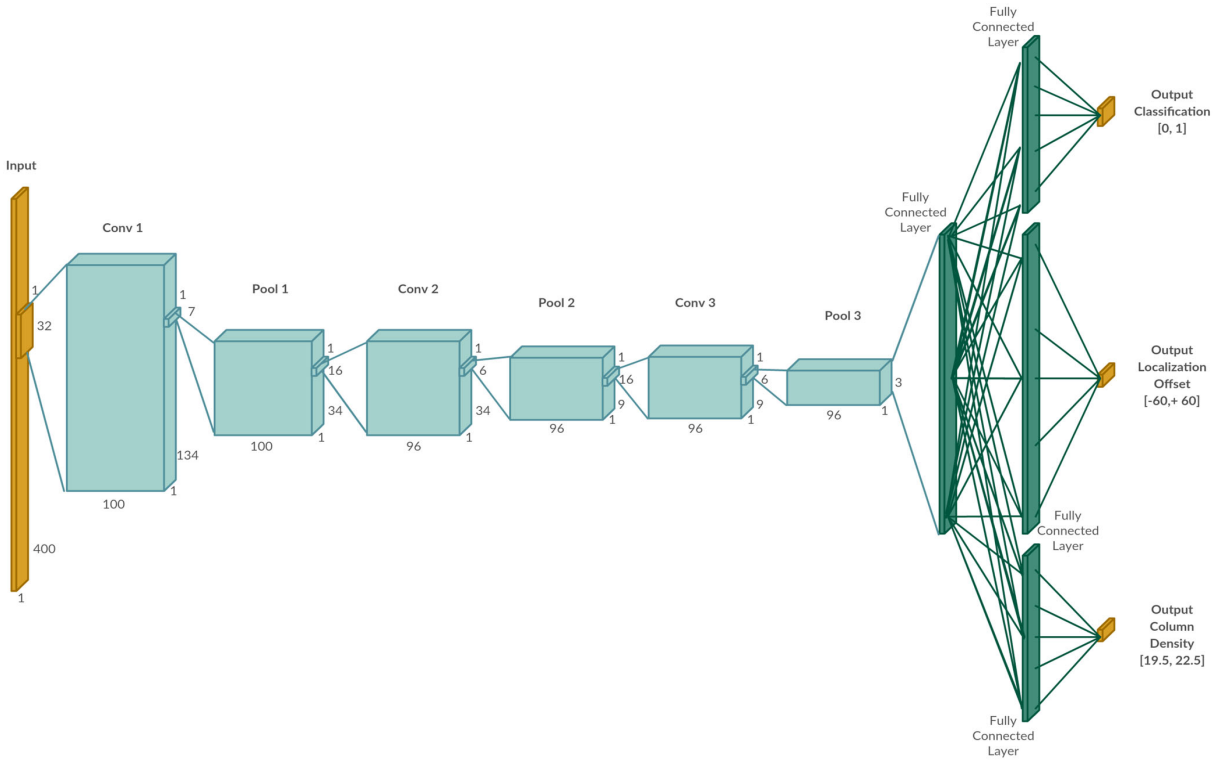

FIGURE 6.2 : Structure of the CNN used in (PARKS, J. X. PROCHASKA, DONG et CAI 2018), which treats the quasar spectra as one-dimensional images, and performs multi-classifications including z, $N_{\rm HI}$, and the existence of DLAs.

Convolutional neural networks

With the rapid increase of survey size and quasar catalog, visual inspection and Voigt profile fitting can no longer fulfill the requirements of precise estimation of both z, $N_{\rm HI}$, and the existence of DLAs. The use of convolutional neural networks (CNN), was proposed to investigate the detection of DLAs (Parks, J. X. Prochaska, Dong et Cai 2018). CNN has been widely used in cognition and classification tasks in imaging processing and time series data (Albawi, Mohammed et Al-Zawi 2017). Since quasar spectra can be seen as one-dimensional imaging data and DLAs can be detected with regular shapes and lengths, CNN is an excellent candidate for the classification of DLAs and the estimation of their physical parameters, e.g., $N_{\rm HI}$ and $z_{\rm DLA}$. Moreover, CNN is also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), where its convolution kernels capture equivalent information from each subsamples along the translation direction. This allows CNN to detect patterns with different space translations, and thus is capable of estimating the redshifts of DLAs. However, CNN usually requires the input spectra to have equal lengths, which is not the case for Lyα forests. Therefore, each quasar spectrum is slashed into a series of spectra pieces with equal length. This slashing is realized by using a sliding window of 400 pixels (to cover the wings of DLAs), and starting from each pixel of the spectrum (see the top plot in Figure 6.1). Eventually, this preprocessing step plays the same role as an additional convolutional layer, but with more precise labels of the physical parameters (z and $N_{\rm HI}$) for each slashed spectrum piece.

Figure 6.1 shows the details of the sliding window and the label definition used in (Parks, J. X. Prochaska, Dong et Cai 2018). A sliding window of 400 pixels is chosen to cover the DLA range in the Lyα forest, and a label of 120 pixels ($[-60, 60]$) is used to find the center of DLAs. Inside the DLA center pixel range, a $N_{\rm HI}$ value and a boolean value are defined to estimate the $N_{\rm HI}$ and the existence of DLAs. For each line-of-sight, i.e. each quasar spectrum, the sliding window analyzes the rest-frame wavelength range $\lambda_{\rm rf} \in [900, 1346]$Å. In eBOSS DR16 data, this yields 1748 pixels for each line-of-sight, and a 1348×400 data matrix for each spectrum (each spectrum is sliced 1348 times, each with 400 pixels). The optimized structure of the CNN used in (Parks, J. X. Prochaska, Dong et Cai 2018) is shown in Figure 6.2, with three convolutional layers (convoluting the data vector into high-dimensional parameter space), three pooling layers (down-sampling the high-dimensional parameters in order to capture the characteristic features), and several connected layers (making decisions according to these parameters and characteristic features).

In order to evaluate the performance for these multi-task classifications, a cross-entropy loss function (a function to quantify the misclassification of an algorithm) is defined to optimize the learning process. It consists of three sub loss functions relevant to classification (existence of DLAs), localization (z), and column density estimation ($N_{\rm HI}$) as follows :

$$\mathcal{L}_{\rm Total} = \mathcal{L}_{\rm c} + \mathcal{L}_{\rm z} + \mathcal{L}_{\rm n}. \tag{6.1}$$

$\mathcal{L}_{\rm c}$ is defined as the standard cross-entropy loss function, used for the classification of DLAs,

$$\mathcal{L}_{\rm c} = -y_c log(\hat{y}_c) - (1 - y_c) log(1 - \hat{y}_c). \tag{6.2}$$

Here y_c is the ground truth label : $y_c = 0$ means no DLA and $y_c = 1$ means the existence of a DLA. $\hat{y}_c$ is the model prediction for the classification in the range of (0,1), which is also called the confidence level. A critical value $C_{\rm min}$ is used to determine the classification of DLAs, where $\hat{y}_c < C_{\rm min}$ indicates a negative DLA classification, and $\hat{y}_c > C_{\rm min}$ yields a positive classification.

The localization loss function is defined as a standard square error term :

$$\mathcal{L}_z = (y_z - \hat{y}_z)^2, \tag{6.3}$$

where y_z stands for the ground truth label with $y_z \in (-60, 60)$, and $\hat{y}_z$ is the model prediction.

The loss function for $N_{\rm HI}$ is designed for $log(N_{\rm HI})$, and in the form of a square error term :

$$\mathcal{L}_n = (\frac{y_c}{y_c + \epsilon})(y_n - \hat{y}_n)^2, \tag{6.4}$$

where y_n stands for the ground truth label for $log(N_{\rm HI})$ with $y_n \in (19.5, 22.5)$ when $y_c = 1$, and $y_n = 0$ when $y_c = 0$. ϵ is a small value to ensure that the ratio $\frac{y_c}{y_c+\epsilon} = 1$ when $y_c \to 1$ and $\frac{y_c}{y_c+\epsilon} = 0$ when $y_c \to 0$.

In practice, the classification loss function $\mathcal{L}_c$ is not used since the localization task can already provide a good result. Finally, a confusion matrix (see Table 2.1) is built based on model prediction and ground truth labels to justify the classification results. The overall classification efficiency is then defined by purity and completeness, as defined in Equation 2.1.

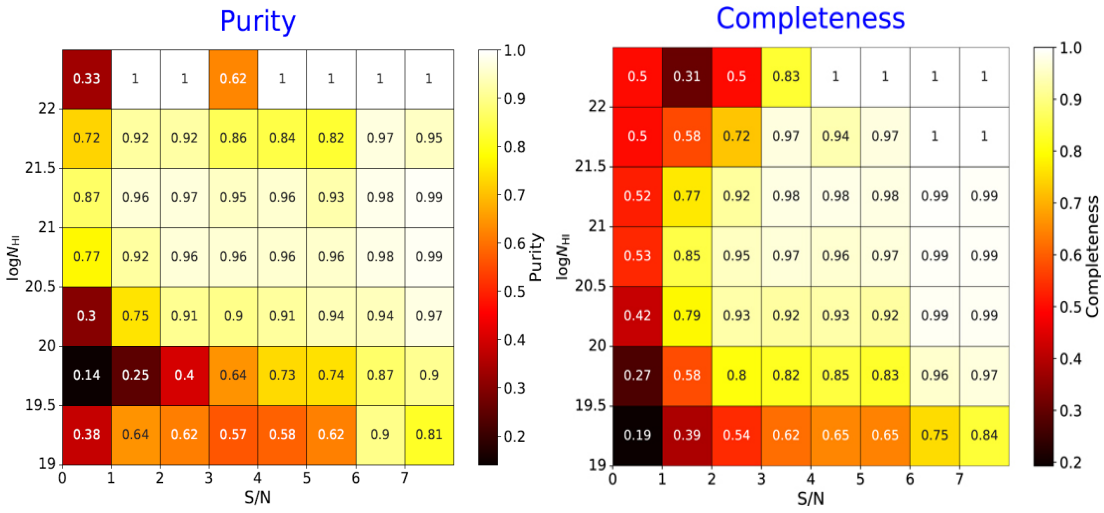

FIGURE 6.3 : The purity and completeness of the CNN classification for DLAs on DESI-Y1 mock spectra (WANG et al. 2022). The critical value $C_{\rm min}$ for the confidence level is chosen as 0.5.

The CNN DLA finder was developed in (PARKS, J. X. PROCHASKA, DONG et CAI 2018) and used to construct the eBOSS DR16 DLA calatog (CHABANIER, ETOURNEAU et al. 2022). It was then re-trained and tested for the DESI collaboration using DESI-Y1 mock spectra (WANG et al. 2022). The results of classification using these mocks are presented in Figure 6.3. These results are based on test samples with different HCD column densities ($log(N_{\rm HI})$) and the signal-to-noise-ratio S/N of quasar spectra. The estimation of $N_{\rm HI}$ and z are also shown in Figure 6.4. The critical value $C_{\rm min}$ for the confidence level is chosen as 0.5, in order to balance the purity and completeness. The results show that the model works with desirable efficiency for most of the DLAs with $20{\rm cm}^{-2} < log(N_{\rm HI}) < 22.5{\rm cm}^{-2}$, and $3 <$ S/N. It also achieves good performance for the estimation of $N_{\rm HI}$ and z with standard deviations of the estimated errors of $\sigma_{\Delta z} = 0.002$ and $\sigma_{\Delta N_{\rm HI}} = 0.17$. However, for low S/N quasar spectra and HCDs with lower $N_{\rm HI}$, the absorption due to HCDs are mixed with Lyα forests, thus are difficult to classify.

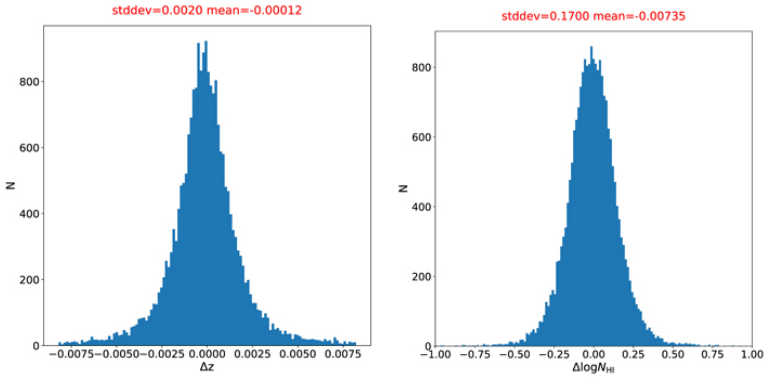

FIGURE 6.4 : The estimation of $N_{\rm HI}$ and z of the CNN model for DLAs on DESI-Y1 mock spectra, using test sample chosen with S/N > 3 and $log(N_{\rm HI}) > 20.0{\rm cm}^{-2}$ (WANG et al. 2022). The critical value $C_{\rm min}$ for the confidence level is chosen as 0.5.

Gaussian processes

Another DLA finder based on Bayesian model selection with Gaussian processes (GP) was proposed in (GARNETT, S. HO, BIRD et J. SCHNEIDER 2017; M.-F. HO, BIRD et GARNETT 2020) and applied to the eBOSS DR16 data (M.-F. HO, BIRD et GARNETT 2021). In this section, I will describe briefly the revised version of this algorithm introduced in (M.-F. HO, BIRD et GARNETT 2021). The observation data for each quasar spectrum can be defined as $\mathcal{D} = (\lambda, \mathbf{y})$, where $\lambda = \lambda_{\rm obs}/(1 + z_{\rm QSO})$ stands for the wavelength vector in the rest frame, and $\mathbf{y}$ refers to the associated observed flux vector. Suppose that a set of models $\mathcal{M}_i$ can be applied to different scenarios of the quasar spectra, e.g., no DLAs, one or more DLAs, etc. The posterior probability of a model $\mathcal{M}$, can be evaluated as a fraction of the sum of the posterior probabilities of all models, based on Bayesian theory :

$$\text{Posterior}(\mathcal{M}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M})\text{Prior}(\mathcal{M})}{\sum_i p(\mathcal{D}|\mathcal{M}_i)\text{Prior}(\mathcal{M}_i)}, \tag{6.5}$$

where $p(\mathcal{D}|\mathcal{M})$ is the model evidence of the observed data $\mathcal{D}$ given model $\mathcal{M}$, Prior($\mathcal{M}$) is the prior probability of model $\mathcal{M}$, and subscripts i denote the model evidence or prior for each scenario. In practice, several models were developed for scenarios with null DLAs ($\mathcal{M}_{0\ \rm DLA}$), with 1-4 DLAs ($\mathcal{M}_{{\rm DLA}(i)_{i=1}^{4}}$), and the model with sub-DLAs ($\mathcal{M}_{\rm sub}$, sub-DLAs have $19.5{\rm cm}^{-2} \leq \log(N_{\rm HI}) \leq 20.0{\rm cm}^{-2}$).

For each quasar spectrum, the observed data $\mathcal{D}'$ with instrumental noise ν and quasar redshift $z_{\rm QSO}$ can be defined as $\mathcal{D}' = (\lambda, \mathbf{y}, \nu, z_{\rm QSO})$. A likelihood based on a Gaussian process can then be introduced to describe the probability of this data given all observation quantities :

$$p(\mathbf{y}|\lambda, \nu, z_{\rm QSO}, \mathcal{M}_i) = \mathcal{N}(\mathbf{y}; \mu, \mathbf{\Sigma}, \mathcal{M}_i), \tag{6.6}$$

where μ is the mean and $\mathbf{\Sigma}$ is the covariance matrix of the Gaussian process, which takes into account the covariances of instrumental noise and correlations between different spectrum pixels, etc. The probability for different models $\mathcal{M}_i$ can be further derived by integrating their different associated nuisance parameters. The model evidence (probability) of the null model (only Lyα

forests and no DLAs) can be defined as

$$p(\mathcal{D}|\mathcal{M}_{0\ \mathrm{DLA}}, \nu, z_{\mathrm{QSO}}) \propto p(\mathbf{y}|\lambda, \nu, z_{\mathrm{QSO}}, \mathcal{M}_{0\ \mathrm{DLA}}) = \int p(\mathbf{y}|\lambda, \nu, z_{\mathrm{QSO}}, \theta, \mathcal{M}_{0\ \mathrm{DLA}}) p(\theta) d\theta, \tag{6.7}$$

where θ refers to all the associated nuisance parameters such as the parameters related to the redshift dependence of the Lyα forests (details see (M.-F. Ho, Bird et Garnett 2021)), and $p(\theta)$ is the prior probability of these parameters.

Based on the GP null model, the likelihood of other DLA models can be derived by applying Voigt profiles (see Section 6.2.3) in addition to the null model probability. According to Equation 6.7, the model evidence of GP DLA models can be built with an extra integration of the DLA nuisance parameters (θ_{DLA}=$\{z_{\mathrm{DLA}}, \mathrm{N_{HI}}\}$) and Voigt profile parameters θ_{Voigt} (see Section 6.2.3 and details in M.-F. Ho, Bird et Garnett 2021), i.e.,

$$p(\mathbf{y}|\lambda, \nu, z_{\mathrm{QSO}}, \mathcal{M}_{\mathrm{DLA}}) = \int p(\mathbf{y}|\lambda, \nu, z_{\mathrm{QSO}}, \theta_{\mathrm{DLA}}, \mathcal{M}_{0\ \mathrm{DLA}}) p(\theta_{\mathrm{DLA}}|z_{\mathrm{QSO}}, \mathcal{M}_{\mathrm{DLA}}) d\theta_{\mathrm{DLA}}, \tag{6.8}$$

where

$$p(\theta_{\mathrm{DLA}}|z_{\mathrm{QSO}}, \mathcal{M}_{\mathrm{DLA}}) = \int p(z_{\mathrm{DLA}}|z_{\mathrm{QSO}}, \theta_{\mathrm{Voigt}}, \mathcal{M}_{\mathrm{DLA}}) p(\mathrm{N_{HI}}|\theta_{\mathrm{Voigt}}, \mathcal{M}_{\mathrm{DLA}}) d\theta_{\mathrm{Voigt}}. \tag{6.9}$$

These probability functions (note that these posterior probabilities are different from the confidence level of the CNN finder) can then be applied to quasar spectra and associated DLA catalogs can be constructed. This GP DLA finder was applied for the DR12 and DR16 DLA catalogs (Garnett, S. Ho, Bird et J. Schneider 2017 ; M.-F. Ho, Bird et Garnett 2020 ; M.-F. Ho, Bird et Garnett 2021). It was also tested using DESI-Y1 mock spectra (Wang et al. 2022) and implemented for future DESI data. The results of classification are presented in Figure 6.5, with the same test sample used for the CNN finder, and with $log(N_{\mathrm{HI}}) > 20.0\mathrm{cm}^{-2}$ since the GP model is trained only on these DLAs. It shows an overall good performance of both purity and completeness on most DLA samples. However, the performance drops significantly for DLAs with lower N_{HI} and lower S/N, the same as the CNN DLA finder. The estimation of N_{HI} and z are shown in Figure 6.6. It achieves good performance for the estimation of N_{HI} and z with standard deviations of the estimation errors of $\sigma_{\Delta z} = 0.0016$ and $\sigma_{\Delta N_{\mathrm{HI}}} = 0.1284$.

The comparison between these two models suggests that for a range of $20.0\mathrm{cm}^{-2} < log(N_{\mathrm{HI}}) < 22.5\mathrm{cm}^{-2}$, the CNN finder yields a better purity and completeness, while the GP finder gives a better estimation of N_{HI} and z. However, the GP finder gives better purity on low S/N samples with S/N < 3. A combination of both models is then implemented for the construction of the DESI DLA catalog (see Section 6.1.3). A more direct comparison of these two finders using the eBOSS DR16 data was also performed in (M.-F. Ho, Bird et Garnett 2021) as shown in Table 6.1. It turns out that these two models have a significant disagreement in the classification of DLAs, which needs to be improved in the future.

6.1.3 DLA catalogs

In this section I describe the DLA catalogs built with the above algorithms, using the eBOSS DR16 data and the DESI data.

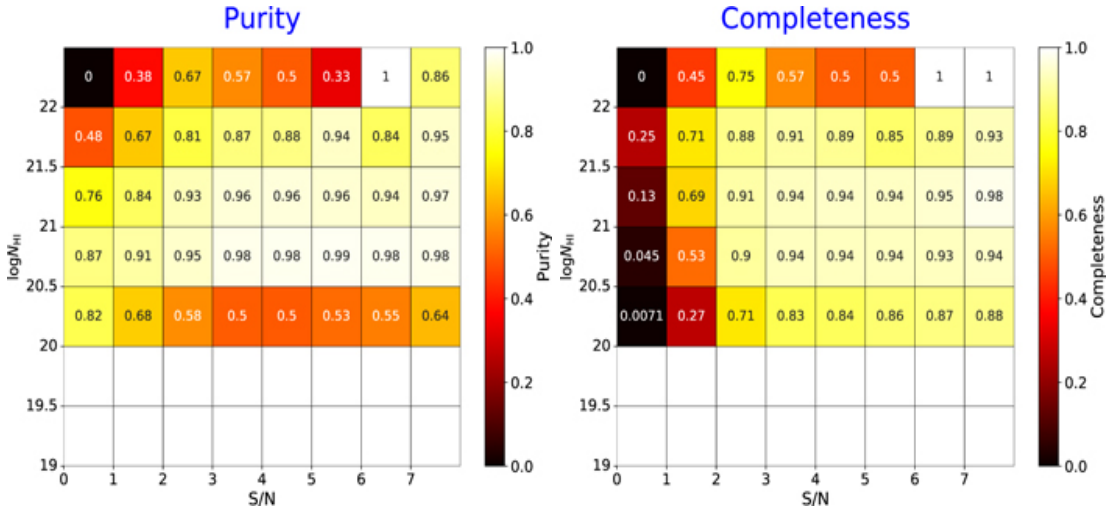

FIGURE 6.5 : The purity and completeness of the GP model classification for DLAs on DESI-Y1 mock spectra (WANG et al. 2022). The test sample is chosen with $log(N_{\rm HI}) > 20.0{\rm cm}^{-2}$ since the GP model is trained only on these DLAs.

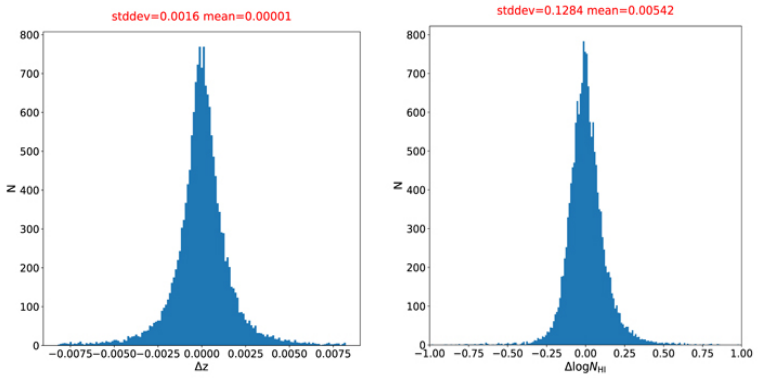

FIGURE 6.6 : The estimation of $N_{\rm HI}$ and z from the GP model for DLAs on DESI-Y1 mock spectra (WANG et al. 2022). The test sample is chosen with S/N > 3 and $log(N_{\rm HI}) > 20.0{\rm cm}^{-2}$.

GP finder \ CNN finder	0 DLA	1 DLA	2 DLA	3 DLA
0 DLA	142759	5686	93	2
1 DLA	2397	8007	208	1
2 DLA	117	234	333	5
3 DLA	8	6	11	4

TABLEAU 6.1 : Confusion Matrix of the comparison between the CNN DLA finder and the GP DLA finder, using the eBOSS DR16 dataset with $20.3\text{cm}^{-2} < log(N_{\text{HI}})$. The number $0, 1, 2, 3$ denotes the number of DLAs detected for one Lyα forest. The confidence level of the CNN and posterior probability of GP are both chosen > 0.98. Credits : M.-F. HO, BIRD et GARNETT 2021.

The eBOSS DR16 DLA catalog

The standard DLA catalog used in the eBOSS DR16 Lyα analysis was built with the CNN algorithm described in Section 6.1.2 (CHABANIER, ETOURNEAU et al. 2022). A total number of $176,807$ HCDs absorbers were found with $z_{\text{DLA}} \geqslant 2$ within $112,155$ sightlines collected from $263,201$ DR16 quasar spectra, described in Section 5.2.1. This number was reduced to $117,458$ HCDs if BAL quasars (introduced in Section 3.2.2) were rejected with BAL_PROB > 0, and a number of $57,136$ DLAs with $\log(N_{\text{HI}}) \geqslant 20.3\text{cm}^{-2}$ were discovered in this dataset.

Figure 6.7 shows the histogram of N_{HI} (x-axis in log scale, y-axis shows the number of HCDs) and z_{DLA} of the eBOSS DR16 DLA catalog, with DLA confidence level $C_{\text{min}} > 0.9$ (the purity increases with a larger C_{min}). The probability distribution of N_{HI} for these DLAs is also shown in the middle. The black curve gives the input distribution into the mocks given by the IGM physics package `pyigm` at redshift $z = 2.5$ (Described in Section 4.1.1). These plots reveal that the CNN algorithm missed a lot of high N_{HI} DLAs (due to the limited number of high N_{HI} DLAs in the data sample). It successfully detects a certain amount of high redshift DLAs, which exceeds the conservative choice in the eBOSS DR16 mocks.

The DESI DLA catalog

The DESI DLA catalogs are built using a combination of the CNN DLA finder and the GP DLA finder. As was described in the previous section, the CNN DLA finder shows higher purity and completeness, while the GP DLA finder provides a better estimation of z_{DLA} and N_{HI}. The CNN DLA finder also provides better classifications for higher redshift ($z > 4$) DLAs (JIAQI et al. in preparation). The final DLA catalog collects the DLAs that are detected by both finders. This is declared to be the case if a DLA found by the GP finder is within 800 kms^{-1} of the DLA center detected by the CNN finder. For DLAs with $z_{\text{DLA}} > 4$, only DLAs found by the CNN finder are collected. The critical value of confidence level for CNN finder is chosen as $C_{\text{min}} = 0.2$ for spectra with S/N > 3, and $C_{\text{min}} = 0.3$ for $0 < \text{S/N} < 3$ to maximize both purity and completeness. For GP finder, C_{min} is chosen as 0.9 (note that this C_{min} is the posterior probability, not the same confidence level as the CNN model), the same as M.-F. HO, BIRD et GARNETT 2021.

The final DESI EDR DLA catalog contains 9,240 DLAs with $log(N_{\text{HI}}) > 20.3$ cm^{-2}, out of 134,626 Lyα quasars with $z_{\text{QSO}} > 1.8$. This is smaller than the number of DLAs discovered in the eBOSS DR16 DLA catalog, where > 15% quasars were found hosting DLAs. This could be a consequence of the high purity of the DESI DLA finder.

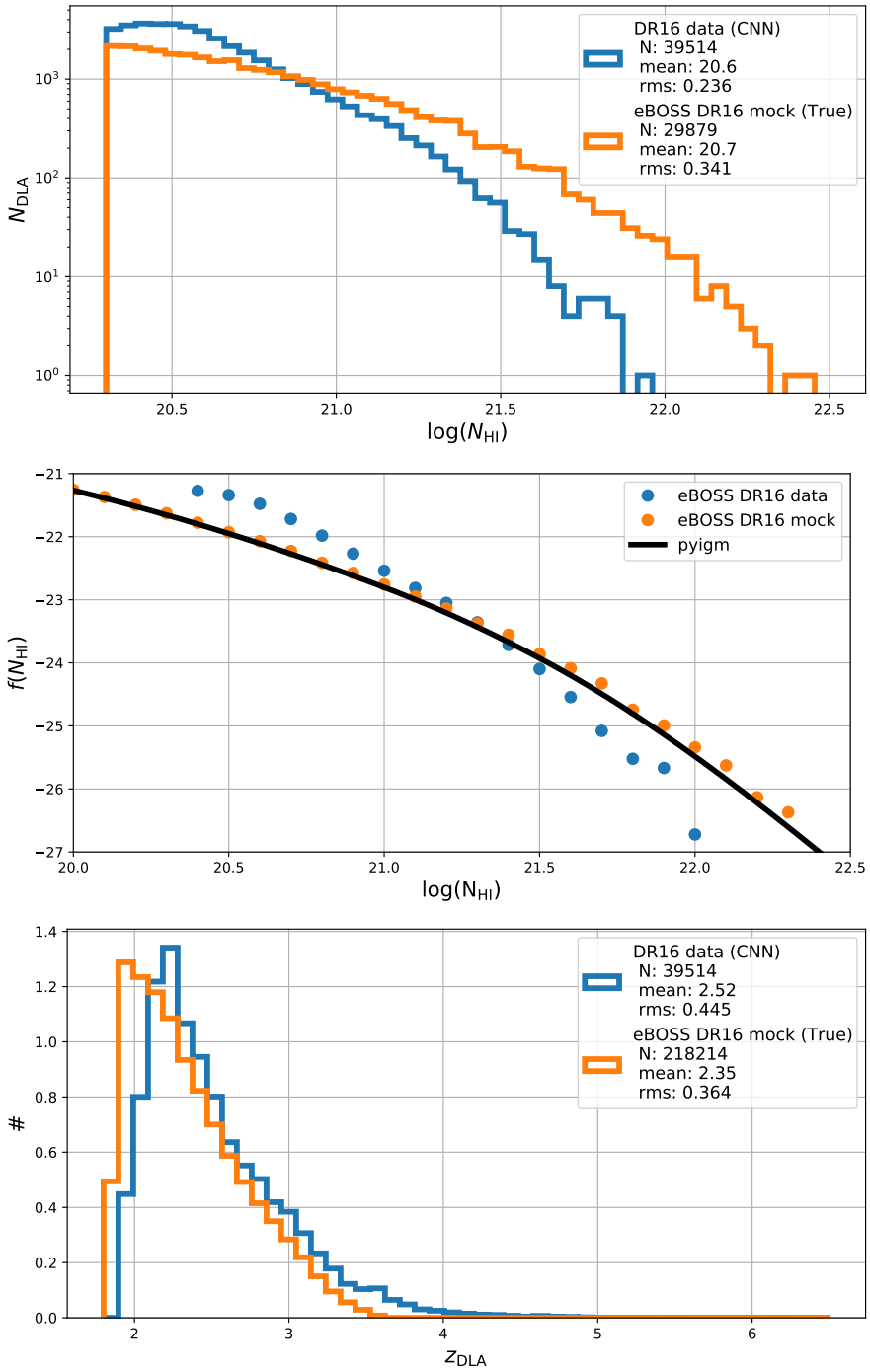

FIGURE 6.7 : The upper plot and the lower plot show the histogram of $N_{\rm HI}$ and $z_{\rm DLA}$ of the eBOSS DR16 DLA catalog, with DLA confidence level $C_{\rm min} > 0.9$. The $N_{\rm HI}$ distribution is without normalization while the $z_{\rm DLA}$ distribution is normalized to 1. The plot in the middle shows the probability distribution of $N_{\rm HI}$ for these DLAs. The black curve gives the input distribution into the mocks given by the IGM physics package `pyigm` (see Section 4.1.1).

The comparison of DLA catalogs

I hereby present a comparison of the $z_{\rm DLA}$ distribution and $N_{\rm HI}$ distribution $f(n)$ (see Equation 6.29) of these two DLA catalogs in Figure 6.8. The two plots show that these two catalogs agree well for the $z_{\rm DLA}$ distribution. However, they show a significant difference in the $f(n)$ distribution, while DESI EDR data agree well with the `pyigm` theoretical prediction (see Section 4.1.1). Since the CNN DLA finder is retrained using DESI mocks, that have more high $\rm N_{HI}$ DLAs, it is as expected to see the eBOSS DR16 DLA catalog missed some high $\rm N_{HI}$ DLAs. On the other hand, the CNN finder is trained to achieve higher purity and completeness when applying to DESI data, so the eBOSS DR16 catalog will contain more false classified DLAs.

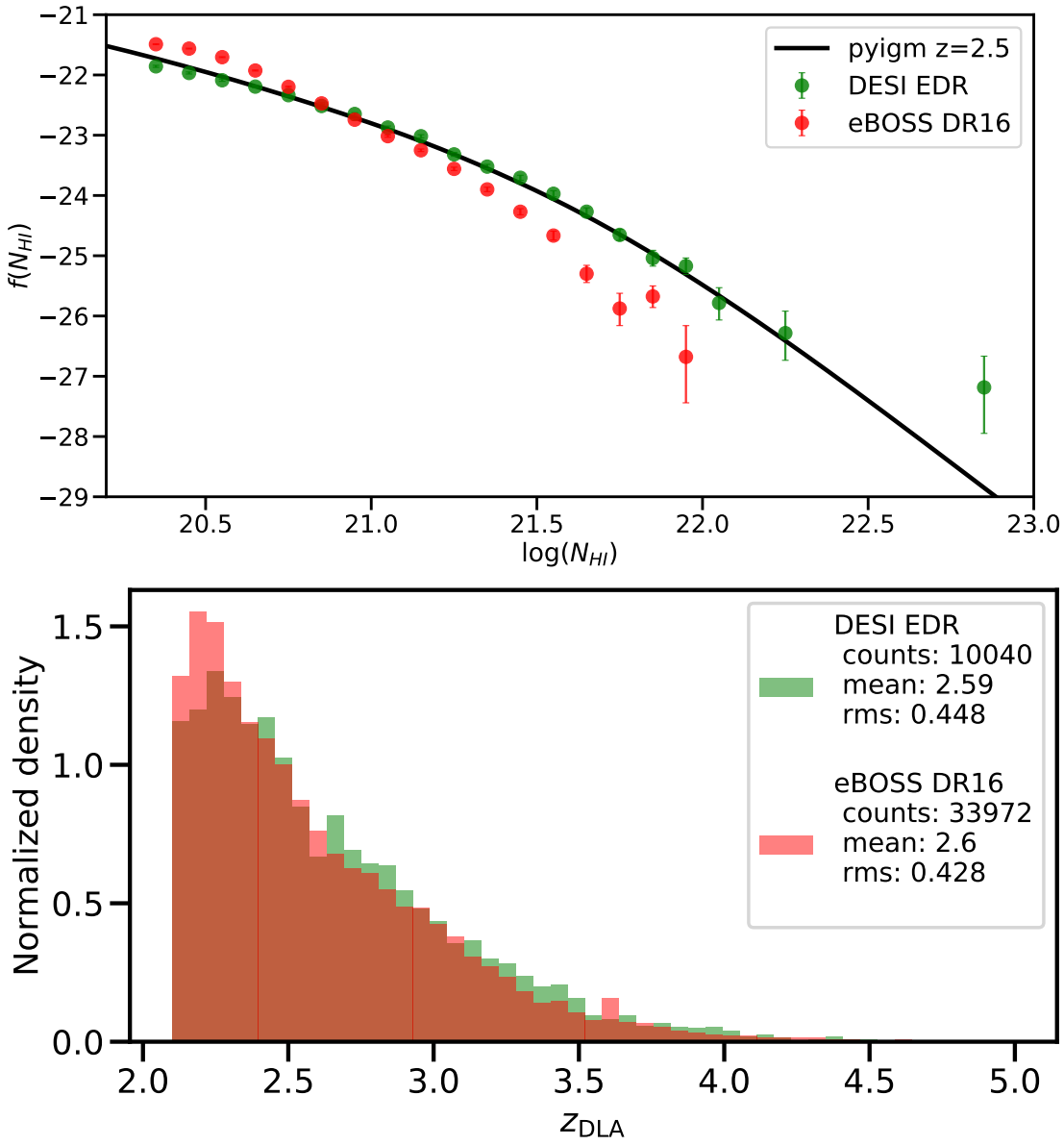

FIGURE 6.8 : The upper plot shows the probability distribution of HCD column densities $f(n)$ (see Equation 6.29) of the eBOSS DR16 DLA catalog and DESI EDR DLA catalog. The black curve gives the theoretical prediction from the IGM physics package `pyigm` (see Section 4.1.1). The vertical error bars show the Poisson uncertainty at 1 sigma. The lower plot shows a normalized comparison of the histogram of z_{DLA}.

6.2 Modeling of HCDs

Large DLAs detected using the above algorithms can be masked following the method introduced in Section 5.3. However, there remains a much larger number of small HCDs, that are undetectable and contribute to the total Lyα correlation function. In this Section, I will present the models that were used in the previous analyses (J. E. Bautista et al. 2017 ; de Sainte Agathe et al. 2019a ; Des Bourboux, Rich et al. 2020) to characterize this effect, and the new model that I developed, the so-called `Voigt` model.

6.2.1 Modeling of Lyα and HCD correlation functions

In this section, I describe the state-of-the-art modeling of the Lyα correlation function, including the damping effect caused by unmasked HCDs. The model is built following the approach proposed by (Font-Ribera et Miralda-Escudé 2012). As introduced in Section 3.2 in Equation 3.1, the flux transmission field of Lyα forests can be expressed in the form of its fluctuation,

$$F(x) = \bar{F}[1 + \delta_F(x)], \tag{6.10}$$

where x is a point in configuration space, $\bar{F}$ is the mean of F at a certain redshift, and δ_F is the fluctuation. In a simple case where the metal absorption lines are neglected, we can consider the total flux transmission as a combination of the Lyα absorption by low-density Hydrogen atoms F_α and HCD absorption $F_{\rm H}$: thus $F(x) = F_{\rm H}(x)F_\alpha(x)$. These two components can be expressed as a function of their fluctuations in a similar way :

$$\begin{aligned} F_\alpha(x) &= \bar{F}_\alpha(x)[1 + \delta_\alpha(x)] \\ F_{\rm H}(x) &= \bar{F}_{\rm H}(x)[1 + \delta_{\rm H}(x)]. \end{aligned} \tag{6.11}$$

We therefore have :

$$F(x) = \bar{F}_\alpha[1 + \delta_\alpha(x)]\bar{F}_{\rm H}[1 + \delta_{\rm H}(x)], \tag{6.12}$$

and

$$1 + \delta_F(x) = \frac{F(x)}{\bar{F}} = \frac{[1 + \delta_\alpha(x)][1 + \delta_{\rm H}(x)]}{1 + C}, \tag{6.13}$$

where $\bar{F}$ is derived as

$$\bar{F} = \langle F(x)\rangle_x = \bar{F}_\alpha\bar{F}_{\rm H}(1 + C). \tag{6.14}$$

Here $C = \langle\delta_\alpha(x)\delta_{\rm H}(x)\rangle$ is a non-zero constant that refers to the cross-correlation of Lyα forests and HCDs, at zero distance separation, which is not computable through practical numerical computation. However, according to the numerical estimation shown in Figure 6.9, C (blue curve) could be relatively small, and thus be neglected compared to the Lyα total correlations (red curve).

Lyα auto-correlation function

Averaging over all the two-point correlation pairs of two $\delta_F(x)$ at a separation $\mathbf{r_{12}} = \mathbf{x_1} - \mathbf{x_2}$, the auto-correlation of Lyα forests can be derived as

$$\begin{aligned} 1 + \xi_F(\mathbf{r_{12}}) &= \langle [1 + \delta_F(\mathbf{x_1})][1 + \delta_F(\mathbf{x_2})] \rangle \\ &= \frac{1 + 2C + \xi_\alpha(\mathbf{r_{12}}) + 2\xi_{\alpha\mathrm{H}}(\mathbf{r_{12}}) + \xi_\mathrm{H}(\mathbf{r_{12}}) + 2\xi_{\alpha\alpha\mathrm{H}}(\mathbf{r_{12}}) + 2\xi_{\alpha\mathrm{HH}}(\mathbf{r_{12}}) + \xi_{\alpha\alpha\mathrm{HH}}(\mathbf{r_{12}})}{(1 + C)^2}, \end{aligned} \tag{6.15}$$

where

$$\begin{aligned} \xi_\alpha(\mathbf{r_{12}}) &= \langle \delta_\alpha(\mathbf{x_1}) \delta_\alpha(\mathbf{x_2}) \rangle \\ \xi_{\alpha\mathrm{H}}(\mathbf{r_{12}}) &= \langle \delta_\alpha(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_2}) \rangle \\ \xi_\mathrm{H}(\mathbf{r_{12}}) &= \langle \delta_\mathrm{H}(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_2}) \rangle \\ \xi_{\alpha\alpha\mathrm{H}}(\mathbf{r_{12}}) &= \langle \delta_\alpha(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_1}) \delta_\alpha(\mathbf{x_2}) \rangle \\ \xi_{\alpha\mathrm{HH}}(\mathbf{r_{12}}) &= \langle \delta_\alpha(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_2}) \rangle \\ \xi_{\alpha\alpha\mathrm{HH}}(\mathbf{r_{12}}) &= \langle \delta_\alpha(\mathbf{x_1}) \delta_\mathrm{H}(\mathbf{x_1}) \delta_\alpha(\mathbf{x_2}) \delta_\mathrm{H}(\mathbf{x_2}) \rangle. \end{aligned} \tag{6.16}$$

Figure 6.9 shows the measurement of the above correlation function (note that $\xi_{\alpha\mathrm{H}}$ is multiplied by 5, $\xi_{\alpha\alpha\mathrm{H}}$ and ξ_H are multiplied by 10 in order to better visualize the comparison, $\xi_{\alpha\mathrm{HH}}$ and $\xi_{\alpha\alpha\mathrm{HH}}$ are not presented since they are too small to compare) using Lyα forests and HCDs in Lyα mocks (see Section 4.1.2). It turns out that the HCD correlations contribute to $\sim 20\%$ of the total correlations, mainly on small scales where $r < 80h^{-1}$Mpc. The three-point and four-point correlations are relatively small so they can be neglected in the first-order approximation. In

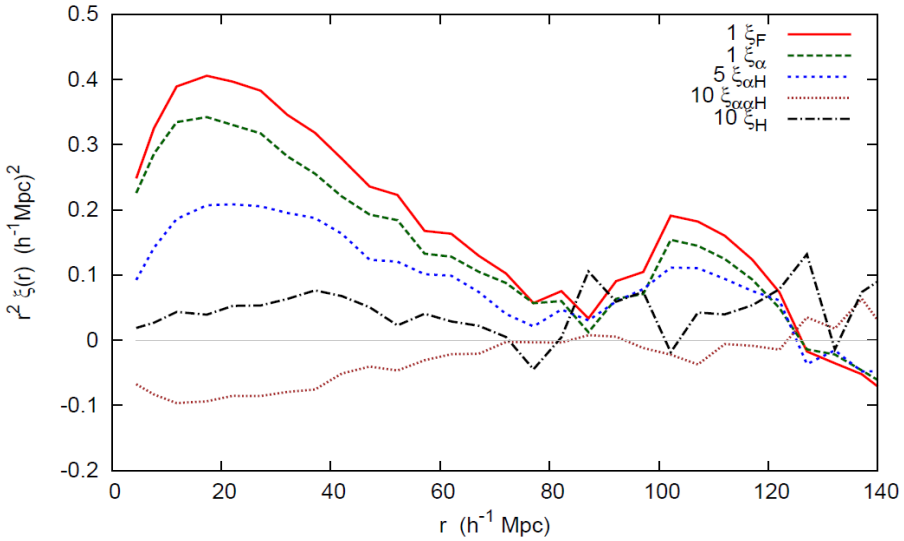

FIGURE 6.9 : The measurement of the auto(cross)-correlation function of Lyα forests and HCDs in London mocks (introduced in Section 4.1.2). Credits of this figure : (FONT-RIBERA et MIRALDA-ESCUDÉ 2012).

the following sections we only consider the two-point correlation contributions to the total Lyα

correlation function, which yields

$$\xi^F_{\mathrm{Ly}\alpha}(\mathbf{r_{12}}) = \frac{\xi_\alpha(\mathbf{r_{12}}) + 2\xi_{\alpha\mathrm{H}}(\mathbf{r_{12}}) + \xi_{\mathrm{H}}(\mathbf{r_{12}}) - C^2}{(1+C)^2} \tag{6.17}$$

For each correlation function, we take the Fourier transform to get their associated flux power spectra

$$\begin{aligned} P_{F\times F}(\vec{k}) &= \frac{P_{\mathrm{Ly}\alpha\ \times\mathrm{Ly}\alpha\ } + 2P_{\mathrm{Ly}\alpha\times\mathrm{HCD}} + P_{\mathrm{HCD}\times\mathrm{HCD}} - C^2}{(1+C)^2} \\ &= P_{\mathrm{Ly}\alpha\ \times\mathrm{Ly}\alpha\ } + 2P_{\mathrm{Ly}\alpha\times\mathrm{HCD}} + P_{\mathrm{HCD}\times\mathrm{HCD}}, \end{aligned} \tag{6.18}$$

where we neglect the constant C. As introduced in Equation 3.31, the Lyα auto-power spectrum is (neglecting the binning effect)

$$P^F_{\mathrm{Ly}\alpha\times\mathrm{Ly}\alpha}(\vec{k}) = b^2_{\mathrm{Ly}\alpha}(1+\beta_{\mathrm{Ly}\alpha}\mu_k^2)^2 P_{\mathrm{L}}(\vec{k}) D_{\mathrm{NL,Ly}\alpha}(\vec{k}). \tag{6.19}$$

We can also consider HCDs as biased tracers of the matter density field, treating them as point-like objects as quasars, while with additional absorption profiles. These absorption profiles in the flux wavelength space, parameterized by Voigt profiles (see Section 6.2.3), will result in a suppression of the flux power spectrum after the Fourier Transform. This suppression is seen at high $k_{||}$, and can be modeled by a non-linear function $F_{\mathrm{HCD}}(k_{||})$. We can then derive the HCD flux power spectrum as :

$$P_{\mathrm{HCD}\times\mathrm{HCD}}(\vec{k}) = {b^F_{\mathrm{HCD}}}^2(1+\beta_{\mathrm{HCD}}\mu_k^2)^2 P_{\mathrm{L}}(\vec{k}) D_{\mathrm{NL,HCD}}(\vec{k}) F^2_{\mathrm{HCD}}(k_{||}), \tag{6.20}$$

taking into account the non-linear effects $D_{\mathrm{NL,HCD}}$ at small scales. However, these non-linear effects of HCDs are negligible as they produce a suppression at smaller scales than $F_{\mathrm{HCD}}(k_{||})$, as explained in Section 6.5. For simplicity, we do not consider them in the following sections ($D_{\mathrm{NL,HCD}}(\vec{k}) = 1$). Different phenomenological models have been applied to characterize F_{HCD} in previous Lyα analyses, e.g., a `Sinc` function was used in the BOSS DR12 analysis (J. E. Bautista et al. 2017) :

$$F_{\mathrm{HCD}}(k_{||}) = \mathrm{sinc}(k_{||} L_{\mathrm{HCD}}), \tag{6.21}$$

and a `Exp` model was used in the eBOSS DR16 analysis (Des Bourboux, Rich et al. 2020) as :

$$F_{\mathrm{HCD}}(k_{||}) = \exp(-k_{||} L_{\mathrm{HCD}}). \tag{6.22}$$

Both models have one free parameter L_{HCD} that characterizes the suppression scale of HCDs. This parameter has a minimal impact on the BAO peak parameters (Cuceu, Font-Ribera et Joachimi 2020), thus was fixed to $L_{\mathrm{HCD}} = 10h^{-1}$Mpc in the eBOSS DR16 analysis. The physical meaning of the Exp model is explored in the next section.

The Ly$\alpha \times$ HCD flux cross-power spectrum is

$$P^F_{\mathrm{Ly}\alpha\times\mathrm{HCD}}(\vec{k}) = b_{\mathrm{Ly}\alpha} b^F_{\mathrm{HCD}}(1+\beta_{\mathrm{Ly}\alpha}\mu_k^2)(1+\beta_{\mathrm{HCD}}\mu_k^2) P_{\mathrm{L}}(\vec{k}) \sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k}) F_{\mathrm{HCD}}(k_{||}). \tag{6.23}$$

We can then derive the first-order total flux power spectrum as (Font-Ribera et Miralda-

Escudé 2012; Des Bourboux, Rich et al. 2020) :

$$\begin{aligned}
P_{\mathrm{F\times F}}(\vec{k}) &= P_{\mathrm{Ly}\alpha\times\mathrm{Ly}\alpha} + 2P_{\mathrm{Ly}\alpha\times\mathrm{HCD}} + P_{\mathrm{HCD}\times\mathrm{HCD}} \\
&= \left(b_{\mathrm{Ly}\alpha}(1+\beta_{\mathrm{Ly}\alpha}\mu_k^2)\sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k}) + b^F_{\mathrm{HCD}}(1+\beta_{\mathrm{HCD}}\mu_k^2)F_{\mathrm{HCD}}(k_{||})\right)^2 P_{\mathrm{L}}(\vec{k}) \\
&= \left(b_{\mathrm{Ly}\alpha}\sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k}) + b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})\right)^2 \\
&\qquad \left(1 + \frac{b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha}\mu_k^2\sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k}) + b^F_{\mathrm{HCD}}\beta_{\mathrm{HCD}}\mu_k^2F_{\mathrm{HCD}}(k_{||})}{b_{\mathrm{Ly}\alpha}\sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k}) + b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})}\right)^2 P_{\mathrm{L}}(\vec{k}) \\
&\approx b'^2(1+\beta'\mu_k^2)^2P_{\mathrm{L}}(\vec{k})D_{\mathrm{NL,Ly}\alpha}(\vec{k}),
\end{aligned} \tag{6.24}$$

with (considering $b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||}) \approx b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})\sqrt{D_{\mathrm{NL,Ly}\alpha}}(\vec{k})$)

$$\begin{aligned}
b'_{\mathrm{Ly}\alpha} &= b_{\mathrm{Ly}\alpha} + b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||}), \\
\beta'_{\mathrm{Ly}\alpha} &= \frac{b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b^F_{\mathrm{HCD}}\beta_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})}{b_{\mathrm{Ly}\alpha} + b^F_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})}.
\end{aligned} \tag{6.25}$$

We therefore recover the model described in Equation 3.34 and 3.31. One can figure out that the correlations contributed by HCDs are characterized by the product of $b^F_{\mathrm{HCD}}\times F_{\mathrm{HCD}}(k_{||})$. Note that here I call the bias of HCDs b^F_{HCD} for the `Exp` model since I will introduce a new model, the `Voigt` model, and use b_{HCD} as it accounts for the halo bias of HCDs. We then have an equivalence of these two models at $k_{||} = 0$:

$$b_{\mathrm{HCD}}F^{\mathrm{Voigt}}_{\mathrm{HCD}}(k_{||}=0) = b^F_{\mathrm{HCD}}F^{\mathrm{Exp}}_{\mathrm{HCD}}(k_{||}=0), \tag{6.26}$$

with

$$F^{\mathrm{Voigt}}_{\mathrm{HCD}}(k_{||}) = A\int f(n)\widetilde{W}(k_{||},n)dn. \tag{6.27}$$

Here $\widetilde{W}(k_{||})$ is the Fourier transform of a Voigt profile, b_{HCD} is the bias of HCDs where they are considered as discrete matter tracers, and is therefore positive. A is the mean number of HCDs per Lyα forest, and weighted over the whole wavelength range (we refer readers to the next section for more details), defined as :

$$A = \frac{\sum\limits_{\mathrm{HCDs}} w_\lambda}{\sum\limits_{\mathrm{Forests}} w_\lambda}. \tag{6.28}$$

Here the nominator sums over all the HCDs in the sample, and w_λ is the pixel weight (see Equation 3.9) at the center wavelength of each HCD. The denominator sums over all the Lyα forests. In our mocks where the mean number of HCDs ($n > 17.2$) per Lyα forest is around 0.3, the weighted $A \sim 0.12$ for the optical wavelength range $\lambda \in [3600, 5500]$ Å with $\Delta\lambda = 10$ Å. In Equation 6.27, $f(n)$ gives the normalized $\mathrm{H_I}$ column density probability distribution of HCDs :

$$f(n) = \frac{1}{\mathcal{N}_{\mathrm{HCD}}}\frac{d\mathcal{N}_{\mathrm{HCD}}}{dn}, \tag{6.29}$$

where $d\mathcal{N}_{\mathrm{HCD}}$ is the number of HCDs with $\mathrm{H_I}$ column densities in the range $[n, n+dn]$, $\mathcal{N}_{\mathrm{HCD}}$

is the total number of HCDs.

Note that this formula in Equation 6.27) was inspired by Font-Ribera et Miralda-Escudé 2012 and was firstly proposed in Rogers, Bird, Peiris, Pontzen, Font-Ribera et Leistedt 2018a. Simulations are performed in this work to measure the shape of the function. However, a detailed analytical derivation was not provided, and the formula missed the normalization factor A. The related HCD bias is the negative flux bias, not the halo bias of HCDs.

Lyα-QSO cross-correlation function

Considering the same first-order approximation of bias expansion as above, the cross-power spectrum between quasars and the transmitted flux fraction field can be computed with only two-point correlations : $P_{F\times\mathrm{QSO}} = P_{\mathrm{Ly}\alpha\times\mathrm{QSO}} + P_{\mathrm{HCD}\times\mathrm{QSO}}$ (see (Font-Ribera et Miralda-Escudé 2012) for the discussion of three- and four-point correlations). As was used in previous Lyα analysis (Des Bourboux, Rich et al. 2020), the modeling of the Ly$\alpha\times$QSO flux cross-power spectrum $P_{\mathrm{Ly}\alpha\times\mathrm{QSO}}$ is defined as

$$P_{\mathrm{Ly}\alpha\times\mathrm{QSO}}(\vec{k}) = b_{\mathrm{Ly}\alpha} b_{\mathrm{QSO}}(1+\beta_{\mathrm{Ly}\alpha}\mu_k^2)(1+\beta_{\mathrm{QSO}}\mu_k^2)P_{\mathrm{L}}(\vec{k})D_{\mathrm{NL,QSO}}(\vec{k}), \tag{6.30}$$

where P_{L} is the linear matter power spectrum, $D_{\mathrm{NL,QSO}}(\vec{k}) = \frac{1}{1+(k_{||}\sigma_v)^2}$ accounts for non-linear quasar peculiar velocities (W. J. Percival et White 2009), with a free parameter σ_v characterizing the quasar velocity dispersion, $b_{\mathrm{Ly}\alpha}$ (b_{QSO}) is the bias of Lyα forests (quasars), and $\beta_{\mathrm{Ly}\alpha}$ (β_{QSO}) characterises the redshift space distortion effect of Lyα forests (quasars).

The HCD $\times$ QSO flux cross-power spectrum is given by

$$\begin{aligned} P_{\mathrm{HCD}\times\mathrm{QSO}}(\vec{k}) = b_{\mathrm{HCD}}^F b_{\mathrm{QSO}}(1+\beta_{\mathrm{HCD}}\mu_k^2)(1+\beta_{\mathrm{QSO}}\mu_k^2) \\ P_{\mathrm{L}}(\vec{k})F_{\mathrm{HCD}}(k_{||}d)\sqrt{D_{\mathrm{NL,QSO}}}(\vec{k})\sqrt{D_{\mathrm{NL,HCD}}}(\vec{k}), \end{aligned} \tag{6.31}$$

where b_{HCD}^F and β_{HCD} are the bias and redshift space distortion parameters for HCDs, respectively. $F_{\mathrm{HCD}}(k_{||}d)$ describes the non-local small-scale suppression due to HCDs. $D_{\mathrm{NL,HCD}}(\vec{k})$ accounts for the non-linear effects of HCDs at small scales and will be neglected in the following equation.

Summing up Equation 6.30 and Equation 6.31, we can write the cross-power spectrum between quasars and the transmitted flux fraction field as :

$$P_{F\times\mathrm{QSO}} = b'_{\mathrm{Ly}\alpha} b_{\mathrm{QSO}}(1+\beta'_{\mathrm{Ly}\alpha}\mu_k^2)(1+\beta_{\mathrm{QSO}}\mu_k^2)P_{\mathrm{L}}(\vec{k})D_{\mathrm{NL,QSO}}(\vec{k})G(\vec{k}). \tag{6.32}$$

Here the function $G(\vec{k})$ accounts for the binning effect introduced in Equation 3.30.

Total correlation function

If we take into account the binning effect on separation grids for both the auto- and cross-correlation function, we obtain the formulas for the total flux power spectrum introduced in Equation 3.31 :

$$P_{\mathrm{Total}}(\vec{k}) = b'_i b'_j(1+\beta'_i\mu_k^2)(1+\beta'_j\mu_k^2)P_{\mathrm{L}}(\vec{k})G(\vec{k})D_{\mathrm{NL}}(\vec{k}), \tag{6.33}$$

where the indices i and j refer to different tracers : $i = j = \mathrm{Ly}\alpha$ for the Lyα auto-correlation and $i = \mathrm{Ly}\alpha$, $j = \mathrm{QSO}$ for the Lyα-quasar cross-correlation, $D_{\mathrm{NL}} = D_{\mathrm{NL,Ly}\alpha}$ for the auto-correlation and $D_{\mathrm{NL}} = D_{\mathrm{NL,QSO}}$ for the cross-correlation.

6.2.2 The L$\beta\gamma$ model

Since the `Sinc` model (see Equation 6.21) and the `Exp` model (see Equation 6.22) were built to characterize the suppression of the damping effect of HCDs at high $k_{||}$, a more general formula can be further introduced to investigate the shape of the suppression. Therefore, in this thesis, I define a new model, to the so-called L$\beta\gamma$ model, to realize this exploration :

$$F_{\rm HCD}(k_{||}) = \frac{1}{(1+(k_{||}L_{\rm HCD})^{\beta})^{\gamma}}, \tag{6.34}$$

where $L_{\rm HCD}$ is used to determine the scale of the suppression, and β and γ are two parameters used to control the slope of the non-linear function. In this case, the shape of the slope is entirely dominated by the product $\beta\times\gamma$, thus making them strongly correlated with each other. A revised version of the L$\beta\gamma$ model can be used in order to break this degeneracy :

$$F_{\rm HCD}(k_{||}) = \frac{1}{(1+(k_{||}L_{\rm HCD})^{\beta/\gamma})^{\gamma}}. \tag{6.35}$$

The L$\beta\gamma$ model is used to better fit the Lyα correlation functions, and I will present these analyses in the next section. However, the physical explanation of this model is still under study and is not included in this thesis. In this regard, I developed another analytical model, the `Voigt` model, to compare with the L$\beta\gamma$ model. I will describe this model in the next section.

6.2.3 The `Voigt` model

As mentioned above, the physical understanding of the L$\beta\gamma$ model is not clear, while giving a good fit to the Lyα correlation function (see the next section). Then I studied this non-local HCD effect from a theoretical point of view, and developed this analytical model, the `Voigt` model. In this section, I will present the theoretical basis of this model, which starts by modeling the HCD absorption profiles using Voigt profiles (an example of using this method is described in Garcia 2006).

Voigt profile

The optical depth of HCDs can be parametrized by a Voigt profile :

$$\phi(v,b,\gamma_{lu}) = \int \frac{dv}{\sqrt{2\pi}\sigma_v}\exp(-v^2/b^2)\times\frac{4\gamma_{lu}}{16\pi^2[v-(1-v/c)\gamma_{lu}]^2+\gamma_{lu}^2}. \tag{6.36}$$

It is a convolutional product of a Gaussian profile and a Lorenztian profile, corresponding to the thermal Doppler effect and the collisional cross-section of Neutral Hydrogen in the IGM, respectively (see Section 1.2.4). Here l and u represent the lower energy level and the upper energy level of the relevant electron transition, respectively. v stands for the one-dimensional relative velocity of Neutral Hydrogen atoms :

$$v = c(\frac{\lambda}{\lambda_{lu}}\frac{1}{1+z_{\rm DLA}}-1), \tag{6.37}$$

with λ the observed wavelength. The broadening effect due to thermal motion is characterized by a parameter $b = \sqrt{2}\sigma_v$, related to the standard deviation of the Gaussian profile :

$$\sigma_v = \sqrt{\frac{2kT}{m_p}}, \tag{6.38}$$

where k is the Boltzmann constant and T is the temperature of the Neutral hydrogen gas. The width of the Lorenztian profile is described by the parameter γ_{lu} :

$$\gamma_{lu} = \frac{\Gamma\lambda_{lu}}{4\pi}, \tag{6.39}$$

where the damping constant $\Gamma = 6.265 \times 10^8 s^{-1}$ for Lyman series transitions, λ_{lu} is the characteristic wavelength of Lyman series, e.g., $\lambda_{12} = 1215.6707$Å for Lyα. The relevant optical depth for each Lyα transition is :

$$\tau(\lambda; z_{\rm DLA}, N_{\rm HI}) = N_{\rm HI}\frac{\pi e^2 f_{lu}\lambda_{lu}}{m_e c}\phi(v, b, \gamma), \tag{6.40}$$

where f_{lu} is the oscillator strength of each Lyman series transition, e is the elementary charge and m_e is the electron mass.

HCD flux field

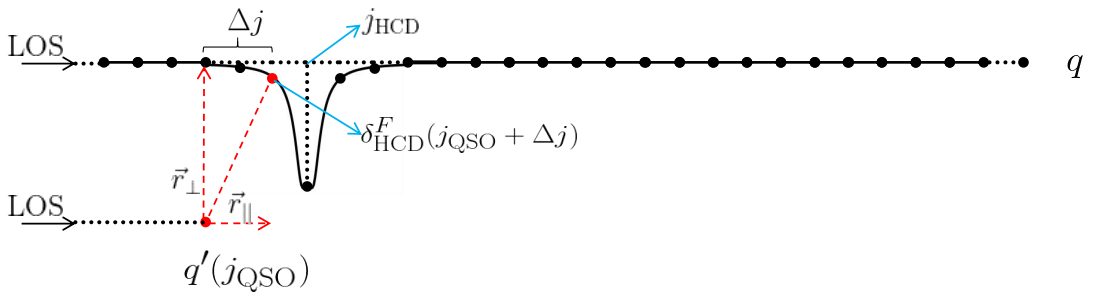

FIGURE 6.10 : The visualization of the quasar-HCD cross-correlation for two lines -of-sight defined by two quasars q and q'. The existence of an HCD centered at $j_{\rm HCD}$ gives a non-local effect for all the Lyα absorptions along this line-of-sight (the absorption wing is very vague and does not become 0 within a limited pixel range, it affects all the pixels along this line-of-sight, thus is called non-local). We compute the cross-correlation of quasar $q'(j_{\rm QSO})$ with the HCD absorptions with a separation pixel Δj.

The existence of HCDs in quasar spectra has a non-local effect for all the Lyα absorptions along a certain line-of-sight (the absorption wing is very vague and does not become 0 within a limited pixel range, it affects all the pixels along this line-of-sight, thus is called non-local), as presented in Figure 6.10. This non-local effect is modeled by considering the total flux transmission field as a product of the Lyα absorption and HCD absorption, as introduced in Equation 6.11. I will describe in this subsection the modeling of the HCD non-local effect by using Voigt profiles to parametrize the HCD absorptions.

We start by considering that HCDs are distributed in three dimensions following a number function $N_{\mathrm{HCD}}(j,i)$, which represents the number of HCDs at pixel j in the wavelength space, along the i_{th} line-of-sight. It is defined as

$$N_{\mathrm{HCD}}(j,i) = \begin{cases} 1 & \text{if HCD,} \\ 0 & \text{if no HCD.} \end{cases} \tag{6.41}$$

Based on a few intuitive assumptions, we can then model the HCD flux field.

- The flux absorption in the quasar spectrum due to HCDs is characterized by a Voigt profile in λ space, which is a convolutional product of a Gaussian profile and a Lorentzian profile, describing the thermal Doppler broadening effect and cross-section in the inter-galactic medium (IGM), respectively.
- The amount of overlapping HCDs is negligible.
- The fluctuations of $N_{\mathrm{HCD}}(j,i)$ along each line-of-sight can be expressed as

$$N_{\mathrm{HCD}}(j,i) = \langle N\rangle_{j,i}(1+\delta_{\mathrm{HCD}}(j,i)), \tag{6.42}$$

  where $\delta_{\mathrm{HCD}}(j)$ is a biased tracer of the over densities of the underlying dark matter density field with a positive bias b_{HCD}.

We then define the transmitted flux fraction field of HCDs, along the i_{th} line-of-sight as :

$$\mathrm{Flux}_{\mathrm{HCD}}(j,i,n) = \sum_{j_{\mathrm{HCD}}} V(j-j_{\mathrm{HCD}},n)N_{\mathrm{HCD}}(j_{\mathrm{HCD}},i). \tag{6.43}$$

Here j is the j_{th} pixel in the wavelength space along the i_{th} line-of-sight, $V(j-j_{\mathrm{HCD}},n)$ represents the Voigt absorption profile of an HCD with a given HI column density $n = \log_{10} N_{\mathrm{HI}}$, and the position centered at j_{HCD}.

The expectation value of the first-order fluctuations of the HCD transmission field, summing over all the HCDs along the same line-of-sight and taking into account the probability of HI column densities, can be derived as :

$$\begin{aligned} \delta^{F}_{\mathrm{HCD}}(j,i) &= \sum_{n}\Big(\frac{\mathrm{Flux}_{\mathrm{HCD}}(j,i,n)}{\langle \mathrm{Flux}_{\mathrm{HCD}}\rangle_{j,i}} - 1\Big)f(n) \\ &= \sum_{n}\sum_{j_{\mathrm{HCD}}}\Big(\frac{V(j-j_{\mathrm{HCD}},n)\langle N\rangle_{j_{\mathrm{HCD}},i}(1+\delta_{\mathrm{HCD}}(j_{\mathrm{HCD}},i))}{\langle V\circledast N\rangle_j\langle N\rangle_i} - 1\Big)f(n) \\ &= \sum_{n}\sum_{j_{\mathrm{HCD}}} W(j-j_{\mathrm{HCD}},n)\delta_{\mathrm{HCD}}(j_{\mathrm{HCD}},i)f(n), \end{aligned} \tag{6.44}$$

where the superscript F denotes the flux fluctuation, $W(j-j_{\mathrm{HCD}},n) = \frac{V(j-j_{\mathrm{HCD}},n)}{\langle V\circledast N\rangle_j}$, and $\langle V\circledast N\rangle_j \approx 1$ in our assumption that HCDs only cover a limit range of pixels compared to the entire Lyα forests. The function $f(n)$ is the HI column density probability distribution function of HCDs.

Lyα-QSO cross-correlation function

In this section I describe the cross-correlation function corresponding to the cross-power spectrum 6.2.1 between quasars and the transmitted flux fraction field, which is visualized in Figure 6.10.

For a total number of $\mathcal{N}_{\rm QSO}$ quasars ($\mathcal{N}_{\rm QSO}$ parallel lines -of-sight), the cross-correlation function between quasars and the transmitted flux fraction field with separation pixels ($r_{||} = \Delta_j$, $r_\perp$) can be given by :

$$\xi_{F\times{\rm QSO}}(r_{||} = \Delta_j, r_\perp) = \xi_{\ {\rm Ly}\alpha\times{\rm QSO}}(\Delta_j, r_\perp) + \xi_{\ {\rm HCD}\times{\rm QSO}}(\Delta_j, r_\perp). \tag{6.45}$$

Considering quasars located at pixel positions $j_{\rm QSO}$ (quasar flux field is just 1), the first term can be derived as :

$$\xi_{\ {\rm Ly}\alpha\times{\rm QSO}}(\Delta_j, r_\perp) = \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \delta^F_{{\rm Ly}\alpha}(j_{\rm QSO} + \Delta_j, i), \tag{6.46}$$

which is summed over all quasar-Lyα forest pairs with transverse separation $r_\perp$, and parallel pixel separation Δj. The second term in Equation 6.45, which is the cross-correlation of quasars and the transmitted flux fraction field of HCDs, can be derived as :

$$\begin{aligned}
\xi_{\ {\rm HCD}\times{\rm QSO}}(\Delta_j, r_\perp) &= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \delta^F_{\rm HCD}(j_{\rm QSO} + \Delta j, i) \\
&= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \sum_{j_{\rm HCD}} \sum_n \delta_{\ {\rm HCD}}(j_{\rm HCD}, i) W(j_{\rm QSO} + \Delta j - j_{\rm HCD}, n) f(n) \\
&= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \sum_{j_{\rm HCD}} \sum_n \xi^{\rm Kaiser}_{\ {\rm HCD}\times{\rm QSO}}(j_{\rm HCD} - j_{\rm QSO}, i) W(\Delta j + j_{\rm QSO} - j_{\rm HCD}, n) f(n) \\
&= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_n \sum_j \xi^{\rm Kaiser}_{\ {\rm HCD}\times{\rm QSO}}(j, i) W(\Delta j - j, n) f(n) \\
&= A \sum_n (\xi^{\rm Kaiser}_{\ {\rm HCD}\times{\rm QSO}} \circledast W)(\Delta j, i, n) f(n).
\end{aligned} \tag{6.47}$$

Here we sum over all quasar-HCD pairs with transverse separation $r_\perp$, and parallel pixel separation Δj. W is a Voigt profile defined in Equation 6.44, $\delta^F_{\rm HCD}(j, i) = \frac{F_{\rm HCD}(j,i)}{\langle F_{\rm HCD}\rangle} - 1$ is the fluctuation of the HCD flux field (we refer readers to Equation 6.44 in for more details), and $\delta_{\rm HCD}(j, i)$ is the fluctuation of the number of HCDs at each pixel position.

We apply a Fourier transform for both sides of Equation 6.46 and Equation 6.47. The first equation gives the Ly$\alpha \times$ QSO flux cross-power spectrum, that can be modeled by the standard modelization of Kaiser (Kaiser 1987) with the nonlinear effects of Lyα forests $D_{\rm NL,Ly\alpha}(\vec{k})$ (Arinyo-i-Prats, Miralda-Escudé, Viel et Cen 2015) :

$$\begin{aligned}
P_{{\rm Ly}\alpha\times{\rm QSO}}(\vec{k}) &= \sum_{\vec{r}} e^{-i(r_\perp \cdot k_\perp + k_{||} r_{||})} \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{pairs} \delta^F_{\ {\rm Ly}\alpha}(j_{\rm QSO} + \Delta_j, i) \\
&= b_{{\rm Ly}\alpha} b_{\rm QSO} (1 + \beta_{{\rm Ly}\alpha} \mu_k^2)(1 + \beta_{\rm QSO} \mu_k^2) P_{\rm L}(\vec{k}) \sqrt{D_{\rm NL,Ly\alpha}}(\vec{k}).
\end{aligned} \tag{6.48}$$

Here $\vec{k} = (k_\perp, k_{||})$, $\vec{r} = (r_\perp, r_{||})$, $r_{||} = d\Delta_j$, d is the bin width of the correlation function.

Equation 6.47 gives the HCD $\times$ QSO flux cross-power spectrum :

$$P_{\rm HCD\times QSO}(\vec{k}) = P^{\rm Kaiser}_{\rm HCD\times QSO}(\vec{k}) F^{\rm Voigt}_{\rm HCD}(k_{||}). \tag{6.49}$$

Here $F^{\rm Voigt}_{\rm HCD}(k_{||}) = A \int \widetilde{W}(k_{||}, n) f(n) dn$, as defined in Equation 6.27, $\widetilde{W}(k_{||})$ is the Fourier trans-

form of a Voigt profile, A is defined in Equation 6.28 with $A = \langle N \rangle_{j_{\rm HCD}} = \frac{\mathcal{N}_{\rm HCD}}{\mathcal{N}_{\rm QSO}}$ where $\mathcal{N}_{\rm HCD}$ and $\mathcal{N}_{\rm QSO}$ represent the total number of HCDs, QSOs (Lyα forests), respectively. In practice, a weighted $A = \frac{\sum_{\rm HCDs} w_\lambda}{\sum_{\rm Forests} w_\lambda} \approx 0.12$ (see Equation 6.28) is used to take into account the Lyα analysis pipeline variances.

$P^{\rm Kaiser}_{\rm HCD\times QSO}(\vec{k})$ is the Fourier transform of the cross-correlation of quasars and the centers of HCDs, modeled using the Kaiser formula with the nonlinear effects of HCDs $D_{\rm NL,HCD}(\vec{k})$:

$$\begin{aligned} P^{\rm Kaiser}_{\rm HCD\times QSO}(\vec{k}) &= \sum_{\vec{r}} e^{-i(r_\perp \cdot k_\perp + k_\parallel r_\parallel)} \frac{1}{\mathcal{N}_{\rm QSO}} \sum_{j_{\rm QSO}} \delta_{\rm HCD}(j_{\rm QSO} + \Delta j) \\ &= \sum_{\vec{r}} e^{-i(r_\perp \cdot k_\perp + k_\parallel r_\parallel)} \xi^{\rm Kaiser}_{\rm HCD\times QSO}(\Delta_j) \\ &= b_{\rm HCD} b_{\rm QSO}(1+\beta_{\rm HCD}\mu_k^2)(1+\beta_{\rm QSO}\mu_k^2) P_{\rm L}(\vec{k}) \sqrt{D_{\rm NL,HCD}}(\vec{k}). \end{aligned} \tag{6.50}$$

Flux-flux auto-correlation function

We present in this section the derivation of the flux-flux auto-correlation function. Inspired by (Font-Ribera et Miralda-Escudé 2012), we consider a first-order approximation of bias expansion to only take into account the two-point correlations. For a total number of $\mathcal{N}_{\rm QSO}$ quasars ($\mathcal{N}_{\rm QSO}$ parallel lines -of-sight), the auto-correlation function of the transmitted flux fraction field, as a function of the separation pixel Δ_j can be given by :

$$\xi_{F\times F}(\Delta_j) = \xi_{\ {\rm Ly}\alpha\times\ {\rm Ly}\alpha}(\Delta_j) + 2\xi_{\ {\rm Ly}\alpha\times{\rm HCD}}(\Delta_j) + \xi_{\ {\rm HCD}\times{\rm HCD}}(\Delta_j). \tag{6.51}$$

The estimator for the first term is used as the same as the eBOSS DR16 analysis (Des Bourboux, Rich et al. 2020) :

$$\xi_{\ {\rm Ly}\alpha\times\ {\rm Ly}\alpha}(\Delta_j) = \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha}) \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha} + \Delta_j), \tag{6.52}$$

with the weighted formula in Equation 3.12.

Following a similar derivation as Equation 6.47, we write the Lyα-HCD cross-correlation function as :

$$\begin{aligned} \xi_{\rm Ly\alpha\times HCD}(\Delta_j) &= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha}) \delta^F_{\ {\rm HCD}}(j_{\ {\rm Ly}\alpha} + \Delta j) \\ &= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \sum_{j_{\rm HCD}} \sum_n \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha}) \delta_{\ {\rm HCD}}(j_{\rm HCD}) W(j_{\ {\rm Ly}\alpha} + \Delta j - j_{\rm HCD}, n) f(n) \\ &= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_{\rm pairs} \sum_{j_{\rm HCD}} \sum_n \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha}) \xi^{\rm Kaiser}_{{\rm Ly}\alpha\times\ {\rm HCD}}(j_{\rm HCD} - j_{\ {\rm Ly}\alpha}) W(j_{\ {\rm Ly}\alpha} + \Delta j - j_{\rm HCD}, n) f(n) \\ &= \frac{1}{\mathcal{N}_{\rm pairs}} \sum_n \sum_{j'} \delta^F_{\ {\rm Ly}\alpha}(j_{\ {\rm Ly}\alpha}) \cdot (\xi^{\rm Kaiser}_{{\rm Ly}\alpha\times\ {\rm HCD}} \circledast W)(\Delta j - j', n) f(n) \\ &= A \sum_n (\delta^F_{\ {\rm Ly}\alpha} \circledast (\xi^{\rm Kaiser}_{{\rm Ly}\alpha\times\ {\rm HCD}} \circledast W))(\Delta j, n) f(n), \end{aligned} \tag{6.53}$$

and the HCD-HCD auro-correlation function as :

$$\begin{aligned}\xi_{\mathrm{HCD}\times\mathrm{HCD}}(\Delta_j) &= \frac{1}{\mathcal{N}_{\mathrm{pairs}}}\sum_{\mathrm{pairs}}\delta^F_{\mathrm{HCD}}(j_{\mathrm{HCD}})\delta^F_{\mathrm{HCD}}(j_{\mathrm{HCD}}+\Delta j)\\ &= A(\sum_n(\xi^{\mathrm{Kaiser}}_{\mathrm{HCD}\times\ \mathrm{HCD}}\circledast W)(\Delta j,n)f(n))^2,\end{aligned} \tag{6.54}$$

The Fourier transform of Equation 6.52, 6.53 and 6.54 will give the formulas of the power spectra defined in Equation 6.19, 6.23 and 6.20.

Physical understanding of the Exp model

Based on the `Voigt` model, we can understand the physical significance of the `Exp` model (see Equation 6.22). In this model, the absorption profiles were assumed to be Lorentzian profiles, and $f(n)$ was not taken into account. This results in a simplified formula of Equation 6.27 :

$$F^{\mathrm{Exp}}_{\mathrm{HCD}}(k_{||}) = |-\widetilde{\mathcal{L}}(L_0=0,\gamma=\frac{L_{\mathrm{HCD}}}{\pi})| = \exp(-k_{||}L_{\mathrm{HCD}}). \tag{6.55}$$

Here $\widetilde{\mathcal{L}}$ is the Fourier transform of a Lorentzian profile at a location $L_0 = 0$, and a full width at half maximum (FWHM) $\gamma = \frac{L_{\mathrm{HCD}}}{\pi}$. Since this simplified formula is an exponential function, we call it the `Exp` model. Note that in this model we use the absolute value of $\mathcal{L}$, and do not consider positions of HCDs, thus giving no normalization information to determine the HCD bias. Therefore, this model determines a negative HCD bias b^F_{HCD}.

In this regard, my theoretical development on the `Voigt` model gives a physical ground to understand the phenomenological model used in previous Lyα analyses.

Three- and four-point correlations

Similar derivations as the ones in Section 6.2.3 can also be performed for the three-point and four-point correlations introduced in Equation 6.15, yielding :

$$\begin{aligned}\xi_{\alpha\alpha\mathrm{H}}(\mathbf{r}_{12}) &= \xi_{\mathrm{Ly}\alpha\times\ \mathrm{Ly}\alpha\times\mathrm{HCD}}(\Delta_j)\\ &= \frac{1}{\mathcal{N}_{\mathrm{pairs}}}\sum_{\mathrm{pairs}}(\delta^F_{\mathrm{Ly}\alpha}(j_{\ \mathrm{Ly}\alpha}))^2\delta^F_{\mathrm{HCD}}(j_{\ \mathrm{Ly}\alpha}+\Delta j)\\ &= \frac{1}{\mathcal{N}_{\mathrm{pairs}}}\sum_{\mathrm{pairs}}\sum_{j_{\mathrm{HCD}}}\sum_n(\delta^F_{\mathrm{Ly}\alpha}(j_{\ \mathrm{Ly}\alpha}))^2\delta_{\ \mathrm{HCD}}(j_{\mathrm{HCD}})W(j_{\ \mathrm{Ly}\alpha}+\Delta j-j_{\mathrm{HCD}},n)f(n)\\ &= \frac{1}{\mathcal{N}_{\mathrm{pairs}}}\sum_{\mathrm{pairs}}\sum_{j_{\mathrm{HCD}}}\sum_n(\delta^F_{\mathrm{Ly}\alpha}(j_{\ \mathrm{Ly}\alpha}))^2\xi^{\mathrm{Kaiser}}_{\mathrm{Ly}\alpha\times\ \mathrm{HCD}}(j_{\mathrm{HCD}}-j_{\ \mathrm{Ly}\alpha})W(j_{\ \mathrm{Ly}\alpha}+\Delta j-j_{\mathrm{HCD}},n)f(n)\\ &= \frac{1}{\mathcal{N}_{\mathrm{pairs}}}\sum_n\sum_{j'}((\delta^F_{\mathrm{Ly}\alpha}(j_{\ \mathrm{Ly}\alpha}))^2\cdot(\xi^{\mathrm{Kaiser}}_{\mathrm{Ly}\alpha\times\ \mathrm{HCD}}\circledast W)(\Delta j-j',n)f(n)\\ &= A\sum_n((\delta^F_{\mathrm{Ly}\alpha})^2\circledast(\xi^{\mathrm{Kaiser}}_{\mathrm{Ly}\alpha\times\ \mathrm{HCD}}\circledast W))(\Delta j,n)f(n).\end{aligned} \tag{6.56}$$

The Fourier Transform of Equation 6.56 gives the associated bispectrum as :

$$B_{\alpha\alpha\mathrm{H}}(\vec{k}) = b^2_{\mathrm{Ly}\alpha}b_{\mathrm{HCD}}(1+\beta_{\mathrm{Ly}\alpha}\mu^2_k)^2(1+\beta_{\mathrm{HCD}}\mu^2_k)B_{\mathrm{L}}(\vec{k})D_{\mathrm{NL,Ly}\alpha}(\vec{k})F_{\mathrm{HCD}}(k_{||}), \tag{6.57}$$

where $F_{\rm HCD}(k_{||})$ is defined in Equation 6.27, and $B_{\rm L}(\vec{k})$ is the linear bispectrum of the dark matter density field.

Furthermore, we derive the other three- and four-point correlations as

$$\begin{aligned} B_{\alpha{\rm HH}}(\vec{k}) &= b_{{\rm Ly}\alpha} b_{\rm HCD}^2 (1+\beta_{{\rm Ly}\alpha}\mu_k^2)(1+\beta_{\rm HCD}\mu_k^2)^2 B_{\rm L}(\vec{k}) \sqrt{D_{{\rm NL},{\rm Ly}\alpha}}(\vec{k}) F_{\rm HCD}^2(k_{||}), \\ T_{\alpha\alpha{\rm HH}}(\vec{k}) &= b_{{\rm Ly}\alpha}^2 b_{\rm HCD}^2 (1+\beta_{{\rm Ly}\alpha}\mu_k^2)^2(1+\beta_{\rm HCD}\mu_k^2)^2 T_{\rm L}(\vec{k}) D_{{\rm NL},{\rm Ly}\alpha}(\vec{k}) F_{\rm HCD}^2(k_{||}), \end{aligned} \tag{6.58}$$

where $T_{\rm L}(\vec{k})$ is the linear trispectrum of the dark matter density field.

	Mocks with DLAs masked				Mocks without DLAs masked			
	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$QSO
Model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model
χ^2	1523.56	1522.01	3279.49	3279.05	1533.46	1533.46	3267.73	3471.22
$N_{\rm data}$	1574	1574	3148	3148	1574	1574	3148	3148
$N_{\rm par}$	7	9	8	10	7	9	8	10
P	0.78	0.78	0.04	0.04	0.72	0.71	0.05	0.0
$\alpha_{\parallel}$	1.01±0.009	1.01±0.009	1.0±0.009	1.0±0.009	1.0±0.009	1.0±0.009	1.0±0.01	1.0±0.009
$\alpha_{\perp}$	0.98±0.014	0.98±0.014	1.0±0.011	1.0±0.011	0.984±0.015	0.984±0.015	0.997±0.012	0.999±0.012
$b_{\eta,{\rm LY}\alpha}$	-0.204±0.002	-0.21±0.001	-0.179±0.021	-0.166±0.011	-0.208±0.002	-0.208±0.003	-0.192±0.003	-0.119±0.011
$\beta_{{\rm LY}\alpha}$	1.67±0.02	1.55±0.03	2.54±1.3	3.49±0.99	1.58±0.05	1.57±0.06	1.66±0.04	8.0±4.18
$b^F_{\rm HCD}$	-0.019±0.001	-0.005±0.003	-0.066±0.041	-0.091±0.017	-0.026±0.004	-0.025±0.007	-0.043±0.003	-0.129±0.009
$\beta_{\rm HCD}$	0.47±0.09	0.49±0.09	0.5±0.09	0.52±0.06	0.48±0.09	0.48±0.09	0.63±0.08	0.51±0.07
$L_{\rm HCD}$	2.29±0.75	2.37±3.62	3.86±3.02	1.07±0.38	9.48±2.57	2.19±1.73	13.02±2.18	1.36±0.15
β		1.79±1.24		0.87±0.21		6.95±5.82		3.80±0.38
γ		11.62±20.21		2.16±0.73		5.97±4.68		-5.22±0.94

TABLEAU 6.2 : Best fit parameters for eBOSS Saclay mocks with or without masking DLAs, using the $L\beta\gamma$ model and the `Exp` model, for Lyα auto-correlation function and Lyα-quasar cross-correlation, respectively.

6.3 Fitting results : The $L\beta\gamma$ model and the Exp model

In this section, we present the fitting results of the Lyα correlation function from eBOSS Saclay mocks with HCDs (see Section 4.1.1) and eBOSS DR16 data, using the $L\beta\gamma$ model.

6.3.1 Fitting results to eBOSS Saclay mocks

We summarize in Figure 6.11 and Table 6.2 the fits of the Lyα auto-correlation function and Lyα-quasar cross-correlation to a stack of 10 eBOSS Saclay mocks with HCDs, using the $L\beta\gamma$ model. These results indicate that the $L\beta\gamma$ model gives a comparable χ^2 to the `Exp` model with a free $L_{\rm HCD}$. Moreover, they show small discrepancies between the constraints for Lyα parameters, e.g., $b_{\eta,{\rm Ly}\alpha}$ and $\beta_{{\rm Ly}\alpha}$. They both suggest a small $b^F_{\rm HCD}$, which is 5-10 times smaller than $b_{{\rm Ly}\alpha} = b_{\eta,{\rm Ly}\alpha}/\beta_{{\rm Ly}\alpha}$. This validates our assumption that HCDs contribute a small correction to the total Lyα correlations. Moreover, for mocks without DLAs masked, these two models show no difference in the constraints of Lyα parameters.

6.3.2 Fitting results to eBOSS DR16 data

Figure 6.12 shows the fits of the Lyα auto-correlation function and Lyα-quasar cross-correlation to the eBOSS DR16 data, using the $L\beta\gamma$ model and the `Exp` model. Unlike the very similar fitting results of these two models on eBOSS Saclay mocks, the results on DR16 data show that the $L\beta\gamma$ model gives a slightly better fitting in the range of $20h^{-1}{\rm Mpc} < r < 80h^{-1}{\rm Mpc}$, than the `Exp` model with a free $L_{\rm HCD}$ or a fixed $L_{\rm HCD} = 10h^{-1}{\rm Mpc}$. The curves of the $L\beta\gamma$ model go through the points very well for $0.8 < \mu < 1$, where the HCDs effect has an important impact. The numerical fits are summarized in Table 6.3, from which we can conclude that :

- The $L\beta\gamma$ model gives better constraints on $b_{\eta,{\rm LY}\alpha}$ and $\beta_{{\rm LY}\alpha}$, while giving smaller χ^2 for the fitting. If we compare to the eBOSS DR16 analysis (DES BOURBOUX, RICH et al. 2020) where $L_{\rm HCD}$ was fixed to $10h^{-1}$Mpc, this improvement is obvious.
- Different HCD models show no influence on the constraints on BAO parameters, i.e., $\alpha_{\parallel}$ and $\alpha_{\perp}$. This matches our expectations since the BAO scale ($\sim 100h^{-1}$Mpc) is beyond the HCD scale.

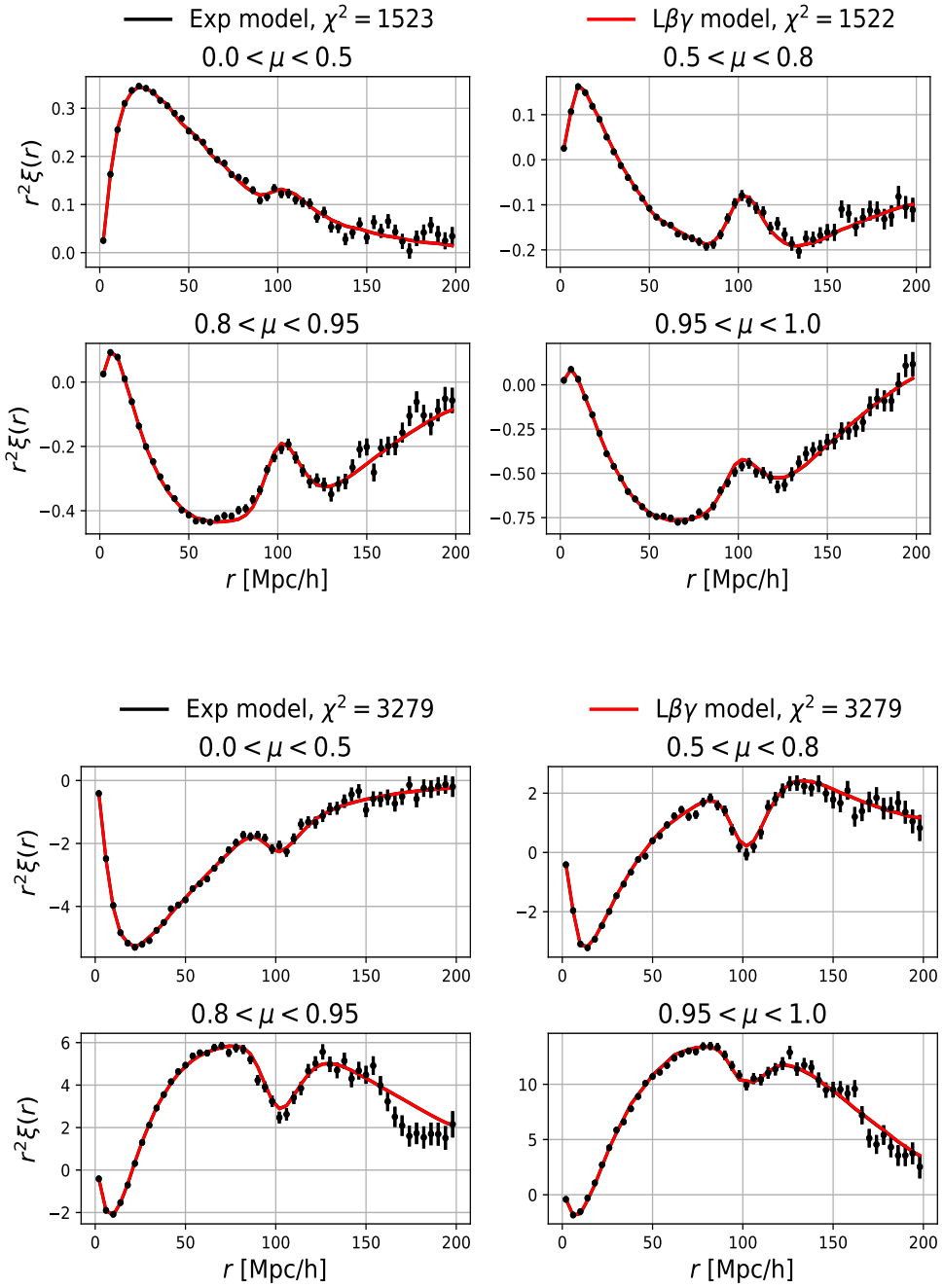

FIGURE 6.11 : eBOSS Saclay mocks : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for pixels in the Lyα region. The correlations are multiplied by r^2 to better see the BAO peak. The black curves (fully overlapped with red curves) show the best-fit models using the Exp model. The red curves give the best-fit models using the $L\beta\gamma$ model, in four wedges of $|\mu| = |\frac{r_\parallel}{r}|$. The fitted range is chosen as $r \in [20, 180]h^{-1}$Mpc.

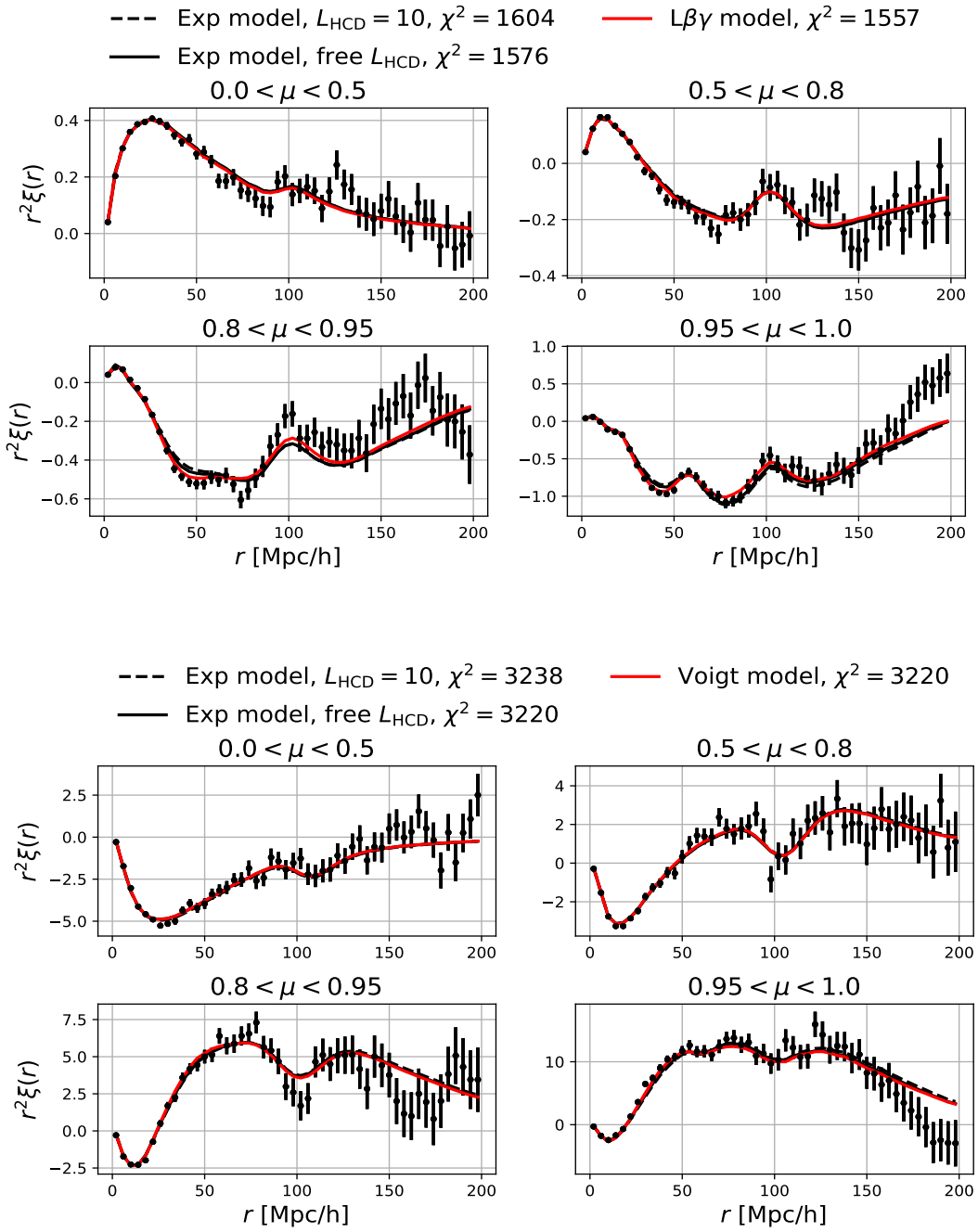

FIGURE 6.12 : eBOSS DR16 data : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for pixels in the Lyα region. The correlations are multiplied by r^2 to better see the BAO scale. The black curves show the best-fit models using the **Exp** model, with a free $L_{\rm HCD}$ (solid) or a fixed $L_{\rm HCD} = 10h^{-1}$Mpc (dashed). The red curves give the best-fit models using the $L\beta\gamma$ model, in four wedges of $|\mu| = |\frac{r_\parallel}{r}|$. The fitted range is chosen as $r \in [10, 180]h^{-1}$Mpc.

	Data with DLAs masked				Data without DLAs masked			
	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$QSO	LY$\alpha \times$LYα	LY$\alpha \times$LYα	LY$\alpha \times$QSO	LY$\alpha \times$QSO
Model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model	Exp model	$L\beta\gamma$ model
χ^2	1576.22	1557.3	3220.27	3220.41	1594.96	1576.2	3219.44	3213.87
$N_{\rm data}$	1590	1590	3180	3180	1590	1590	3180	3180
$N_{\rm par}$	14	16	13	13	14	16	13	15
P	0.49	0.61	0.25	0.25	0.36	0.48	0.25	0.27
$\alpha_{\parallel}$	1.05±0.034	1.04±0.034	1.06±0.032	1.05±0.032	1.04±0.034	1.03±0.034	1.05±0.034	1.05±0.035
$\alpha_{\perp}$	0.981±0.042	0.99±0.042	0.932±0.039	0.935±0.039	0.974±0.044	0.984±0.044	0.948±0.042	0.946±0.043
$b_{\eta,{\rm LY}\alpha}$	-0.175±0.013	-0.175±0.008	-0.228±0.016	-0.208±0.01	-0.173±0.013	-0.167±0.01	-0.231±0.019	-0.25±0.014
$\beta_{{\rm LY}\alpha}$	3.23±1.26	2.87±0.4	1.91±0.33	2.13±0.16	5.25±3.29	6.05±2.26	1.92±0.34	1.77±0.16
$b^F_{\rm HCD}$	-0.105±0.022	-0.088±0.009	-0.034±0.024	-0.05±0.0	-0.139±0.02	-0.13±0.01	-0.047±0.027	-0.02±0.009
$\beta_{\rm HCD}$	0.53±0.08	0.54±0.07	0.52±0.09	0.7±0.0	0.51±0.08	0.52±0.07	0.51±0.09	0.51±0.09
$L_{\rm HCD}$	2.28±0.63	13.47±1.11	0.95±2.87	12.33±2.10	2.59±0.52	14.52±0.99	-0.01±1.66	7.33±0.98
β		138.24±71.03		60.59±165.80		114.74±44.51		108.67±93.42
γ		0.003±0.001		0.011±0.032		0.003±0.001		0.151±0.288

TABLEAU 6.3 : Best fit parameters for eBOSS DR16 data, using the $L\beta\gamma$ model and the `Exp` model, for the Lyα auto-correlation function and the Lyα-quasar cross-correlation, respectively.

- The two parameters determining the shape of the slope in the $L\beta\gamma$ model, β and γ, are very badly constrained and show a very strong correlation. This could be further investigated, and tested with other extensions of the model as, for example, suggested by Equation 6.35.

6.4 Fitting results : The `Voigt` model

In this section we summarize the fittings of the Lyα correlations using the `Voigt` model, and compare them with the `Exp` model used in the eBOSS DR16 analysis. The fittings are performed on eBOSS DR16 data and different types of Saclay mocks described in section 4.4, each with a stack of 10 mocks.

6.4.1 Fitting results of eBOSS Saclay mocks

We present the fitting of the auto- and cross-correlation functions for Lyα mocks in Figure 6.13 and the prediction of the parameters, $\{b_{\rm HCD}, b_{\rm HCD}{*}\beta_{\rm HCD}\}$, in Figure 6.14. Table 6.4 and Table 6.5 show the results for all the fitted parameters. The fitted range is chosen as $r \in [20, 180]h^{-1}$Mpc. The figures and tables show that :

- The `Voigt` model gives good fitting for different types of Saclay mocks, going well through the data points, except in the case where the mocks have HCDs with $n = 21.0$. This may be due to the distortions of the quasar continuum fitting in the Lyα analysis pipeline, as HCDs with such wide damping absorption wings will have a significant distortion on the continuum fitting, and then on the Lyα correlation function.
- Regarding the Lyα auto-correlation function, the `Voigt` model gives good constraints on the predicted parameters $b_{\rm HCD}$ and $\beta_{\rm HCD}$ (input $b_{\rm HCD} = 2$ and $\beta_{\rm HCD} = 0.5$, these values are used to match the measured DLA bias in Pérez-Ràfols, Miralda-Escudé, Arinyo-i-Prats, Font-Ribera et Mas-Ribas 2018), within 2σs for all the cases except the $n = 21$ case.
- Regarding the Lyα auto-correlation function, the `Voigt` model provides good predictions for the Lyα bias and redshift distortion parameter, with a $\beta_{{\rm Ly}\alpha}$ around 1.6, comparable with the results from Lyα simulations. Besides, looking at the constraints on $b_{\eta,{\rm LY}\alpha}$ and $\beta_{{\rm Ly}\alpha}$ from auto- and cross-correlation functions, the results obtained using `Voigt` model are also in much better agreement compared to other two models.
- The modeling of HCDs does not affect the BAO peak position. This is as expected, as the HCDs mostly affect the scales below the BAO scale.

Comparison with the Exp model

We make a comparison between the `voigt` model and the `Exp` model, in the mocks generated with a distribution $f(n)$ of HCDs (eboss-0.2 mocks, see Section 4.4). As the large DLAs are detectable and maskable, we mask out the DLAs with $n > 20.3$ as a further comparison with the no masking case. Numerical fits are shown in Table 6.5, and the comparison of the constraints on $b'\beta' = b_{{\rm Ly}\alpha}\beta_{{\rm Ly}\alpha} + b_{\rm HCD}\beta_{\rm HCD}F_{\rm HCD}(k_{||})$ fitted from Lyα auto-correlations and Lyα-quasar cross-correlation function are shown in Figure 6.15 and Figure 6.16, respectively. For the `Exp` model, we take into account the constraints of one more free parameter, $L_{\rm HCD}$. These results indicate that :

- The `voigt` model and the `Exp` model present comparable fitting results, i.e., similar χ^2, and parameter values $\{\alpha_{||}, \alpha_{\perp}, b_{\eta,{\rm Ly}\alpha}, \beta_{{\rm Ly}\alpha}, \beta_{\rm HCD}\}$. However, the `voigt` model gives a physical measurement of $b_{\rm HCD}$, that can not be constrained from the `Exp` model. The auto- and cross-correlation functions results for $b_{\eta,{\rm Ly}\alpha}$ and $\beta_{{\rm Ly}\alpha}$ are also more consistent with the `Voigt` model.

	Mocks with HCDs with same n=log $N_{\rm HI}$							
	LY$\alpha\times$LYα	LY$\alpha\times$LYα	LY$\alpha\times$LYα	LY$\alpha\times$LYα	LY$\alpha\times$QSO	LY$\alpha\times$QSO	LY$\alpha\times$QSO	LY$\alpha\times$QSO
n	19.5	20.0	20.5	21.0	19.5	20.0	20.5	21.0
Model	Voigt model	Voigt model	Voigt model	Voigt model	Voigt model	Voigt model	Voigt model	Voigt model
$N_{\rm mocks}$	10	10	10	10	10	10	10	10
χ^2	1583.58	1584.23	1583.63	1682.56	3204.53	3236.97	3261.56	3286.47
$N_{\rm data}$	1574	1574	1574	1574	3148	3148	3148	3148
$N_{\rm par}$	6	6	6	6	7	7	7	7
P	0.39	0.38	0.39	0.02	0.21	0.11	0.07	0.03
$\alpha_{\parallel}$	1.01±0.009	1.01±0.009	1.01±0.01	1.0±0.01	1.0±0.009	1.0±0.01	1.0±0.01	0.998±0.01
$\alpha_{\perp}$	0.998±0.013	1.0±0.015	1.0±0.017	1.01±0.019	1.0±0.011	1.0±0.012	1.01±0.013	1.01±0.014
$b_{\eta,{\rm LY}\alpha}$	-0.206±0.002	-0.204±0.003	-0.201±0.003	-0.187±0.003	-0.193±0.001	-0.192±0.001	-0.192±0.001	-0.192±0.001
$\beta_{{\rm LY}\alpha}$	1.61±0.04	1.6±0.05	1.57±0.06	1.33±0.05	1.66±0.02	1.68±0.02	1.69±0.02	1.68±0.02
$b_{\rm HCD}$	1.95±0.292	2.13±0.157	2.2±0.092	1.92±0.052	2.15±0.15	2.24±0.094	2.25±0.064	2.16±0.048
$\beta_{\rm HCD}$	0.48±0.09	0.48±0.08	0.5±0.07	0.75±0.05	0.37±0.07	0.4±0.06	0.46±0.05	0.59±0.05

TABLEAU 6.4 : Best fit parameters for stack of Lyα mocks (eboss-0.2+ mocks, see Section 4.4) created with HCDs with the same column densities n = 19.5, 20.0, 20.5, 21.0, using the `Voigt` model, for Lyα auto-correlation function and Lyα-quasar cross-correlation, respectively.

	Mocks with DLAs masked				Mocks without DLAs masked			
	LY$\alpha\times$LYα	LY$\alpha\times$LYα	LY$\alpha\times$QSO	LY$\alpha\times$QSO	LY$\alpha\times$LYα	LY$\alpha\times$LYα	LY$\alpha\times$QSO	LY$\alpha\times$QSO
Model	Exp model	Voigt model	Exp model	Voigt model	Exp model	Voigt model	Exp model	Voigt model
$N_{\rm mocks}$	10	10	10	10	10	10	10	10
χ^2	1523.56	1523.22	3279.49	3378.02	1533.46	1534.1	3267.73	3331.4
$N_{\rm data}$	1574	1574	3148	3148	1574	1574	3148	3148
$N_{\rm par}$	7	6	8	7	7	6	8	7
P	0.78	0.79	0.04	0.0	0.72	0.73	0.05	0.01
$\alpha_{\parallel}$	1.01±0.009	1.01±0.009	1.0±0.009	1.0±0.009	1.0±0.009	1.0±0.009	1.0±0.01	1.0±0.009
$\alpha_{\perp}$	0.98±0.014	0.98±0.013	1.0±0.011	1.0±0.011	0.984±0.015	0.984±0.015	0.997±0.012	0.998±0.012
$b_{\eta,{\rm LY}\alpha}$	-0.204±0.002	-0.206±0.001	-0.179±0.021	-0.192±0.001	-0.208±0.002	-0.205±0.002	-0.192±0.003	-0.191±0.001
$\beta_{{\rm LY}\alpha}$	1.67±0.02	1.64±0.04	2.54±1.3	1.66±0.02	1.58±0.05	1.64±0.04	1.66±0.04	1.68±0.02
$b^F_{\rm HCD}$	-0.019±0.001		-0.066±0.041		-0.026±0.004		-0.043±0.003	
$b_{\rm HCD}$		1.93±0.347		2.21±0.215		1.82±0.146		2.06±0.114
$\beta_{\rm HCD}$	0.47±0.09	0.48±0.09	0.5±0.09	0.38±0.08	0.48±0.09	0.48±0.08	0.63±0.08	0.41±0.07
$L_{\rm HCD}$	2.29±0.75		3.86±3.02		9.48±2.57		13.02±2.18	

TABLEAU 6.5 : Best fit parameters for stack of Lyα mocks (eboss-0.2 mocks, see Section 4.4), using the `Voigt` model and the `Exp` model, for Lyα auto-correlation function and Lyα-quasar cross-correlation, respectively.

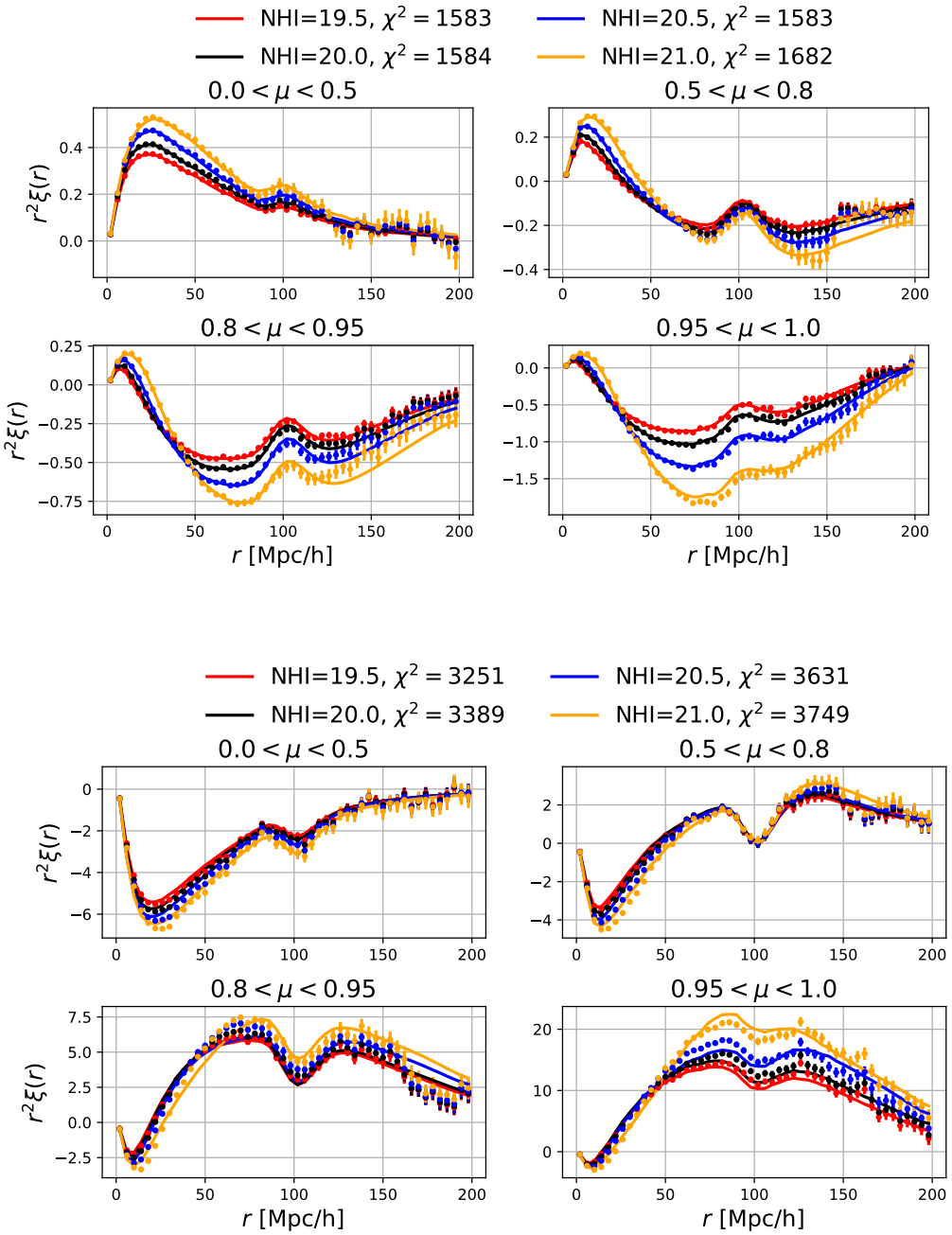

FIGURE 6.13 : Lyα mocks : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for pixels in the Lyα region. The correlations are multiplied by r^2 to better see the BAO peak. The points of different colors give the measured correlation for mocks with HCDs with different column densities $n = 19.5, 20, 20.5, 21$, and the curves give the best fit models using the Voigt model, in four wedges of $|\mu| = |\frac{r_\parallel}{r}|$. The fitted range is chosen as $r \in [20, 180]h^{-1}$Mpc.

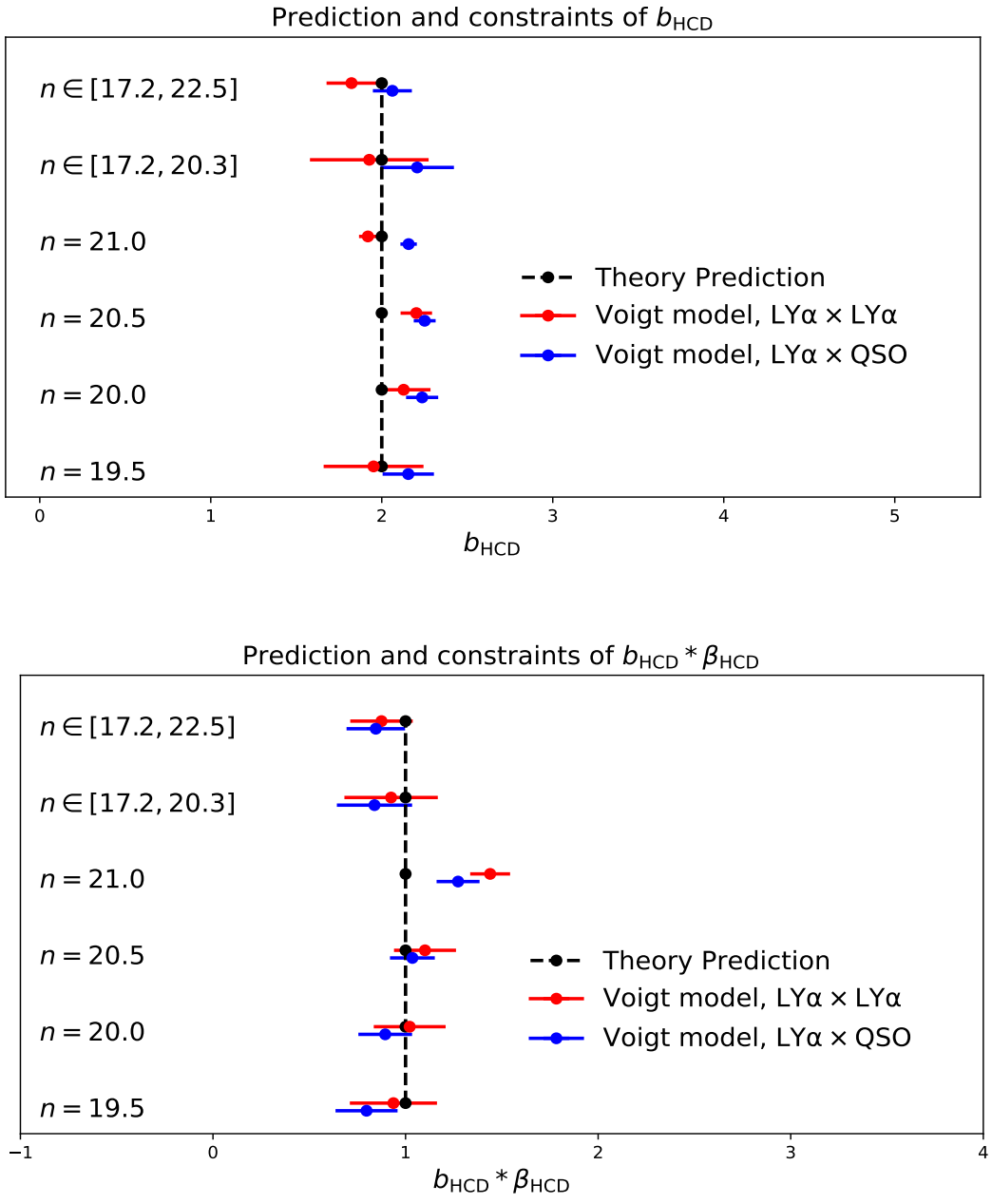

FIGURE 6.14 : Theoretical prediction and experimental constraints for b_{HCD} and $b_{\mathrm{HCD}} * \beta_{\mathrm{HCD}}$ using the Voigt model. The black points show the true value of $b_{\mathrm{HCD}} = 2.0$ and $b_{\mathrm{HCD}} * \beta_{\mathrm{HCD}} = 1.0$. The red points and blue points are fitted from auto- and cross-correlations respectively, for mocks with HCDs with different column densities. The bottom four experiments use mocks with the same column density HCDs, and the top two experiments use a distribution of HCDs, described in 4.4.

- The comparison of constraints on $b'\beta' = b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}}\beta_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})$ shows that the `voigt` model gives tighter constraints, while the worse constraint for $n \in [17.2, 20.3]$ may be due to the smaller number of HCDs and tinier effect on Lyα correlation function.

Correlation between parameters

In order to understand the parameter constraints of different models and their correlations, we compute the Gaussian likelihood with Gaussian distributed parameters, as described in Section 5.2.4. We only use mocks with realistic HCD distributions (with/without masking) and the Lyα auto-correlation function. We fix β_{HCD} since it is mainly determined by priors. The effective Lyα bias is used, which is less correlated with $\beta_{\mathrm{Ly}\alpha}$. It is defined as :

$$b_{\mathrm{eff,Ly}\alpha} = b_{\mathrm{Ly}\alpha}\sqrt{1 + \frac{2}{3}\beta_{\mathrm{Ly}\alpha} + \frac{1}{5}\beta^2_{\mathrm{Ly}\alpha}}. \tag{6.59}$$

Moreover, it gives the amplitude of the monopole of the correlation function. The combination ($b_{\mathrm{eff,Ly}\alpha}$,$\beta_{\mathrm{Ly}\alpha}$) then separates clearly the isotropic part ($b_{\mathrm{eff,Ly}\alpha}$) from the anisotropic part ($\beta_{\mathrm{Ly}\alpha}$). Figure 6.17 shows the triangle plot of the Lyα parameters of interest $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha},|b^F_{\mathrm{HCD}}|,L_{\mathrm{HCD}}\}$ for the `Exp` model, and Figure 6.18 shows the constraints on $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha},|b_{\mathrm{HCD}}|\}$ for the `Voigt` model. A comparison of these two models is summarized in Figure 6.19 and Figure 6.20, showing the constraints on $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha}\}$ and $\{|b_{\mathrm{eff,LY}\alpha}|,|b_{\eta,\mathrm{LY}\alpha}|\}$. We can infer from these results that :

- The `Voigt` model gives tighter constraints on both $b_{\mathrm{eff,Ly}\alpha}$, $b_{\eta,\mathrm{LY}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$. The correlation between $b_{\mathrm{eff,Ly}\alpha}$ and $b_{\eta,\mathrm{LY}\alpha}$ is much weaker compared to the `Exp` model.
- b^F_{HCD} and L_{HCD} are strongly anti-correlated in the `Exp` model, indicating that the important factor is the product of these two parameters. We get rid of this issue in the `Voigt` model since the amplitude information of the HCD power spectrum is analytically computed.

6.4.2 Fitting of eBOSS DR16 data

I present in Figure 6.21 the fitting of the auto- and cross-correlation functions to the eBOSS DR16 data, using the `voigt` model and the `Exp` model. Table 6.6 show the results for all the fitted parameters, and the masking of large DLAs is also taken into account as a control group.

- The `voigt` model measures $b_{\eta\mathrm{Ly}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$ with a smaller uncertainty than the `Exp` model, for both the Lyα auto-correlation and the Lyα-quasar cross-correlation functions. There is good consistency between auto- and cross-correlation functions for $\beta_{\mathrm{Ly}\alpha}$, while not for $b_{\eta\mathrm{Ly}\alpha}$.
- The two models of HCDs do not affect the measurement of BAO significantly.
- Using the Lyα auto-correlation function from eBOSS DR16 data, we measure a much larger b_{HCD} than what we have in the mocks. This could be due to the fact that the input HI column density distribution in the `Voigt` model is not realistic enough, since it is fitted from a limited number of DLAs with a wide range of HI column densities. However, it also suggests that the model could potentially be used to constrain the HI column density distribution of HCDs in the range of $17 < n < 20$, which is technically hard to measure from direct observation.

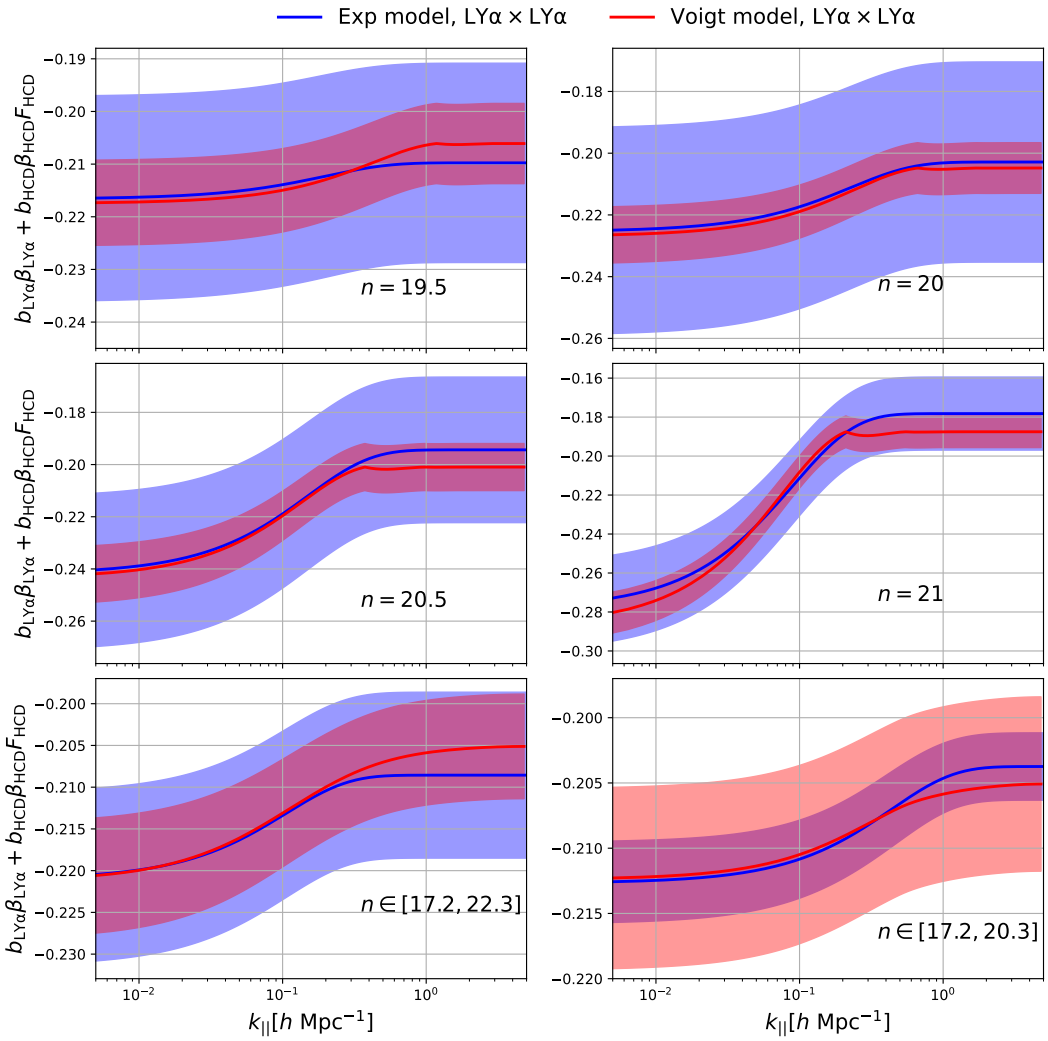

FIGURE 6.15 : The comparison between the `voigt` model and the `Exp` model for `Saclay` mocks with different types of HCDs. The upper four plots show the comparison in mocks with HCDs with the same column densities, from 19.5 to 21. The lower two plots show the mocks with a distribution of HCDs. The blue curves and red curves show the 1σ constraints on $b'\beta' = b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}}\beta_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})$, of the `Exp` model and the `Voigt` model, respectively, fitted using the Lyα auto-correlations.

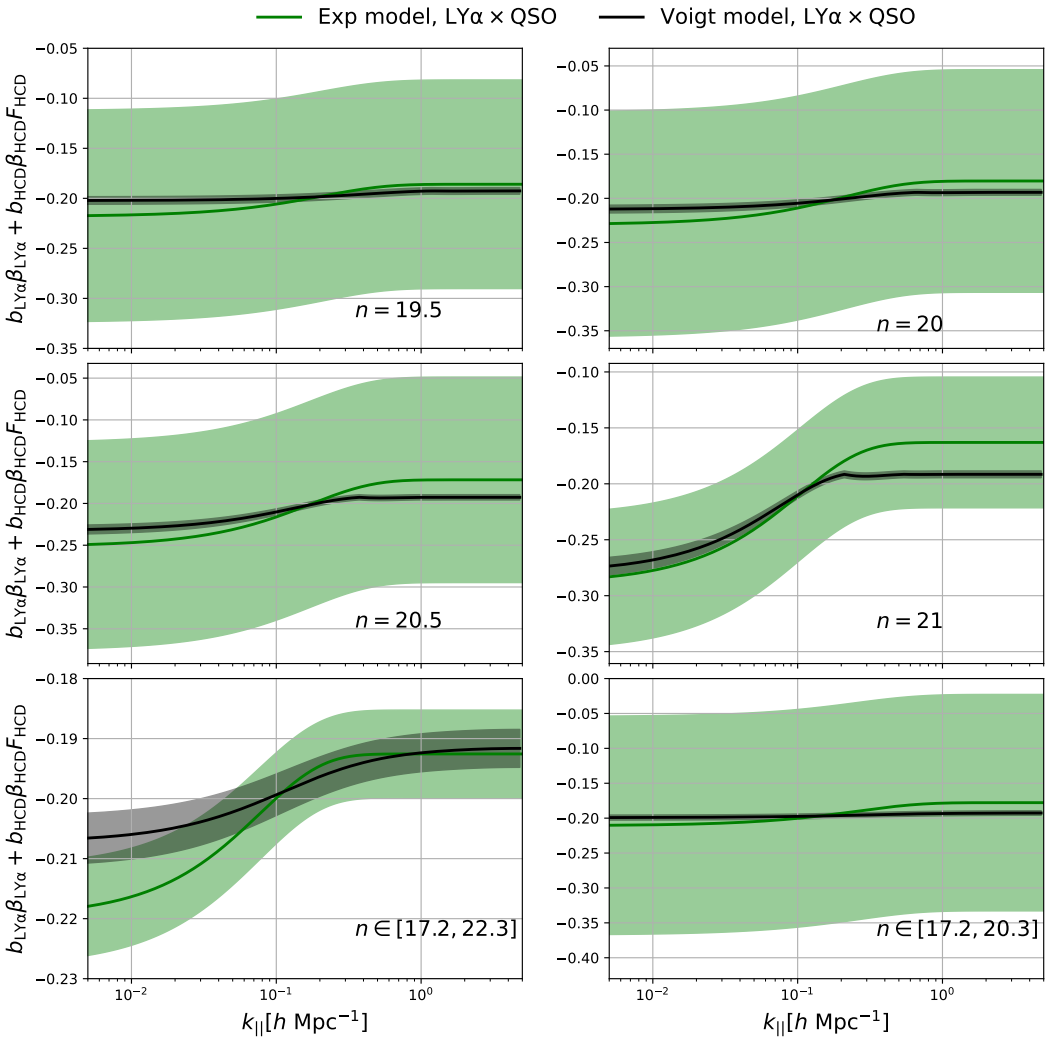

FIGURE 6.16 : The comparison between the `voigt` model and the `Exp` model for `Saclay` mocks with different types of HCDs. The upper four plots show the comparison in mocks with HCDs with the same column densities, from 19.5 to 21. The lower two plots show the mocks with a distribution of HCDs. The blue curves and red curves show the 1σ constraints on $b'\beta' = b_{\mathrm{Ly}\alpha}\beta_{\mathrm{Ly}\alpha} + b_{\mathrm{HCD}}\beta_{\mathrm{HCD}}F_{\mathrm{HCD}}(k_{||})$, of the `Exp` model and the `Voigt` model, respectively, fitted using the Lyα-quasar cross-correlation function.

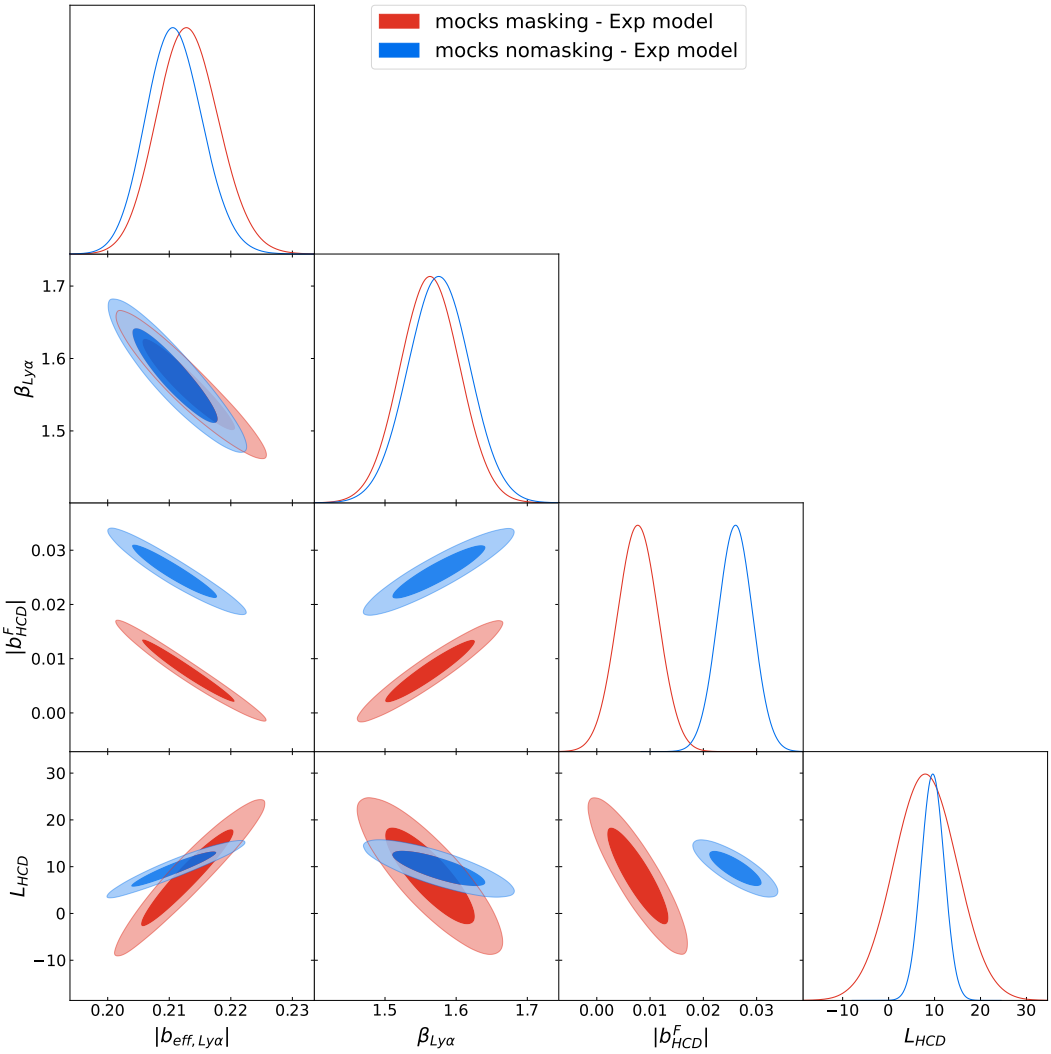

FIGURE 6.17 : Triangle plot for the Lyα parameters constraints $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha},|b^F_{\mathrm{HCD}}|,L_{\mathrm{HCD}}\}$ using the `Exp` model.

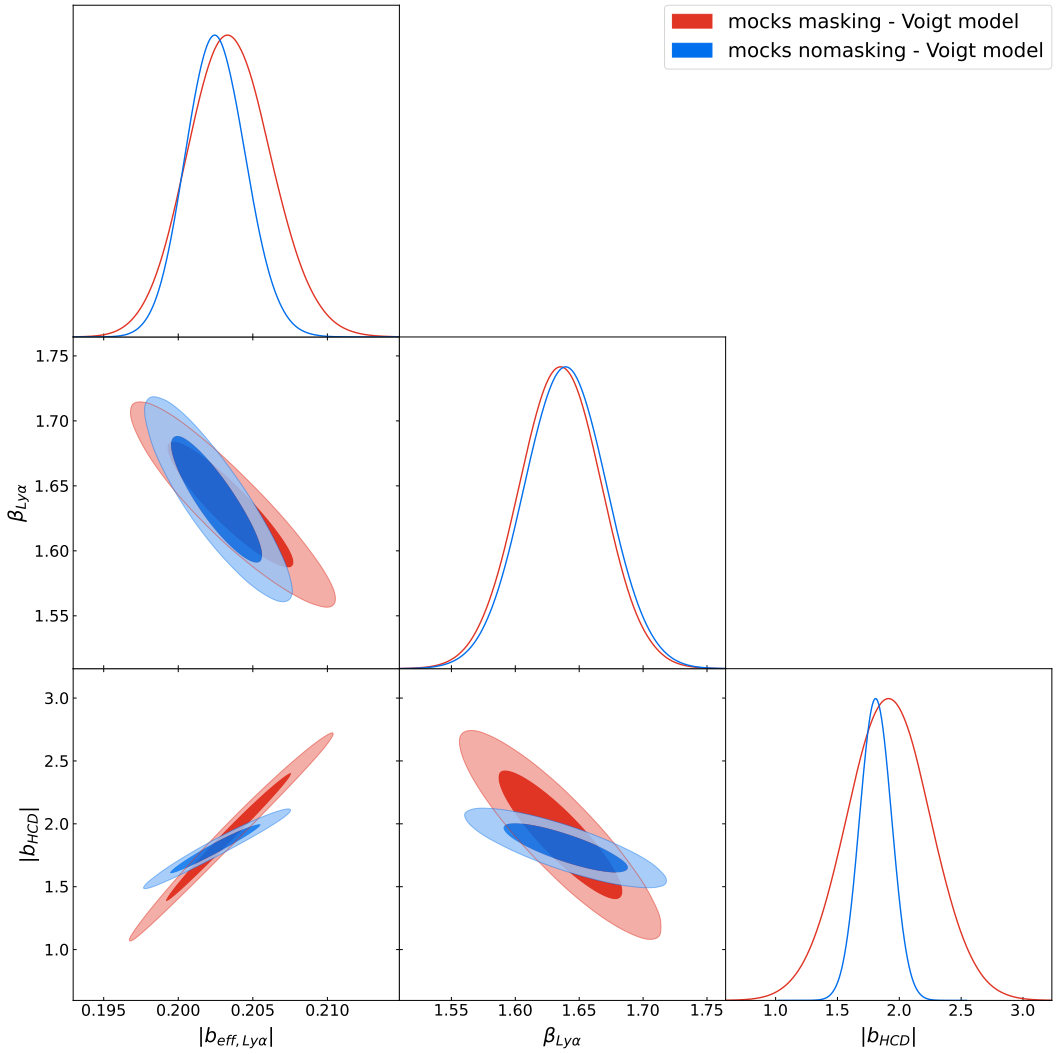

FIGURE 6.18 : Triangle plot for the Lyα parameters constraints $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha},|b_{\mathrm{HCD}}|\}$ using the `Voigt` model.

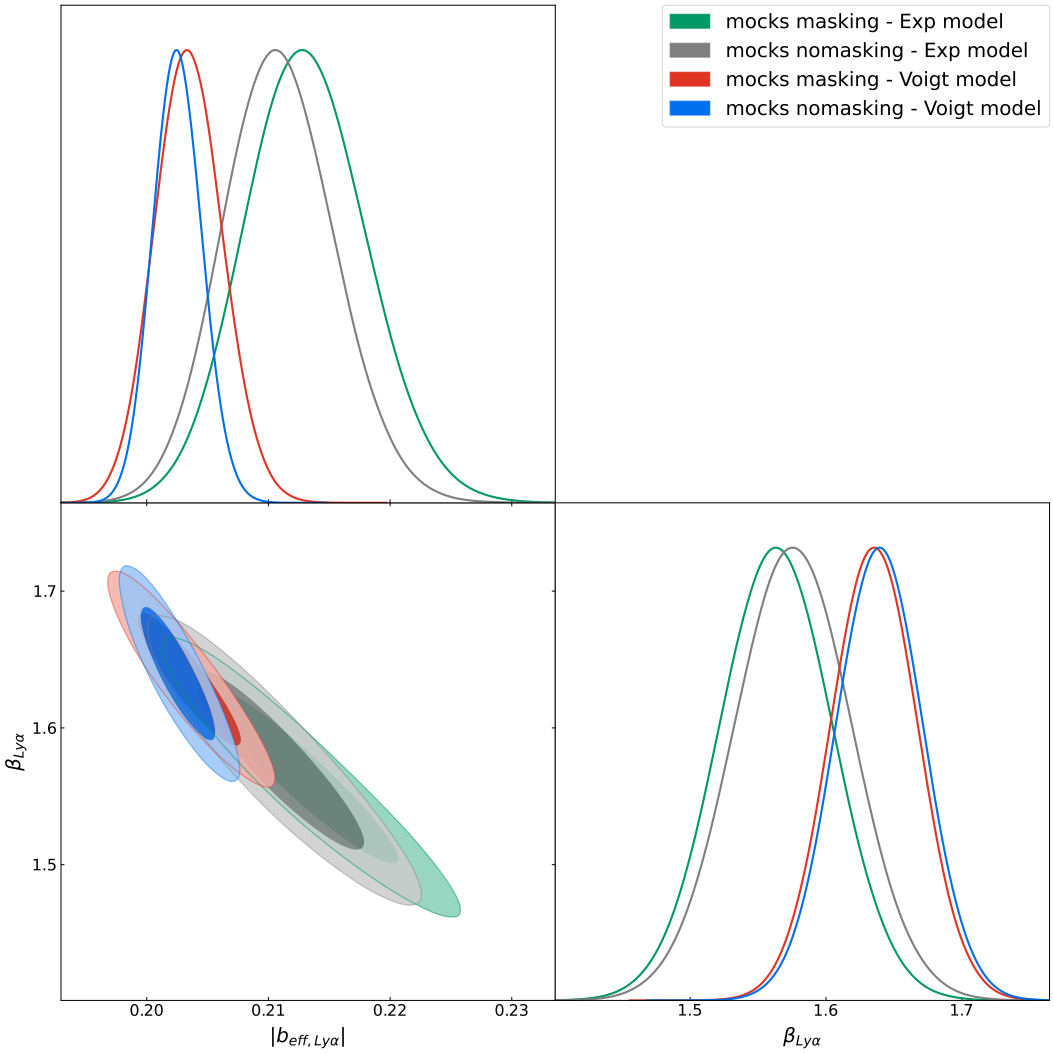

FIGURE 6.19 : The comparison of the Lyα parameters constraints $\{|b_{\mathrm{eff,Ly}\alpha}|,\beta_{\mathrm{Ly}\alpha}\}$ between the `Voigt` model and the `Exp` model.

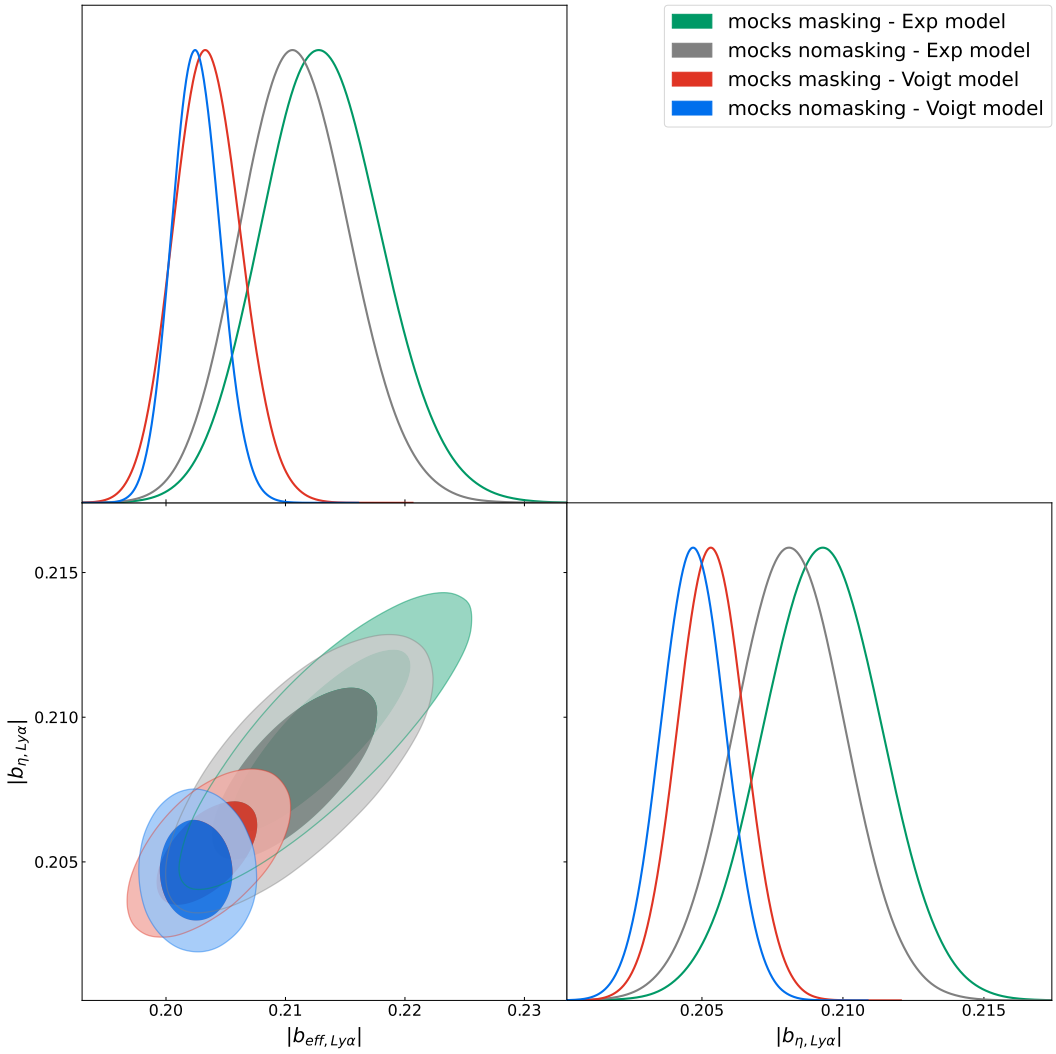

FIGURE 6.20 : The comparison of the Lyα parameters constraints $\{|b_{\mathrm{eff,LY}\alpha}|,|b_{\eta,\mathrm{LY}\alpha}|\}$ between the `Voigt` model and the `Exp` model.

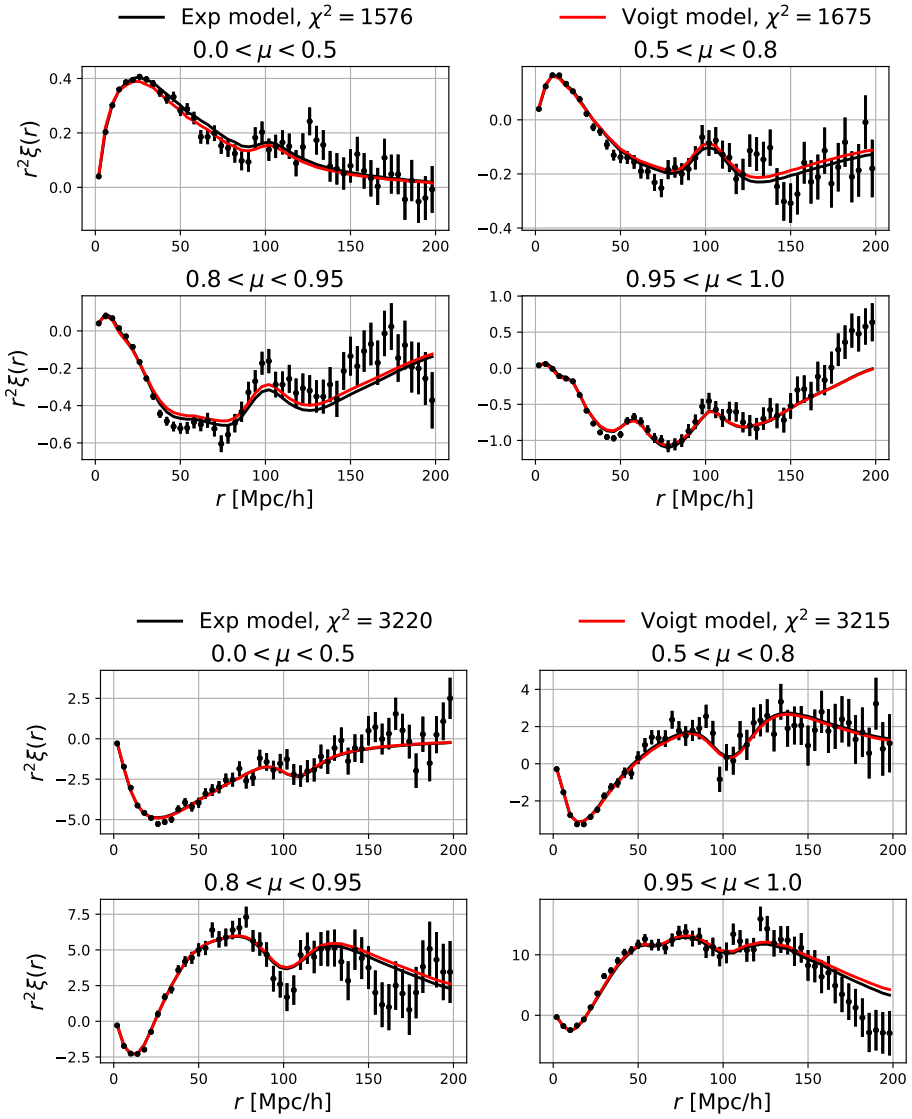

FIGURE 6.21 : eBOSS DR16 data : Lyα auto-correlation function (top four panels) and Lyα-quasar cross-correlation (bottom four panels), for pixels in the Lyα region. The correlations are multiplied by r^2 to better see the BAO scale. The black curves show the best-fit models using the Exp model, with and without a fixed $L_{\rm HCD} = 10h^{-1}$Mpc. The red curves give the best-fit models using the Voigt model, in four wedges of $|\mu| = |\frac{r_\parallel}{r}|$. The fitted range is chosen as $r \in [10, 180]h^{-1}$Mpc.

	Data with DLAs masked				Data without DLAs masked			
	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$QSO	Ly$\alpha \times$QSO	Ly$\alpha \times$Lyα	Ly$\alpha \times$Lyα	Ly$\alpha \times$QSO	Ly$\alpha \times$QSO
Model	Exp model	Voigt model	Exp model	Voigt model	Exp model	Voigt model	Exp model	Voigt model
χ^2	1576.22	1624.34	3220.27	3221.4	1594.96	1630.65	3219.44	3224.94
$N_{\rm data}$	1590	1590	3180	3180	1590	1590	3180	3180
$N_{\rm par}$	14	13	13	12	14	13	13	12
P	0.49	0.2	0.25	0.25	0.36	0.17	0.25	0.24
$\alpha_{\|\|}$	1.05±0.034	1.04±0.033	1.06±0.032	1.06±0.032	1.04±0.034	1.04±0.033	1.05±0.034	1.05±0.034
$\alpha_\perp$	0.981±0.042	0.985±0.041	0.932±0.039	0.933±0.039	0.974±0.044	0.973±0.044	0.948±0.042	0.947±0.042
$b_{\eta,{\rm Ly}\alpha}$	-0.175±0.013	-0.179±0.004	-0.228±0.016	-0.237±0.014	-0.173±0.013	-0.189±0.005	-0.231±0.019	-0.27±0.018
$\beta_{{\rm Ly}\alpha}$	3.23±1.26	1.71±0.11	1.91±0.33	1.91±0.21	5.25±3.29	1.84±0.14	1.92±0.34	1.56±0.16
$b^F_{\rm HCD}$	-0.105±0.022		-0.034±0.024		-0.139±0.02		-0.047±0.027	
$b_{\rm HCD}$		7.3±0.611		3.78±1.92		4.79±0.326		-0.424±1.4
$\beta_{\rm HCD}$	0.53±0.08	0.67±0.08	0.52±0.09	0.51±0.09	0.51±0.08	0.67±0.08	0.51±0.09	0.5±0.09
$L_{\rm HCD}$	2.28±0.63		0.95±2.87		2.59±0.52		-0.01±1.66	

TABLEAU 6.6 : Best fit parameters for eBOSS DR16 data, using the `Voigt` model and the `Exp` model, for Lyα auto-correlation function and Lyα-quasar cross-correlation, respectively.

6.5 Non-linear effects of HCDs

I present in Figure 6.22 a comparison of all the non-linear effects on the Lyα power spectrum, i.e., the `Voigt` model showing the non-linear effect due to the Voigt-profile HCD absorption, the non-linear effect of Lyα forests at small scales (hereafter the Arinyo effect (ARINYO-I-PRATS, MIRALDA-ESCUDÉ, VIEL et CEN 2015)), the binning effect due to the correlation function bins, and the quasar nonlinear velocities that affect the Lyα-quasar cross-correlation. Note that the non-linear effect of HCDs at small scales ($D_{\rm NL,HCD}$ in Equation 6.20) produce a suppression at higher $k_{||}$ than that of the Lyα forests, and thus are ignored in this comparison. It turns out that for the Lyα auto-correlation function, the `Voigt` model predicts a suppression at smaller $k_{||}$ than the Arinyo effect and the binning effect. However, the binning effect drops down quickly, thus covering the tail of the HCD effect. Therefore, in order to get rid of the impact of the binning effect, we should make a further study on the Lyα correlation function with smaller binsize, e.g., = 2 or $1h^{-1}$Mpc. For Lyα-quasar cross-correlations, the quasar nonlinear velocities give a non-negligible suppression, comparable to the HCD effect. This could be one of the explanations of our worse constraints of $b_{\rm HCD}$ from the cross correlations.

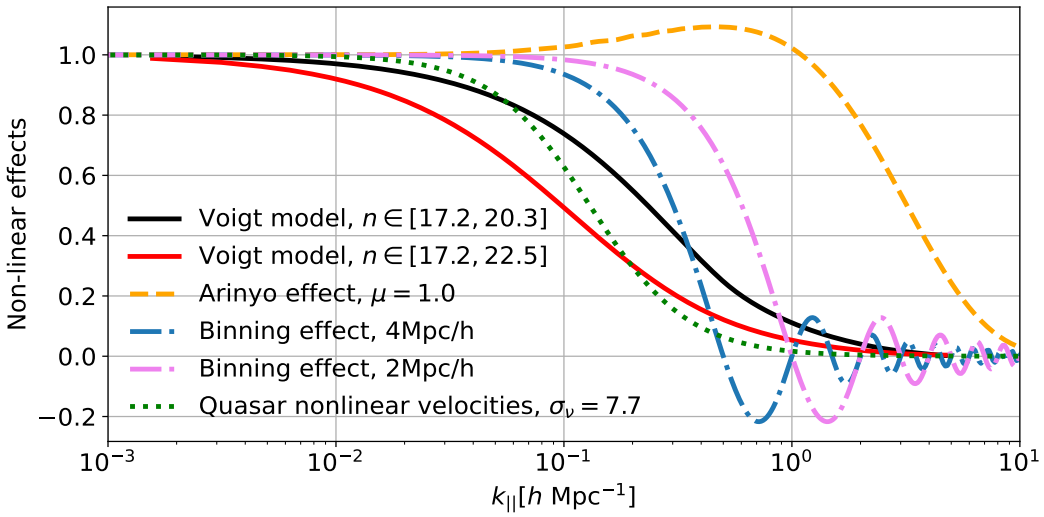

FIGURE 6.22 : The comparison of the `Voigt` model (Solid lines), the Arinyo effect (Dashed line), the binning effect (Dotted-dashed lines), and the quasar nonlinear velocities (Dotted line).

6.6 Summary and Prospects

The damping wings of the HCD absorption profile result in a suppression on the Lyα forest power spectrum, and a broadband impact on the correlation function. I have developed a three-parameter empirical fitting function, the $L\beta\gamma$ model, to characterize this non-local damping effect of HCDs on the Lyα correlation function and power spectrum. In this model, the L_{HCD} parameter determines the scale of the suppression, which is physically related to the size of HCDs. The combination of β and γ determines the shape of the slope of the damping function $F_{\mathrm{HCD}}(k_{||})$. The $L\beta\gamma$ model shows no difference with the `Exp` model when applied to eBOSS Saclay mocks with HCDs, while showing encouraging improvement when applied to eBOSS DR16 data in the range of $20h^{-1}\mathrm{Mpc} < r < 80h^{-1}\mathrm{Mpc}$. This suggests that the $L\beta\gamma$ model is probably modeling something else than HCDs, which has a non-negligible impact on the Lyα correlations.

I further developed a theoretical model, i.e., the `Voigt` model, based on the Voigt absorption profile that parametrizes the damping wings of HCDs, and takes into account the HI column density probability distribution of HCDs. It has no additional free parameters, providing a physical measurement of the bias and RSD parameters of HCDs, and a good constraint on the Lyα parameters. My theoretical development on this model also gives a physical ground to understand the phenomenological models used in previous Lyα analyses. Based on the `Voigt` model, we understand clearly the physical meaning of the `Exp` model, where all the absorption profiles are assumed to be Lorentzian with the same width, and their localization information is not taken into account. We perform a suite of verification of the model, based on the fitting of Lyα forest correlations, computed from mocks with different HI column density probability distributions of HCDs. It turns out that the model works well for HCDs with small n, while giving small discrepancies for mocks with large HCDs with $n \geqslant 20.5$. These could possibly be affected by the continuum fitting distortions due to the large DLA absorption in the Lyα analysis pipeline (could be checked with the true transmission mocks in the future). Compared with the previous `Exp` model used in the eBOSS DR16 analysis, the `Voigt` model gives comparable χ^2, with tighter constraints on $b_{\mathrm{eff,Ly}\alpha}$, $b_{\eta,\mathrm{LY}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$. The correlation between $b_{\mathrm{eff,Ly}\alpha}$ and $b_{\eta,\mathrm{LY}\alpha}$ is much wea-

ker compared to the `Exp` model. The auto- and cross-correlation functions results for $b_{\eta,\mathrm{Ly}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$ are also more consistent with the `Voigt` model. Moreover, the fitting from the eBOSS DR16 data confirms that the modeling of HCDs does not affect the measurement of BAO significantly. However, we measure a much larger b_{HCD} in eBOSS DR16 data than what we have in the mocks, which could possibly be explained that the input HI column density probability distribution of HCDs (obtained using limited DLAs in the literature) is not accurate enough to constrain b_{HCD} of eBOSS or DESI data. Moreover, a future study can be explored, taking into account a more realistic HI column density distribution, regarding the dependence of the hosting halo mass. The HI column density distribution in the range of $17 < n < 20$, could also potentially be constrained with this model, which is technically hard to measure from direct observation.

Bibliography of the current chapter

Sundius, T. (1973). "Computer fitting of Voigt profiles to Raman lines". In: *Journal of Raman Spectroscopy* 1.5, pp. 471–488.

Kaiser, N. (1987). "Clustering in real space and in redshift space". In: *Monthly Notices of the Royal Astronomical Society* 227.1, pp. 1–21.

Prochaska, J. X. and S. Herbert-Fort (2004). "The sloan digital sky survey damped Lyα survey: data release 1". In: *Publications of the Astronomical Society of the Pacific* 116.821, p. 622.

Viel, M., M. Haehnelt, R. Carswell, and T.-S. Kim (2004). "The effect of (strong) discrete absorption systems on the Lyman α forest flux power spectrum". In: *Monthly Notices of the Royal Astronomical Society* 349.3, pp. L33–L37.

McDonald, P., U. Seljak, R. Cen, P. Bode, and J. P. Ostriker (2005). "Physical effects on the Lyα forest flux power spectrum: damping wings, ionizing radiation fluctuations and galactic winds". In: *Monthly Notices of the Royal Astronomical Society* 360.4, pp. 1471–1482.

Prochaska, J. X., S. Herbert-Fort, and A. M. Wolfe (2005). "The SDSS damped Lyα survey: data release 3". In: *The Astrophysical Journal* 635.1, p. 123.

Garcıa, T. T. (2006). "Voigt profile fitting to quasar absorption lines: an analytic approximation to the Voigt–Hjerting function". In: *Monthly Notices of the Royal Astronomical Society* 369.4, pp. 2025–2035.

Noterdaeme, P., P. Petitjean, C. Ledoux, and R. Srianand (2009). "Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II–Data Release 7". In: *Astronomy & Astrophysics* 505.3, pp. 1087–1098.

Percival, W. J. and M. White (2009). "Testing cosmological structure formation using redshift-space distortions". In: *Monthly Notices of the Royal Astronomical Society* 393.1, pp. 297–308.

Font-Ribera, A. and J. Miralda-Escudé (2012). "The effect of high column density systems on the measurement of the Lyman-α forest correlation function". In: *Journal of Cosmology and Astroparticle Physics* 2012.07, p. 028.

Noterdaeme, P., P. Petitjean, W. Carithers, et al. (2012). "Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9". In: *Astronomy & Astrophysics* 547, p. L1.

Arinyo-i-Prats, A., J. Miralda-Escudé, M. Viel, and R. Cen (2015). "The non-linear power spectrum of the Lyman alpha forest". In: *Journal of Cosmology and Astroparticle Physics* 2015.12, p. 017.

Albawi, S., T. A. Mohammed, and S. Al-Zawi (2017). "Understanding of a convolutional neural network". In: *2017 international conference on engineering and technology (ICET)*. Ieee, pp. 1–6.

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z= 2.3 with SDSS DR12 Lyα-Forests". In: *Astronomy & Astrophysics* 603, A12.

Garnett, R., S. Ho, S. Bird, and J. Schneider (2017). "Detecting damped Ly α absorbers with Gaussian processes". In: *Monthly Notices of the Royal Astronomical Society* 472.2, pp. 1850–1865.

Parks, D., J. X. Prochaska, S. Dong, and Z. Cai (2018). "Deep learning of quasar spectra to discover and characterize damped Lyα systems". In: *Monthly Notices of the Royal Astronomical Society* 476.1, pp. 1151–1168.

Pérez-Ràfols, I., J. Miralda-Escudé, A. Arinyo-i-Prats, A. Font-Ribera, and L. Mas-Ribas (2018). "The cosmological bias factor of damped Lyman alpha systems: dependence on metal line strength". In: *Monthly Notices of the Royal Astronomical Society* 480.4, pp. 4702–4709.

Rogers, K. K., S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera, and B. Leistedt (2018a). "Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers". In: *Monthly Notices of the Royal Astronomical Society* 476.3, pp. 3716–3728.

— (2018b). "Simulating the effect of high column density absorbers on the one-dimensional Lyman α forest flux power spectrum". In: *Monthly Notices of the Royal Astronomical Society* 474.3, pp. 3032–3042.

de Sainte Agathe, V. et al. (Sept. 2019a). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In: 629, A85, A85. arXiv: 1904.03400 [astro-ph.CO].

Cuceu, A., A. Font-Ribera, and B. Joachimi (2020). "Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman-α forest". In: *Journal of Cosmology and Astroparticle Physics* 2020.07, p. 035.

Des Bourboux, H. D. M., J. Rich, et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey: baryon acoustic oscillations with Lyα forests". In: *The Astrophysical Journal* 901.2, p. 153.

Fumagalli, M., S. Fotopoulou, and L. Thomson (2020). "Detecting neutral hydrogen at z 3 in large spectroscopic surveys of quasars". In: *Monthly Notices of the Royal Astronomical Society* 498.2, pp. 1951–1962.

Ho, M.-F., S. Bird, and R. Garnett (2020). "Detecting multiple DLAs per spectrum in SDSS DR12 with Gaussian processes". In: *Monthly Notices of the Royal Astronomical Society* 496.4, pp. 5436–5454.

— (2021). "Damped Lyman-α absorbers from Sloan digital sky survey DR16Q with Gaussian processes". In: *Monthly Notices of the Royal Astronomical Society* 507.1, pp. 704–719.

Ting Tan, Y. L. and C. Balland (2021). "Detection of the Damped Lyman-alpha systems in quasar spectra with machine learning algorithms". In: *2021 IAP colloquium.*

Chabanier, S., T. Etourneau, et al. (2022). "The Completed Sloan Digital Sky Survey IV Extended Baryon Oscillation Spectroscopic Survey: The Damped Lyα Systems Catalog". In: *The Astrophysical Journal Supplement Series* 258.1, p. 18.

Wang, B. et al. (2022). "Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems". In: *The Astrophysical Journal Supplement Series* 259.1, p. 28.

Jiaqi, Z. et al. (in preparation). *The DESI Damped Lyα System Survey: Data Release 1.*

Conclusion (English version)

In this manuscript, I presented the work carried out during my Ph.D. study in the cosmology group of the LPNHE (Sorbonne University), which is supported by CNRS and Centre Pierre Binetruy (CPB). I make use of the Lyα mocks developed at CEA Saclay, and collaborated with their cosmology group in the Lyα analysis.

This thesis benefits from spectroscopic observation data from two large cosmological surveys, eBOSS and DESI. The survey validation program of DESI was started almost at the same time as this thesis, thus enabling me to get involved in the data quality checking, target selection pipeline test, collection of the main survey data, and scientific analysis of DESI.

I presented a comparison of DESI EDR and eBOSS DR16 Lyα analysis. Comparable χ^2 and similar parameter correlations were found between their fits. The DESI EDR data show encouraging data quality and the need for more systematic studies to prepare the upcoming enormous DESI dataset.

I work as a core member for the test of Saclay mocks and am responsible for the insertion of HCDs and BALs into DESI mocks. The analysis on Lyα mocks shows that the Lyα analysis pipeline performs well and motivates further development of the model for HCDs and metals.

The presence of HCDs in quasar spectra has a broadband impact on the Lyα correlation functions. The damping wings of HCDs extend out to all scales in the Lyα forest, resulting in a suppression (at the scale of HCD widths) on the Lyα power spectrum. It is therefore essential to have a physical understanding of the non-local effect of HCDs on the Lyα correlation functions, which will be useful for the Lyα full shape analysis, P1D measurements, etc.

I have developed a three-parameter empirical fitting function, the $L\beta\gamma$ model, to characterize this non-local damping effect of HCDs on the Lyα correlation function and power spectrum. In this model, the $L_{\rm HCD}$ parameter determines the scale of the suppression, which is physically related to the size of HCDs. The combination of β and γ determines the shape of the slope of the damping function $F_{\rm HCD}(k_{||})$. The $L\beta\gamma$ model shows encouraging improvement when applied to eBOSS DR16 data in the range of $20h^{-1}{\rm Mpc} < r < 80h^{-1}{\rm Mpc}$. However, this is probably modeling something else than HCDs, which has a non-negligible impact on the Lyα correlations.

I further developed a theoretical model, i.e., the `Voigt` model, based on the Voigt absorption profile that parametrizes the damping wings of HCDs, and takes into account the HI column density probability distribution of HCDs. It provides a physical measurement of the bias and RSD parameters of HCDs, and a good constraint on the Lyα parameters. My theoretical development on this model also gives a physical ground to understand the phenomenological models used in previous Lyα analyses. Based on the `Voigt` model, we understand clearly the physical meaning of the `Exp` model, where all the absorption profiles are assumed to be Lorentzian with the same width, and their localization information is not taken into account. I perform a suite of verification of the model, based on the fitting of Lyα forest correlations, computed from mocks with different HI column density probability distributions of HCDs.

Compared with the previous `Exp` model used in the eBOSS DR16 analysis, the `Voigt` model

gives comparable χ^2, with tighter constraints on $b_{\mathrm{eff,Ly}\alpha}$, $b_{\eta,\mathrm{LY}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$. The correlation between $b_{\mathrm{eff,Ly}\alpha}$ and $b_{\eta,\mathrm{LY}\alpha}$ is much weaker compared to the `Exp` model. The auto- and cross-correlation functions results for $b_{\eta,\mathrm{Ly}\alpha}$ and $\beta_{\mathrm{Ly}\alpha}$ are also more consistent with the `Voigt` model. Moreover, the fitting from the eBOSS DR16 data confirms that the modeling of HCDs does not affect the measurement of BAO significantly.

My work provides an important theoretical tool for future Lyα analyses and could potentially be useful for the study of dark energy models.

Conclusion (French version)

Dans ce manuscrit, j'ai présenté les travaux menés lors de ma thèse dans le groupe de cosmologie du LPNHE (Sorbonne Université), soutenu par le CNRS et le Centre Pierre Binetruy (CPB). J'utilise les simulations Lyα développées au CEA Saclay et j'ai collaboré avec leur groupe de cosmologie à l'analyse Lyα.

Cette thèse bénéficie des données d'observation spectroscopiques de deux grandes campagnes cosmologiques, eBOSS et DESI. Le programme de validation d'enquête de DESI a démarré presque en même temps que cette thèse, me permettant ainsi de m'impliquer dans la vérification de la qualité des données, le test du pipeline de sélection des cibles, la collecte des principales données d'enquête et l'analyse scientifique de DESI.

J'ai présenté une comparaison des analyses DESI EDR et eBOSS DR16 Lyα. Des χ^2 comparables et des corrélations de paramètres similaires ont été trouvées entre leurs ajustements. Les données DESI EDR montrent une qualité de données encourageante et la nécessité d'études plus systématiques pour préparer l'énorme ensemble de données DESI à venir.

Je travaille en tant que membre principal pour le test des simulations Saclay et suis responsable de l'insertion des HCD et BAL dans les simulations DESI. L'analyse des simulations Lyα montre que le pipeline d'analyse Lyα fonctionne bien et motive le développement ultérieur du modèle pour les HCD et les métaux.

La présence de HCD dans les spectres des quasars a un impact à large bande sur les fonctions de corrélation Lyα. Les ailes amortissantes des HCD s'étendent à toutes les échelles de la forêt Lyα, ce qui entraîne une coupure (à l'échelle des largeurs des HCD) sur le spectre de puissance Lyα. Il est donc essentiel d'avoir une compréhension physique de l'effet non local des HCD sur les fonctions de corrélation Lyα, ce qui sera utile pour l'analyse de forme complète Lyα, les mesures P1D, etc.

J'ai développé une fonction d'ajustement empirique à trois paramètres, le modèle $L\beta\gamma$, pour caractériser cet effet d'amortissement non local des HCD sur la fonction de corrélation Lyα et le spectre de puissance. Dans ce modèle, le paramètre $L_{\rm HCD}$ détermine l'échelle du seuil, qui est physiquement liée à la taille des HCD. La combinaison de β et γ détermine la forme de la pente de la fonction d'amortissement $F_{\rm HCD}(k_{||})$. Le modèle $L\beta\gamma$ montre une amélioration encourageante lorsqu'il est appliqué aux données eBOSS DR16 dans la plage de $20h^{-1}$Mpc $< r < 80h^{-1}$Mpc. Cependant, il s'agit probablement d'une modélisation autre que les HCD, ce qui a un impact non négligeable sur les corrélations Lyα.

J'ai ensuite développé un modèle théorique, à savoir le modèle `Voigt`, basé sur le profil d'absorption Voigt qui paramétrise les ailes d'amortissement des HCD et prend en compte la distribution de probabilité de densité de colonne HI des HCD. Il fournit une mesure physique des paramètres de biais et de RSD des HCD, ainsi qu'une bonne contrainte sur les paramètres Lyα. Mon développement théorique sur ce modèle donne également une base physique pour comprendre les modèles phénoménologiques utilisés dans les analyses Lyα précédentes. Sur la base du modèle `Voigt`, nous comprenons clairement la signification physique du modèle `Exp`, où

tous les profils d'absorption sont supposés lorentziens de même largeur, et leurs informations de localisation ne sont pas prises en compte. compte. J'effectue une suite de vérifications du modèle, basée sur l'ajustement des corrélations forestières Lyα, calculées à partir de simulations avec différentes distributions de probabilité de densité de colonnes HI des HCD.

Par rapport au modèle `Exp` précédent utilisé dans l'analyse eBOSS DR16, le modèle `Voigt` donne des χ^2 comparables, avec des contraintes plus strictes sur $b_{\mathrm{eff,Ly}\alpha}$, $b_{\eta,\mathrm{LY}\alpha}$ et $\beta_{\mathrm{Ly}\alpha}$. La corrélation entre $b_{\mathrm{eff,Ly}\alpha}$ et $b_{\eta,\mathrm{LY}\alpha}$ est beaucoup plus faible par rapport au `Exp` modèle. Les résultats des fonctions d'auto-corrélation et de corrélation croisée pour $b_{\eta,\mathrm{Ly}\alpha}$ et $\beta_{\mathrm{Ly}\alpha}$ sont également plus cohérents avec le `Voigt` modèle. De plus, l'ajustement des données eBOSS DR16 confirme que la modélisation des HCD n'affecte pas de manière significative la mesure du BAO.

Mon travail fournit un outil théorique important pour les futures analyses Lyα et pourrait potentiellement être utile pour l'étude des modèles d'énergie noire.

Bibliographie

Jeans, J. H. (1902). "I. The stability of a spherical nebula". In : *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 199.312-320, p. 1-53.

Einstein, A. (1916). "The foundation of the general theory of relativity." In : *Annalen Phys.* 49.7. Sous la dir. de J.-P. Hsu et D. Fine, p. 769-822.

Hubble, E. (1929). "A relation between distance and radial velocity among extra-galactic nebulae". In : *Proceedings of the national academy of sciences* 15.3, p. 168-173.

Robertson, H. P. (1936). "Kinematics and World-Structure III." In : *Astrophysical Journal, vol. 83, p. 257* 83, p. 257.

Walker, A. G. (1937). "On Milne's theory of world-structure". In : *Proceedings of the London Mathematical Society* 2.1, p. 90-127.

Bondi, H. et T. Gold (1948). "The steady-state theory of the expanding universe". In : *Monthly Notices of the Royal Astronomical Society* 108.3, p. 252-270.

Matthews, T. A. et A. R. Sandage (1963). "Optical Identification of 3C 48, 3C 196, and 3C 286 with Stellar Objects." In : *Astrophysical Journal, vol. 138, p. 30* 138, p. 30.

Gunn, J. E. et B. A. Peterson (nov. 1965). "On the Density of Neutral Hydrogen in Intergalactic Space." In : 142, p. 1633-1636.

Penzias, A. A. et R. W. Wilson (1965). "Measurement of the Flux Density of CAS a at 4080 Mc/s." In : *The Astrophysical Journal* 142, p. 1149.

Bahcall, J. N. et P. Peebles (1969). "Statistical tests for the origin of absorption lines observed in quasi-stellar sources". In : *The Astrophysical Journal* 156, p. L7.

Moffat, A. (1969). "A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry". In : *Astronomy and Astrophysics, Vol. 3, p. 455 (1969)* 3, p. 455.

Jackson, J. (1972). "A critique of Rees's theory of primordial gravitational radiation". In : *Monthly Notices of the Royal Astronomical Society* 156.1, 1P-5P.

Sundius, T. (1973). "Computer fitting of Voigt profiles to Raman lines". In : *Journal of Raman Spectroscopy* 1.5, p. 471-488.

Donnelly, T., S. Freedman, R. Lytel, R. Peccei et M. Schwartz (1978). "Do axions exist ?" In : *Physical Review D* 18.5, p. 1607.

Alcock, C. et B. Paczyński (1979). "An evolution free test for non-zero cosmological constant". In : *Nature* 281.5730, p. 358-359.

Greenstein, J. L. et M. Schmidt (1979). "The quasi-stellar radio sources 3c 48 and 3c 273". In : *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, p. 811-818.

Penzias, A. A. et R. W. Wilson (1979). "A measurement of excess antenna temperature at 4080 MHz". In : *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, p. 873-876.

Carswell, R. F., D. C. Morton, M. G. Smith, A. N. Stockton, D. A. Turnshek et R. J. Weymann (1984). "The absorption line profiles in Q1101-264". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 278, March 15, 1984, p. 486-498. Research supported by the Science and Engineering Research Council and Radcliffe Trust.* 278, p. 486-498.

Bethe, H., G. Brown et I. Worpole (1985). "How a Supernova Explodes". In.

Kaiser, N. (1987). "Clustering in real space and in redshift space". In : *Monthly Notices of the Royal Astronomical Society* 227.1, p. 1-21.

Foltz, C., F. Chaffee, P. Hewett, R. Weymann et S. Morris (1990). "On the Fraction of Optically-Selected QSOs with Broad Absorption Lines in Their Spectra". In : *Bulletin of the American Astronomical Society.* T. 22, p. 806.

Coles, P. et B. Jones (1991). "A lognormal model for the cosmological mass distribution". In : *Monthly Notices of the Royal Astronomical Society* 248.1, p. 1-13.

Weymann, R. J., S. L. Morris, C. B. Foltz et P. C. Hewett (1991). "Comparisons of the emission-line and continuum properties of broad absorption line and normal quasi-stellar objects". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 373, May 20, 1991, p. 23-53.* 373, p. 23-53.

Hoyle, F., G. Burbidge et J. V. Narlikar (1993). "A quasi-steady state cosmological model with creation of matter". In : *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 410, no. 2, p. 437-457.* 410, p. 437-457.

Jungman, G., M. Kamionkowski et K. Griest (1996). "Supersymmetric dark matter". In : *Physics Reports* 267.5-6, p. 195-373.

Hui, L. et N. Y. Gnedin (1997). "Equation of state of the photoionized intergalactic medium". In : *Monthly Notices of the Royal Astronomical Society* 292.1, p. 27-42.

Gunn, J. E., M. Carr et al. (1998). "The Sloan digital sky survey photometric camera". In : *The Astronomical Journal* 116.6, p. 3040.

Hamilton, A. (1998). "Linear redshift distortions : A Review". In : *The Evolving Universe : Selected Topics on Large-Scale Structure and on the Properties of Galaxies*, p. 185-275.

Perlmutter, S. et al. (1999). "Measurements of Ω and Λ from 42 high-redshift supernovae". In : *The Astrophysical Journal* 517.2, p. 565.

Lewis, A., A. Challinor et A. Lasenby (2000). "Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models". In : *The Astrophysical Journal* 538.2, p. 473.

Riess, A. G. et al. (2000). "Tests of the accelerating universe with near-infrared observations of a high-redshift type Ia supernova". In : *The Astrophysical Journal* 536.1, p. 62.

York, D. G. et al. (2000). "The sloan digital sky survey : Technical summary". In : *The Astronomical Journal* 120.3, p. 1579.

Hamilton, A. (2001). "Formulae for growth factors in expanding universes containing matter and a cosmological constant". In : *Monthly Notices of the Royal Astronomical Society* 322.2, p. 419-425.

Jing, Y. et G. Börner (2001). "Scaling properties of the redshift power spectrum : theoretical models". In : *The Astrophysical Journal* 547.2, p. 545.

Hall, P. B. et al. (2002). "Unusual broad absorption line quasars from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 141.2, p. 267.

Richards, G. T. et al. (2002). "Spectroscopic target selection in the sloan digital sky survey : The quasar sample". In : *The Astronomical Journal* 123.6, p. 2945.

McDonald, P. (2003). "Toward a Measurement of the Cosmological Geometry at z˜ 2 : Predicting Lyα Forest Correlation in Three Dimensions and the Potential of Future Data Sets". In : *The Astrophysical Journal* 585.1, p. 34.

Prochaska, J. X. et S. Herbert-Fort (2004). "The sloan digital sky survey damped Lyα survey : data release 1". In : *Publications of the Astronomical Society of the Pacific* 116.821, p. 622.

Viel, M., M. Haehnelt, R. Carswell et T.-S. Kim (2004). "The effect of (strong) discrete absorption systems on the Lyman α forest flux power spectrum". In : *Monthly Notices of the Royal Astronomical Society* 349.3, p. L33-L37.

Yoon, S.-C. et N. Langer (2004). "Presupernova evolution of accreting white dwarfs with rotation". In : *Astronomy & Astrophysics* 419.2, p. 623-644.

Eisenstein, D. J., I. Zehavi et al. (2005). "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies". In : *The Astrophysical Journal* 633.2, p. 560.

McDonald, P., U. Seljak, R. Cen, P. Bode et J. P. Ostriker (2005). "Physical effects on the Lyα forest flux power spectrum : damping wings, ionizing radiation fluctuations and galactic winds". In : *Monthly Notices of the Royal Astronomical Society* 360.4, p. 1471-1482.

Prochaska, J. X., S. Herbert-Fort et A. M. Wolfe (2005). "The SDSS damped Lyα survey : data release 3". In : *The Astrophysical Journal* 635.1, p. 123.

Garcia, T. T. (2006). "Voigt profile fitting to quasar absorption lines : an analytic approximation to the Voigt–Hjerting function". In : *Monthly Notices of the Royal Astronomical Society* 369.4, p. 2025-2035.

McDonald, P., U. Seljak, S. Burles et al. (2006). "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey". In : *The Astrophysical Journal Supplement Series* 163.1, p. 80.

Trump, J. R. et al. (2006). "A catalog of broad absorption line quasars from the sloan digital sky survey third data release". In : *The Astrophysical Journal Supplement Series* 165.1, p. 1.

Eisenstein, D. J., H.-j. Seo, E. Sirko et D. N. Spergel (2007). "Improving cosmological distance measurements by reconstruction of the baryon acoustic peak". In : *The Astrophysical Journal* 664.2, p. 675.

Eisenstein, D. J., H.-J. Seo et M. White (2007). "On the robustness of the acoustic scale in the low-redshift clustering of matter". In : *The Astrophysical Journal* 664.2, p. 660.

Mazzali, P. A., F. K. Ropke, S. Benetti et W. Hillebrandt (2007). "A common explosion mechanism for type Ia supernovae". In : *Science* 315.5813, p. 825-828.

Percival, W. J., S. Cole et al. (2007). "Measuring the baryon acoustic oscillation scale using the sloan digital sky survey and 2dF galaxy redshift survey". In : *Monthly Notices of the Royal Astronomical Society* 381.3, p. 1053-1066.

Dai, X., F. Shankar et G. R. Sivakoff (2008). "2MASS reveals a large intrinsic fraction of BALQSOs". In : *The Astrophysical Journal* 672.1, p. 108.

Croom, S. M. et al. (2009). "The 2dF-SDSS LRG and QSO Survey : the spectroscopic QSO catalogue". In : *Monthly Notices of the Royal Astronomical Society* 392.1, p. 19-44.

Noterdaeme, P., P. Petitjean, C. Ledoux et R. Srianand (2009). "Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II–Data Release 7". In : *Astronomy & Astrophysics* 505.3, p. 1087-1098.

Percival, W. J. et M. White (2009). "Testing cosmological structure formation using redshift-space distortions". In : *Monthly Notices of the Royal Astronomical Society* 393.1, p. 297-308.

Peter, P. et J.-P. Uzan (2009). *Primordial cosmology*. Oxford University Press.

Bolton, A. S. et D. J. Schlegel (2010). "Spectro-perfectionism : an algorithmic framework for photon noise-limited extraction of optical fiber spectroscopy". In : *Publications of the Astronomical Society of the Pacific* 122.888, p. 248.

Kim, J. E. et G. Carosi (2010). "Axions and the strong C P problem". In : *Reviews of Modern Physics* 82.1, p. 557.

Percival, W. J., B. A. Reid et al. (2010). "Baryon acoustic oscillations in the Sloan Digital Sky Survey data release 7 galaxy sample". In : *Monthly Notices of the Royal Astronomical Society* 401.4, p. 2148-2168.

Wright, E. L. et al. (2010). "The Wide-field Infrared Survey Explorer (WISE) : mission description and initial on-orbit performance". In : *The Astronomical Journal* 140.6, p. 1868.

Bovy, J. et al. (2011). "Think outside the color box : Probabilistic target selection and the SDSS-XDQSO Quasar targeting catalog". In : *The Astrophysical Journal* 729.2, p. 141.

Le Goff, J. et al. (2011). "Simulations of BAO reconstruction with a quasar Ly-α survey". In : *Astronomy & Astrophysics* 534, A135.

Bolton, A. S., D. J. Schlegel et al. (2012). "Spectral classification and redshift measurement for the SDSS-III baryon oscillation spectroscopic survey". In : *The Astronomical Journal* 144.5, p. 144.

Font-Ribera, A., P. McDonald et J. Miralda-Escudé (2012). "Generating mock data sets for large-scale Lyman-α forest correlation measurements". In : *Journal of Cosmology and Astroparticle Physics* 2012.01, p. 001.

Font-Ribera, A. et J. Miralda-Escudé (2012). "The effect of high column density systems on the measurement of the Lyman-α forest correlation function". In : *Journal of Cosmology and Astroparticle Physics* 2012.07, p. 028.

Noterdaeme, P., P. Petitjean, W. Carithers et al. (2012). "Column density distribution and cosmological mass density of neutral gas : Sloan Digital Sky Survey-III Data Release 9". In : *Astronomy & Astrophysics* 547, p. L1.

Sholl, M. J. et al. (2012). "BigBOSS : a stage IV dark energy redshift survey". In : *Ground-based and Airborne Instrumentation for Astronomy IV*. T. 8446. SPIE, p. 1902-1913.

Delubac, T., J. Rich et al. (2013). "Baryon acoustic oscillations in the Lyα forest of BOSS quasars". In : *Astronomy & Astrophysics* 552, A96.

Font-Ribera, A., E. Arnau et al. (2013). "The large-scale quasar-Lyman α forest cross-correlation from BOSS". In : *Journal of Cosmology and Astroparticle Physics* 2013.05, p. 018.

Levi, M. et al. (2013). "The DESI Experiment, a whitepaper for Snowmass 2013". In : *arXiv preprint arXiv :1308.0847*.

Palanque-Delabrouille, N. et al. (2013). "The one-dimensional Lyα forest power spectrum from BOSS". In : *Astronomy & Astrophysics* 559, A85.

Rorai, A., J. F. Hennawi et M. White (2013). "A new method to directly measure the Jeans scale of the intergalactic medium using close quasar pairs". In : *The Astrophysical Journal* 775.2, p. 81.

Rudie, G. C., C. C. Steidel, A. E. Shapley et M. Pettini (2013). "The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift< z>= 2.4". In : *The Astrophysical Journal* 769.2, p. 146.

Smee, S. A. et al. (2013). "The multi-object, fiber-fed spectrographs for the sloan digital sky survey and the baryon oscillation spectroscopic survey". In : *The Astronomical Journal* 146.2, p. 32.

Ak, N. F. et al. (2014). "The dependence of C IV broad absorption line properties on accompanying Si IV and Al III absorption : Relating quasar-wind ionization levels, kinematics, and column densities". In : *The Astrophysical Journal* 791.2, p. 88.

Font-Ribera, A., D. Kirkby et al. (2014). "Quasar-Lyman α forest cross-correlation from BOSS DR11 : Baryon Acoustic Oscillations". In : *Journal of Cosmology and Astroparticle Physics* 2014.05, p. 027.

Prochaska, J. X., P. Madau, J. M. O'Meara et M. Fumagalli (2014). "Towards a unified description of the intergalactic medium at redshift z 2.5". In : *Monthly Notices of the Royal Astronomical Society* 438.1, p. 476-486.

Arinyo-i-Prats, A., J. Miralda-Escudé, M. Viel et R. Cen (2015). "The non-linear power spectrum of the Lyman alpha forest". In : *Journal of Cosmology and Astroparticle Physics* 2015.12, p. 017.

Delubac, T., J. E. Bautista et al. (2015). "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars". In : *Astronomy & Astrophysics* 574, A59.

Myers, A. D., N. Palanque-Delabrouille et al. (2015). "The SDSS-IV extended Baryon oscillation spectroscopic survey : Quasar target selection". In : *The Astrophysical Journal Supplement Series* 221.2, p. 27.

Ade, P. A. et al. (2016). "Planck 2015 results-xiii. cosmological parameters". In : *Astronomy & Astrophysics* 594, A13.

Aghamousa, A. et al. (2016a). "The DESI experiment part I : science, targeting, and survey design". In : *arXiv preprint arXiv :1611.00036*.

— (2016b). "The desi experiment part ii : Instrument design". In : *arXiv preprint arXiv :1611.00037*.

Dawson, K. S. et al. (2016). "The SDSS-IV extended Baryon Oscillation Spectroscopic Survey : overview and early data". In : *The Astronomical Journal* 151.2, p. 44.

Harris, D. W. et al. (2016). "The composite spectrum of BOSS quasars selected for studies of the Lyα forest". In : *The Astronomical Journal* 151.6, p. 155.

Shen, Y. et al. (2016). "The Sloan Digital Sky Survey reverberation mapping project : velocity shifts of quasar emission lines". In : *The Astrophysical Journal* 831.1, p. 7.

Albawi, S., T. A. Mohammed et S. Al-Zawi (2017). "Understanding of a convolutional neural network". In : *2017 international conference on engineering and technology (ICET)*. Ieee, p. 1-6.

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z= 2.3 with SDSS DR12 Lyα-Forests". In : *Astronomy & Astrophysics* 603, A12.

Blanton, M. R. et al. (2017). "Sloan digital sky survey IV : Mapping the Milky Way, nearby galaxies, and the distant universe". In : *The Astronomical Journal* 154.1, p. 28.

Garnett, R., S. Ho, S. Bird et J. Schneider (2017). "Detecting damped Ly α absorbers with Gaussian processes". In : *Monthly Notices of the Royal Astronomical Society* 472.2, p. 1850-1865.

Laurent, P. et al. (2017). "Clustering of quasars in SDSS-IV eBOSS : study of potential systematics and bias determination". In : *Journal of Cosmology and Astroparticle Physics* 2017.07, p. 017.

Prochaska, J., N. Tejos, C. Wotta et al. (2017). "pyigm/pyigm : Initial Release for Publications". In : *UCSC, Zenodo, doi* 10.

Raichoor, A., J. Comparat et al. (2017). "The SDSS-IV extended Baryon Oscillation Spectroscopic Survey : final emission line galaxy target selection". In : *Monthly Notices of the Royal Astronomical Society* 471.4, p. 3955-3973.

Ata, M. et al. (2018). "The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample : first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2". In : *Monthly Notices of the Royal Astronomical Society* 473.4, p. 4773-4794.

Balland, C. et al. (2018). "QuasarNET : Human-level spectral classification and redshifting with Deep Neural Networks". In : *arXiv e-prints*, arXiv-1808.

Blomqvist, M., M. M. Pieri et al. (2018). "The triply-ionized carbon forest from eBOSS : cosmological correlations with quasars in SDSS-IV DR14". In : *Journal of Cosmology and Astroparticle Physics* 2018.05, p. 029.

Miller, T. N. et al. (2018). "Fabrication of the DESI corrector lenses". In : *Advances in Optical and Mechanical Technologies for Telescopes and Instrumentation III*. T. 10706. SPIE, p. 256-264.

Parks, D., J. X. Prochaska, S. Dong et Z. Cai (2018). "Deep learning of quasar spectra to discover and characterize damped Lyα systems". In : *Monthly Notices of the Royal Astronomical Society* 476.1, p. 1151-1168.

Pérez-Ràfols, I., J. Miralda-Escudé, A. Arinyo-i-Prats, A. Font-Ribera et L. Mas-Ribas (2018). "The cosmological bias factor of damped Lyman alpha systems : dependence on metal line strength". In : *Monthly Notices of the Royal Astronomical Society* 480.4, p. 4702-4709.

Rogers, K. K., S. Bird, H. V. Peiris, A. Pontzen, A. Font-Ribera et B. Leistedt (2018a). "Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers". In : *Monthly Notices of the Royal Astronomical Society* 476.3, p. 3716-3728.

— (2018b). "Simulating the effect of high column density absorbers on the one-dimensional Lyman α forest flux power spectrum". In : *Monthly Notices of the Royal Astronomical Society* 474.3, p. 3032-3042.

Zarrouk, P. (2018a). "Clustering Analysis in Configuration Space and Cosmological Implications of the SDSS-IV eBOSS Quasar Sample". Thèse de doct. Université Paris-Saclay (ComUE).

— (2018b). "Clustering Analysis in Configuration Space and Cosmological Implications of the SDSS-IV eBOSS Quasar Sample". Thèse de doct., p. 17.

Blomqvist, M., H. D. M. Des Bourboux et al. (2019). "Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14". In : *Astronomy & Astrophysics* 629, A86.

Chabanier, S., N. Palanque-Delabrouille et al. (2019). "The one-dimensional power spectrum from the SDSS DR14 Lyα forests". In : *Journal of Cosmology and Astroparticle Physics* 2019.07, p. 017.

De Sainte Agathe, V. (2019). "Mesure de la position du pic d'oscillations acoustiques baryoniques dans les forêts Lyα et Lyβ des spectres des quasars du relevé eBOSS-SDSS IV". Thèse de doct. Sorbonne université.

de Sainte Agathe, V. et al. (sept. 2019a). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : 629, A85, A85. arXiv : 1904.03400 [astro-ph.CO].

Des Bourboux, H. D. M., K. S. Dawson et al. (2019). "The extended baryon oscillation spectroscopic survey : measuring the cross-correlation between the Mg ii flux transmission field and quasars and galaxies at z= 0.59". In : *The Astrophysical Journal* 878.1, p. 47.

Dey, A. et al. (2019). "Overview of the DESI legacy imaging surveys". In : *The Astronomical Journal* 157.5, p. 168.

Guo, Z. et P. Martini (2019). "Classification of Broad Absorption Line Quasars with a Convolutional Neural Network". In : *The Astrophysical Journal* 879.2, p. 72.

Hamann, F., H. Herbst, I. Paris et D. Capellupo (2019). "On the structure and energetics of quasar broad absorption-line outflows". In : *Monthly Notices of the Royal Astronomical Society* 483.2, p. 1808-1828.

Harpole, A., M. Zingale, I. Hawke et T. Chegini (fév. 2019). *pyro : a framework for hydrodynamics explorations and prototyping*. Version 3.1.

Sainte Agathe, V. de et al. (2019b). "Baryon acoustic oscillations at z= 2.34 from the correlations of Lyα absorption in eBOSS DR14". In : *Astronomy & Astrophysics* 629, A85.

Aghanim, N. et al. (2020). "Planck 2018 results-VI. Cosmological parameters". In : *Astronomy & Astrophysics* 641, A6.

Ahumada, R. 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". In : *The Astrophysical Journal Supplement Series* 249.1, p. 3.

Bautista, J. (2020). "Spectral Reductions, Redshifts, and Catalogs for Cosmology". In : *American Astronomical Society Meeting Abstracts# 235*. T. 235, p. 413-02.

Cuceu, A., A. Font-Ribera et B. Joachimi (2020). "Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman-α forest". In : *Journal of Cosmology and Astroparticle Physics* 2020.07, p. 035.

Des Bourboux, H. D. M., J. Rich et al. (2020). "The completed SDSS-IV extended baryon oscillation spectroscopic survey : baryon acoustic oscillations with Lyα forests". In : *The Astrophysical Journal* 901.2, p. 153.

Dodelson, S. et F. Schmidt (2020). *Modern cosmology*. Academic press.

Farr, J., A. Font-Ribera, H. D. M. Des Bourboux et al. (2020). "LyaCoLoRe : synthetic datasets for current and future Lyman-α forest BAO surveys". In : *Journal of Cosmology and Astroparticle Physics* 2020.03, p. 068.

Farr, J., A. Font-Ribera et A. Pontzen (2020). "Optimal strategies for identifying quasars in DESI". In : *Journal of Cosmology and Astroparticle Physics* 2020.11, p. 015.

Fumagalli, M., S. Fotopoulou et L. Thomson (2020). "Detecting neutral hydrogen at z 3 in large spectroscopic surveys of quasars". In : *Monthly Notices of the Royal Astronomical Society* 498.2, p. 1951-1962.

Ho, M.-F., S. Bird et R. Garnett (2020). "Detecting multiple DLAs per spectrum in SDSS DR12 with Gaussian processes". In : *Monthly Notices of the Royal Astronomical Society* 496.4, p. 5436-5454.

Lyke, B. W. et al. (2020). "The Sloan Digital Sky Survey Quasar Catalog : Sixteenth Data Release". In : *The Astrophysical Journal Supplement Series* 250.1, p. 8.

Yèche, C. et al. (2020). "Preliminary Target Selection for the DESI Quasar (QSO) Sample". In : *Research Notes of the AAS* 4.10, p. 179.

Alam, S. et al. (2021). "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey : Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory". In : *Physical Review D* 103.8, p. 083533.

Di Valentino, E. et al. (2021). "In the realm of the Hubble tension—a review of solutions". In : *Classical and Quantum Gravity* 38.15, p. 153001.

Ho, M.-F., S. Bird et R. Garnett (2021). "Damped Lyman-α absorbers from Sloan digital sky survey DR16Q with Gaussian processes". In : *Monthly Notices of the Royal Astronomical Society* 507.1, p. 704-719.

Ting Tan, Y. L. et C. Balland (2021). "Detection of the Damped Lyman-alpha systems in quasar spectra with machine learning algorithms". In : *2021 IAP colloquium*.

Abareshi, B. et al. (2022). "Overview of the instrumentation for the Dark Energy Spectroscopic Instrument". In : *The Astronomical Journal* 164.5, p. 207.

Abdurro'uf, N. et al. (2022). "The seventeenth data release of the Sloan Digital Sky Surveys : Complete release of MaNGA, MaStar, and APOGEE-2 data". In : *The Astrophysical Journal. Supplement Series* 259.2.

Angulo, R. E. et O. Hahn (2022). "Large-scale dark matter simulations". In : *Living Reviews in Computational Astrophysics* 8.1, p. 1.

Chabanier, S., T. Etourneau et al. (2022). "The Completed Sloan Digital Sky Survey IV Extended Baryon Oscillation Spectroscopic Survey : The Damped Lyα Systems Catalog". In : *The Astrophysical Journal Supplement Series* 258.1, p. 18.

Lan, T.-W. et al. (2022). "The DESI Survey Validation : Results from Visual Inspection of Bright Galaxies, Luminous Red Galaxies, and Emission Line Galaxies". In : *arXiv preprint arXiv :2208.08516*.

Mörtsell, E., A. Goobar, J. Johansson et S. Dhawan (2022). "The Hubble tension revisited : additional local distance ladder uncertainties". In : *The Astrophysical Journal* 935.1, p. 58.

Myers, A. D., J. Moustakas et al. (2022). "The Target Selection Pipeline for the Dark Energy Spectroscopic Instrument". In : *arXiv preprint arXiv :2208.08518*.

Ramirez-Pérez, C., J. Sanchez, D. Alonso et A. Font-Ribera (2022). "CoLoRe : fast cosmological realisations over large volumes with multiple tracers". In : *Journal of Cosmology and Astroparticle Physics* 2022.05, p. 002.

Ravoux, C. (2022). "One-and three-dimensional measurements of the matter distribution from eBOSS and first DESI Lyman-α forest samples". Thèse de doct. Université Paris-Saclay.

Silber, J. H. et al. (2022). "The Robotic Multi-Object Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)". In : *arXiv preprint arXiv :2205.09014*.

Stermer, J. (2022). "Utilisation de catalogues simulés pour les analyses BAO Lyman-alpha du relevé eBOSS". In.

Wang, B. et al. (2022). "Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems". In : *The Astrophysical Journal Supplement Series* 259.1, p. 28.

Youles, S. et al. (2022). "The effect of quasar redshift errors on Lyman-α forest correlation functions". In : *Monthly Notices of the Royal Astronomical Society* 516.1, p. 421-433.

Adame, A. et al. (2023). "The Early Data Release of the Dark Energy Spectroscopic Instrument". In : *arXiv preprint arXiv :2306.06308*.

Alexander, D. M. et al. (2023). "The DESI Survey Validation : Results from Visual Inspection of the Quasar Survey Spectra". In : *The Astronomical Journal* 165.3, p. 124.

Chaussidon, E. et al. (2023). "Target Selection and Validation of DESI Quasars". In : *The Astrophysical Journal* 944.1, p. 107.

Gordon, C. et al. (2023). "3D Correlations in the Lyman-$alpha$ Forest from Early DESI Data". In : *arXiv e-prints*, arXiv-2308.

Guy, J. et al. (2023). "The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument". In : *The Astronomical Journal* 165.4, p. 144.

Hahn, C. et al. (2023). "The DESI Bright Galaxy Survey : Final Target Selection, Design, and Validation". In : *The Astronomical Journal* 165.6, p. 253.

Raichoor, A., J. Moustakas et al. (2023). "Target Selection and Validation of DESI Emission Line Galaxies". In : *The Astronomical Journal* 165.3, p. 126.

Ramirez-Pérez, C., I. Pérez-Ràfols et al. (2023). "The Lyman-alpha forest catalog from the Dark Energy Spectroscopic Instrument Early Data Release". In : *arXiv preprint arXiv :2306.06312*.

Zhou, R. et al. (2023). "Target Selection and Validation of DESI Luminous Red Galaxies". In : *The Astronomical Journal* 165.2, p. 58.

Bailey, S. (in preparation). *Redrock : Spectroscopic Classification and Redshift Fitting for the Dark Energy Spectroscopic Instrumen*.

Etourneau, T. et al. (in preparation).

Herrera-Alcantar, H. K. et al. (in preparation). *DESI Lyman-alpha synthetic spectra*.

Ignasi et al. (in preparation). *Lya catalog paper*.

Jelinsky, P. et al. (2022, in prep). In.

Jiaqi, Z. et al. (in preparation). *The DESI Damped Lyα System Survey : Data Release 1*.

Miller, T. et al. (2022, in prep). In.

Poppett, C. et al. (2022, in prep). In.

Ting, T. et al. (in preparation). *Modeling of the High Column Density systems in The Lyman-α forest.*

Publication

I describe in this section one of my publications during my Ph.D. study, which is related to weak lensing.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **522,** 3766–3783 (2023) https://doi.org/10.1093/mnras/stad1142
Advance Access publication 2023 18

Assessing theoretical uncertainties for cosmological constraints from weak lensing surveys

Ting Tan,[1,2]★ Dominik Zürcher ,[2] Janis Fluri,[2,3] Alexandre Refregier,[2] Federica Tarsitano [2] and Tomasz Kacprzak [2]

[1]*Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies, LPNHE, 4 Place Jussieu, F-75252 Paris, France*
[2]*Institute for Particle Physics and Astrophysics, Department of Physics, ETH Zürich, Wolfgang Pauli Strasse 27, CH-8093 Zürich, Switzerland*
[3]*Data Analytics Lab, Department of Computer Science, ETH Zurich Universitätstrasse 6, CH-8006 Zürich, Switzerland*

Accepted 2023 April 6. Received 2023 April 6; in original form 2022 July 7

ABSTRACT

Weak gravitational lensing is a powerful probe, which is used to constrain the standard cosmological model and its extensions. With the enhanced statistical precision of current and upcoming surveys, high-accuracy predictions for weak lensing statistics are needed to limit the impact of theoretical uncertainties on cosmological parameter constraints. For this purpose, we present a comparison of the theoretical predictions for the non-linear matter and weak lensing power spectra, based on the widely used fitting functions (`mead` and `rev-halofit`), emulators (`EuclidEmulator` , `EuclidEmulator2` , `BaccoEmulator` , and `CosmicEmulator`), and *N*-body simulations (PKDGRAV3). We consider the forecasted constraints on the ΛCDM and wCDM models from weak lensing for stage III and stage IV surveys. We study the relative bias on the constraints and their dependence on the assumed prescriptions. Assuming a ΛCDM cosmology, we find that the relative agreement on the S_8 parameter is between 0.2 and 0.3σ for a stage III-like survey between the above predictors. For a stage IV-like survey the agreement becomes 1.4–3.0σ. In the wCDM scenario, we find broader S_8 constraints, and agreements of 0.18–0.26σ and 0.7–1.7σ for stage III and stage IV surveys, respectively. The accuracies of the above predictors therefore appear adequate for stage III surveys, whereas the fitting functions would need improvements for future stage IV surveys. Furthermore, we find that, of the fitting functions, `mead` provides the best agreement with the emulators. We discuss the implication of these findings for the preparation of future weak lensing surveys, and the relative impact of theoretical uncertainties to other systematics.

Key words: gravitational lensing: weak – large-scale structure of Universe – Cosmological parameters.

1 INTRODUCTION

The next generation of wide field cosmological surveys, such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST[1]; Abell et al. 2009), *Euclid*,[2] and the *Nancy Grace Roman Space Telescope* (*NGRST*[3]; Akeson et al. 2019) will map the matter distribution of the local Universe with an unprecedented accuracy. These high-precision measurements present a challenge for the theoretical modelling of cosmological observables. Cosmic shear is a cosmological observable that relies on the distortions of galaxy shapes caused by weak gravitational lensing (e.g. Bartelmann & Schneider 2001). This effect is due to the gravitational deflection of photons by the matter density field along the line of sight. Cosmic shear measures the inhomogeneities in the cosmic density field with high precision and can be used as an unbiased tracer of the matter distribution. It is sensitive to both, the matter distribution of the Universe and the growth of cosmic structure, which is important for the understanding of the expansion history of the Universe. A commonly used cosmic shear summary statistic is the cosmic shear angular power spectrum, which can be predicted from the matter power spectrum. The modelling of the matter power spectrum on large scales can be derived using perturbation theory (Bernardeau et al. 2002; Crocce & Scoccimarro 2006; Baumann et al. 2012; Crocce, Scoccimarro & Bernardeau 2012; Blas, Garny & Konstandin 2014; Blas et al. 2016; Foreman & Senatore 2016; Nishimichi, Bernardeau & Taruya 2016; Beutler et al. 2017; Cataneo et al. 2019; d'Amico et al. 2020), where the structure formation of the Universe is linear. Some extended perturbation theories (e.g. Chudaykin et al. 2020; D'Amico, Senatore & Zhang 2021) can provide an accurate model up to $k \sim 0.3\,h\,\mathrm{Mpc}^{-1}$. However, at non-linear, smaller scales, non-linear processes have a strong impact on the matter power spectrum, and perturbation theory is no longer valid.

In this work, we compare the theoretical predictions of the non-linear matter power spectrum, and the associated theoretical uncertainties on cosmological parameters from measurements of the cosmic shear angular power spectrum. The comparison includes some widely used models fitted from *N*-body simulations using analytical halo models: `halofit` (Smith et al. 2003) is fitted to low resolution, gravity-only *N*-body simulations, which is known to exhibit a non-

★ E-mail: ting.tan@lpnhe.in2p3.fr
[1]https://www.lsst.org.
[2]https://www.cosmos.esa.int/web/euclid/home.
[3]https://roman.gsfc.nasa.gov/.

negligible mismatch with current state-of-the-art hydrodynamic *N*-body simulations; `rev-halofit` (Takahashi et al. 2012), developed as the revisited version of `halofit` is used in the analysis of the Dark Energy Survey (DES; Amon et al. 2022); and `mead` (Mead et al. 2015), which is used in the analysis of the Kilo-Degree Survey combined with the VISTA Kilo-Degree Infrared Galaxy Survey (Giblin et al. 2021). Apart from the halo model fitting method, emulators are generated from the interpolation of a suite of *N*-body simulations, e.g. `CosmicEmulator` (Heitmann et al. 2009, 2013; Lawrence et al. 2017), `BaccoEmulator` (Angulo et al. 2020; Aricò et al. 2021), `EuclidEmulator` (Knabenhans et al. 2019) and its updated version `EuclidEmulator2` (Collaboration et al. 2020), `COSMOPOWER` (Mancini et al. 2022), and `GP emulator` (Giblin et al. 2019). In this study, `CosmicEmulator`, `BaccoEmulator`, `EuclidEmulator`, and `EuclidEmulator2` are representatively selected in the comparison at the level of the matter power spectrum, and a comparison between `rev-halofit`, `mead` and `EuclidEmulator` is also shown in Knabenhans et al. (2021). In order to estimate the theoretical uncertainties, we look at the weak lensing cosmological parameter constraints, by generating a forecast for a stage III, DES-like survey and a stage IV, *Euclid*-like survey. We take into account the parameters described by the standard ΛCDM cosmological model and the extended wCDM model. As a further investigation, we also discuss the relative impact of theoretical uncertainties compared to other systematics, such as baryonic effects, photometric redshift uncertainty (e.g. Huterer et al. 2006), shear bias (e.g. Bernstein & Jarvis 2002; Hirata et al. 2004), and galaxy intrinsic alignment (e.g. Heavens, Refregier & Heymans 2000).

This paper is organized as follows. In Section 2, we describe the theoretical framework, including three halo-model based fitting functions, `mead`, `halofit`, and `rev-halofit`; four power spectrum emulators extracted from *N*-body simulations: `CosmicEmulator`, `BaccoEmulator`, `EuclidEmulator` and `EuclidEmulator2`, and one *N*-body simulation code PKDGRAV3 (Potter, Stadel & Teyssier 2017). In Section 3, we present the method and the relevant codes used in this study. We summarize our results in Section 4 and our conclusions in Section 5.

2 THEORY

In this section, we describe the theoretical background of the matter power spectrum, weak lensing, and its angular power spectrum, as well as the different predictors of the matter power spectrum that we include in the comparisons.

2.1 Weak lensing

Considering the cosmic density field $\rho(\vec{r})$ at the position $\vec{r}$, the density contrast $\delta(\vec{r})$ is defined as the relative difference of $\rho(\vec{r})$ to the average density $\bar{\rho}$

$$\delta(\vec{r}) = \frac{\rho(\vec{r}) - \bar{\rho}}{\bar{\rho}}. \tag{1}$$

In Fourier space, the density contrast takes the following form:

$$\delta(\vec{k}) = \int \delta(\vec{r}) \exp{(i\vec{k}\cdot\vec{r})} \mathrm{d}^3\mathrm{r}. \tag{2}$$

Furthermore, the matter power spectrum $P(\vec{k})$ is defined as the correlation of the density contrast in Fourier space (Peebles 2020):

$$\langle \delta(\vec{k})\delta(\vec{k}')\rangle = (2\pi)^3 \delta_{\mathrm{D}}^{(3)}(\vec{k}+\vec{k}')P(\vec{k}), \tag{3}$$

where δ_{D} is the three-dimensional Dirac delta function. For full-sky surveys, the cosmic shear angular power spectrum is approximately identical to the convergence power spectrum (Bartelmann & Maturi 2016), which can be defined as a weighted integration along the line-of-sight over the matter power spectrum (Bartelmann & Schneider 2001), and simplified using the Kaiser–Limber approximation (Limber 1953; Kaiser 1992, 1998; LoVerde & Afshordi 2008). We follow the formalism of LoVerde & Afshordi (2008), Giannantonio et al. (2012), Kilbinger et al. (2017), Kitching et al. (2017), and Tarsitano et al. (2021) to compute the cross-correlated shear power spectrum with tomographic redshift bins i and j:

$$C_{\gamma}^{ij}(\ell) = \frac{9}{16}\left(\frac{H_0}{c}\right)^4 \Omega_{\mathrm{m}}^2 \int_0^{\chi_{\mathrm{h}}} \mathrm{d}\chi \mathrm{P}_{\mathrm{NL}}\left(\frac{\ell}{\mathrm{r}}, \chi\right) \frac{\mathrm{g_i}(\chi)\mathrm{g_j}(\chi)}{(\mathrm{ar}(\chi))^2}. \tag{4}$$

Here P_{NL} is the non-linear matter power spectrum, χ is the comoving distance, χ_{h} is the comoving horizon distance, r is the comoving angular diameter distance, Ω_{m} is the total matter density, $a = (1 + z)^{(-1)}$ is the scale factor, and $g(\chi)$ is the lensing efficiency function defined as:

$$g_i(\chi) = 2\int_{\chi}^{\chi_{\mathrm{h}}} \mathrm{d}\chi' \mathrm{n_i}(\chi) \frac{\mathrm{r}(\chi)\mathrm{r}(\chi'-\chi)}{\mathrm{r}(\chi')}, \tag{5}$$

with $n_i(\chi)$ being the normalized number density of the observed galaxies at a comoving distance χ.

2.2 Matter power spectrum

The matter power spectrum is a fundamental statistics to study the large-scale structure of the Universe. As seen above, it is, in particular, useful to predict the cosmic shear angular power spectrum. Therefore, it is necessary to have an accurate theoretical model for the matter power spectrum on all scales. On large scales and mildly non-linear scales, the matter power spectrum can be modelled using perturbation theory and some extended theories. On small scales, which are in the non-linear regime, these approaches are not suited to predict the power spectrum with the necessary precision, while other methods are developed with the use of a halo model or simulations.

2.2.1 Analytical predictions

A common way to model the matter power spectrum on these small scales is to empirically fit physically motivated formulas to measurements from *N*-body simulations, e.g. as done in Hamilton et al. (1991). Furthermore, modelling the density field as a collection of virialized haloes, the matter power spectrum can be approximated analytically using the statistics of haloes, and fitted to simulations or emulators (Ma & Fry 2000; Seljak 2000; Cooray & Sheth 2002).

In this study, we compare three halo-model based fitting functions: `mead`, `halofit`, and `rev-halofit`. `halofit` was built using a series of *N*-body simulations with a total of $N = 256^3$ particles and the box size from 84 to 240 Mpc h^{-1}. Using the halo model, the matter power spectrum is constructed with two terms, the one-halo term proposed by Ma & Fry (2000), Peacock & Smith (2000), Seljak (2000), Scoccimarro et al. (2001) and a two-halo term (Ma & Fry 2000; Seljak 2000; Scoccimarro et al. 2001) to describe the exclusion effects between dark matter haloes. The one-halo term indicates the correlation of the matter field of one single halo, which dominates on small scales, whereas the two-halo term describes the cross-correlation between different haloes, which has a strong impact on larger scales. Assuming that the haloes are distributed according to the halo mass function (Press & Schechter 1974; Sheth & Tormen 1999), the matter power spectrum modelled with this

approach can achieve a high precision on large scales. However, due to the lack of baryons and the relatively low resolution of the *N*-body simulations used in their study, `halofit` does not match high-resolution *N*-body simulations, giving an accuracy at the 5 per cent level at $k = 1\,h\,\mathrm{Mpc}^{-1}$ (Heitmann et al. 2010), and larger differences for $k > 1\,h\,\mathrm{Mpc}^{-1}$, which is insufficient for the non-linear regime. `rev-halofit` is a revised prescription of `halofit`, which provides a more accurate prediction of the matter power spectrum for $k < 30\,h\,\mathrm{Mpc}^{-1}$ and $z < 10$, with a 5 per cent level accuracy at $k = 1\,h\,\mathrm{Mpc}^{-1}$ and 10 per cent level accuracy at $k = 10\,h\,\mathrm{Mpc}^{-1}$. `rev-halofit` uses high-resolution *N*-body simulations for 16 cosmological models around the Wilkinson Microwave Anisotropy Probe (WMAP) best-fitting cosmological parameters. The *N*-body simulations were run with the `Gadget-2` *N*-body code (Springel, Yoshida & White 2001; Springel 2005a), 1024^3 particles in total, and the box size from 320 to 2000 Mpc h^{-1}. The power spectrum is fitted using an improved fitting formula with five more model parameters as compared to `halofit`. Several extended methods have been proposed to improve the halo model (Bird, Viel & Haehnelt 2012; Mohammed & Seljak 2014; Seljak & Vlah 2015). Here we only consider `mead` (Mead et al. 2015), which reaches an accuracy at the 5 per cent level for $k = 10\,h\,\mathrm{Mpc}^{-1}$ and $z < 2$. `mead` introduces more physical parameters in addition to the halo model, and is fitted to the 'Coyote Universe'(Heitmann et al. 2013) suite of high-resolution simulations, the same simulations used for the generation of `CosmicEmulator`. It also includes massive neutrinos (Mead et al. 2016) and baryonic effects e.g. active galactic nuclei (AGNs) feedback, supernovae explosions, and gas cooling. However, we only consider the dark-matter-only case in this study.

2.2.2 *Emulators*

The fitting functions based on halo models described in Section 2.2.1 can provide accurate non-linear power spectrum predictions for large *k*-modes and a wide redshift range, which can be used to predict cosmological observables. However, they also have limitations as the precision is not uniform for different cosmological parameters, and it is difficult for fitting functions to give a high precision below the 1 per cent level compared to high-resolution simulations. Power spectrum emulators are constructed following a different approach in which one interpolates the power spectrum from a set of *N*-body simulations within a certain range of relevant parameters, using interpolation methods, e.g. Gaussian processes regression (Heitmann et al. 2010, 2013; Angulo et al. 2020) or polynomial chaos expansion (Knabenhans et al. 2019; Collaboration et al. 2020). Compared to fitting functions, emulators usually provide consistent precision of the predictions for different *k*-modes. However, emulators also have limitations: First, the covered parameter space is limited, thus making it difficult to perform a likelihood analysis, for which one needs to explore a wide range of parameter values. Secondly, the ranges of *k* and redshift are also limited, making it difficult to compute the weak lensing cosmic shear observables for high ℓs, which require an integration over a large *k* range.

In this study, we compare four emulators: `CosmicEmulator` (Heitmann et al. 2016), `BaccoEmulator` (Angulo et al. 2020), `EuclidEmulator` (Knabenhans et al. 2019), and `EuclidEmulator2` (Collaboration et al. 2020), which are selected as representatives for different interpolation methods, i.e. CosmicEmulator using Gaussian processes regression, EuclidEmulator using polynomial chaos expansion, and BaccoEmulator using Neural network, and Gaussian processes regression. `CosmicEmulator` is fitted using a set of the 'Coyote Universe' simulations and the 'Mira-Titan Universe'(Lawrence et al. 2017) simulations. We use the latest version of the emulator (Heitmann et al. 2016), for which the 'Mira-Titan Universe' simulations were run with 3200^3 particles and a simulation volume of $(2100\,h^{-1}\mathrm{Mpc})^3$. The `CosmicEmulator` successfully achieves high-precision predictions of the power spectrum within the 4 per cent level for $k_{\mathrm{max}} = 5\,h\,\mathrm{Mpc}^{-1}$ and $z < 2$. It allows for the variation of various parameters, including the matter density Ω_{m}, the amplitude of density fluctuations σ_8, the baryon density Ω_{b}, the scalar spectral index n_{s}, the dark energy equation of state parameters w_0, and w_{a}, the dimensionless Hubble parameter h, the neutrino density Ω_ν, and the redshift z. `EuclidEmulator` uses a different emulation method using *N*-body simulations generated with the PKDGRAV3 code (Potter et al. 2017). It uses 100 simulations with 2048^3 particles in a $(1250\,h^{-1}\mathrm{Mpc})^3$ simulation volume. The non-linear correction is encoded as a boost factor adding up to the input linear power spectrum, achieving a precision at the 1 per cent level for predictions within the ranges $k < 1\,h\,Mpc^{-1}$ and $z < 1$. Knabenhans et al. (2019) demonstrated that `EuclidEmulator` agrees with `rev-halofit` at the 8 per cent level. As an updated version of `EuclidEmulator`, `EuclidEmulator2` is extended with dynamical dark energy and massive neutrinos, created with a larger parameter space and a modified version of the PKDGRAV3 *N*-body code. `EuclidEmulator2` provides a consistent accuracy with simulations at the 2 per cent level up to $k_{\mathrm{max}} = 10\,h\,\mathrm{Mpc}^{-1}$ for $z < 2$, and slightly lower accuracy for higher redshift $z \sim 3$. However, as `EuclidEmulator2` uses the amplitude of the primordial power spectrum A_{s} instead of σ_8 as input parameter, we use the following formula to transfer σ_8 into A_{s} (Hand et al. 2018):

$$A_{\mathrm{s}} = \left(\frac{\sigma_8}{\sigma_{8,0}}\right)^2 \times A_{\mathrm{s},0} \tag{6}$$

in our comparison, where $\sigma_{8,0} = 0.826$ and $A_{\mathrm{s},0} = 2.184 \times 10^{-9}$.

`BaccoEmulator` is another state-of-the-art emulator using an updated version of the `L-Gadget3` code (Springel 2005b; Angulo et al. 2012) with 4320^3 particles in a $(1440\,h^{-1}\,\mathrm{Mpc})^3$ simulation volume. It has a 2 per cent level accuracy over the redshift range $0 < z < 1.5$ and $k < 5\,h\,\mathrm{Mpc}^{-1}$.

2.2.3 *N-body simulations*

We also include in this study a comparison with a dark-matter-only *N*-body simulation run with PKDGRAV3, which is based on a binary tree algorithm. This code uses fifth order multipole expansions of the gravitational potential between particles and can achieve fast computational speeds with hardware acceleration. A comparison between PKDGRAV3 and the *N*-body codes, `Gadget-3`, `Gadget-4`, and `Ramses` is presented in Schneider et al. (2016) and Springel et al. (2021). The PKDGRAV3 simulations are the same as the ones used for `EuclidEmulator`, with 2048^3 particles in total and the box size of $L = 1250h^{-1}$Mpc. The details are presented in Knabenhans et al. (2019).

3 METHOD

In this work, we perform a comparison of predictors of the non-linear matter power spectrum, i.e. halo-model based fitting functions and emulators. We estimate the theoretical uncertainties of these predictors on the parameter constraint level by looking at the weak lensing cosmological parameter constraints from a stage III and a stage IV surveys. For each survey, we perform a comparison using

Table 1. Parameter settings for mock surveys: The stage IV survey is created using a four times larger survey area and galaxy density compared to the stage III survey. A deeper Smail redshift distribution is also used in the stage IV survey.

Survey	Stage III	Stage IV
Survey area [deg^2]	5000	20 000
Galaxy density [arcmin^{-2}]	5	20
Redshift distribution	Smail	Smail
Redshift bins	4	4
Redshift range[4]	0.025 ∼ 3.0	0.025 ∼ 3.0

[4]The presented redshift range refers to the considered range used in the generation process of the covariance matrix for the mock surveys. The range differs from the redshift range used for the predictors of the weak lensing power spectrum in Section 4.2, where we use [0.08,2] for the stage III survey and [0.08,3.0] for the stage IV survey.

the standard ΛCDM cosmological model and the extended wCDM model.

3.1 Survey parameters

The estimate of the theoretical uncertainties for cosmological parameters is realized by forecasting the constraints for a stage III and a stage IV surveys. The covariance matrix is estimated from simulations, as described in Section 3.2 below. Table 1 shows the parameter settings used for the generation of the mock galaxy surveys. Martinelli et al. (2021) suggests using $\ell_{\rm max} = 5000$ for stage IV-like surveys to probe deep into non-linear regime. However, in this study we use a more conservative limit of $\ell_{\rm max} = 1000$, and do not take into account baryonic effects.

We use Smail et al. (1995) distributions to model the global redshift distribution of the source galaxies for both the stage III and the stage IV surveys. The corresponding formulas and parameter settings for these two distributions are as follows:

$$n(z)_{\rm stageIII} = z^{\alpha} \exp\left[-\left(\frac{z}{z_0}\right)^{\beta}\right], \tag{7}$$

with $\alpha = 1.5$, $\beta = 1.1$, and $z_0 = 0.31$ and

$$n(z)_{\rm stageIV} = \left(\frac{z}{z_0}\right)^{\alpha} \exp\left[-\left(\frac{z}{z_0}\right)^{\beta}\right], \tag{8}$$

with $\alpha = 2.0$, $\beta = 1.5$, and $z_0 = 0.64$ (Martinelli et al. 2021). In both cases the source galaxies are randomly divided into four tomographic bins with equal number of galaxies in each bin, and a Gaussian convolution is performed so that they follow the schema in Amara & Réfrégier (2007). The four tomographic bins are chosen to reduce the computation time and for simplicity. This is a conservative choice for the estimation of theoretical uncertainty, but could be enough for a forecast comparison. As a result of the auto- and cross-combinations of these four redshift bins, we have 10 combinations of auto- and cross-correlations for the cosmic shear measurements (four auto-correlations and six cross-correlations). Fig. 1 shows the global and tomographic redshift distributions used in this study.

3.2 Covariance matrix

An accurate estimate of the survey covariance matrix is crucial for the correct calculation of the likelihood function. We estimate the covariance matrices for the stage III and stage IV survey setups described in Table 1 from numerical simulations, using the NGSF

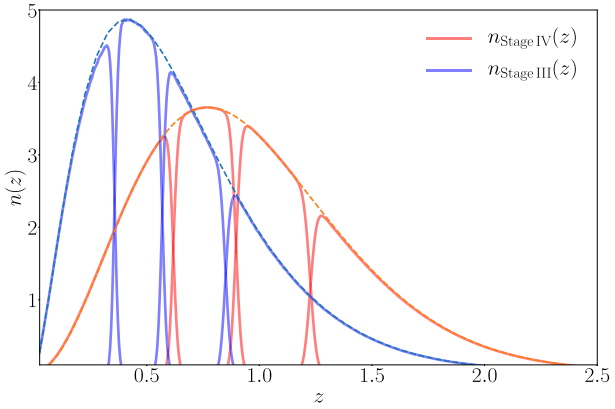

Figure 1. The redshift distributions of the source galaxies. One can see the four tomographic distributions for the stage III and the stage IV surveys. The global distributions, which follows the Smail et al. (1995) model, are shown by the dashed lines.

code described in Zürcher et al. (2021) and Dominiket et al. (2022). We generate a large number ($N = 2000$) of realization of the angular power spectra for each survey setup following the methodology outlined in Zürcher et al. (2021). In the following, we introduce the used N-body simulations, briefly summarize the forward modelling procedure used to generate the angular power spectra and describe the estimation of the covariance matrix. We refer the reader to Zürcher et al. (2021) for a more detailed description of the methodology.

We utilize the 50 independent PKDGRAV3 (Potter et al. 2017) N-body simulations at the fiducial cosmology that were previously used in Dominiket et al. (2022); Zürcher et al. (2021) and generated using the state-of-the-art dark-matter-only N-body code PKDGRAV3. The cosmological parameters in the used simulations are fixed to the (ΛCDM,TT,TE,EE+lowE+lensing) results of Planck 2018 (Aghanim et al. 2020), except for $\Omega_{\rm m}$ and σ_8 which are set to the values found in Troxel et al. (2018). This setup results in $\Omega_{\rm cdm} = 0.26$, $\sigma_8 = 0.84$, $\Omega_{\rm b} = 0.0493$, $n_{\rm s} = 0.9649$, $w = -1$, and $h = 0.6736$. We include three massive neutrino species in all simulations. The neutrinos are modelled as a relativistic fluid (Tram et al. 2019) and a degenerate mass hierarchy with a minimal neutrino mass of $m_\nu = 0.02$ eV per species was chosen. The dark energy density Ω_Λ is adapted for each cosmology to achieve a flat geometry.

Each simulation was run using a unit box with a side-length of 900 Mpc h^{-1} and 768^3 simulated particles. In order to achieve a simulation volume large enough to cover the redshift range up to $z = 3.0$ the unit box was replicated up to 14 times per dimension depending on the cosmology. While such a replication scheme is known to underpredict the variance of very large, superbox modes (Fluri et al. 2019), it has been demonstrated by Dominiket et al. (2022) that the simulations accurately recover the angular power spectra predicted by the theory code CLASS (Lesgourgues 2011) for $\ell \in [30, 2048]$.

The particle shells from each PKDGRAV3 simulation are combined into tomographic full-sky mass maps using the UFALCON software (Sgier et al. 2019). The particle shells are weighted according to the tomographic redshift distributions shown in Fig. 1. The UFALCON software uses the HEALPIX (Gorski et al. 2005) pixelization scheme to pixelize the sphere. A resolution of NSIDE = 1024 was chosen. UFALCON also makes use of the Born approximation, which is known to deteriorate the accuracy of the produced mass maps. However, Petri, Haiman & May (2017) have demonstrated that the introduced bias is negligible for stage III-like and stage IV-like surveys.

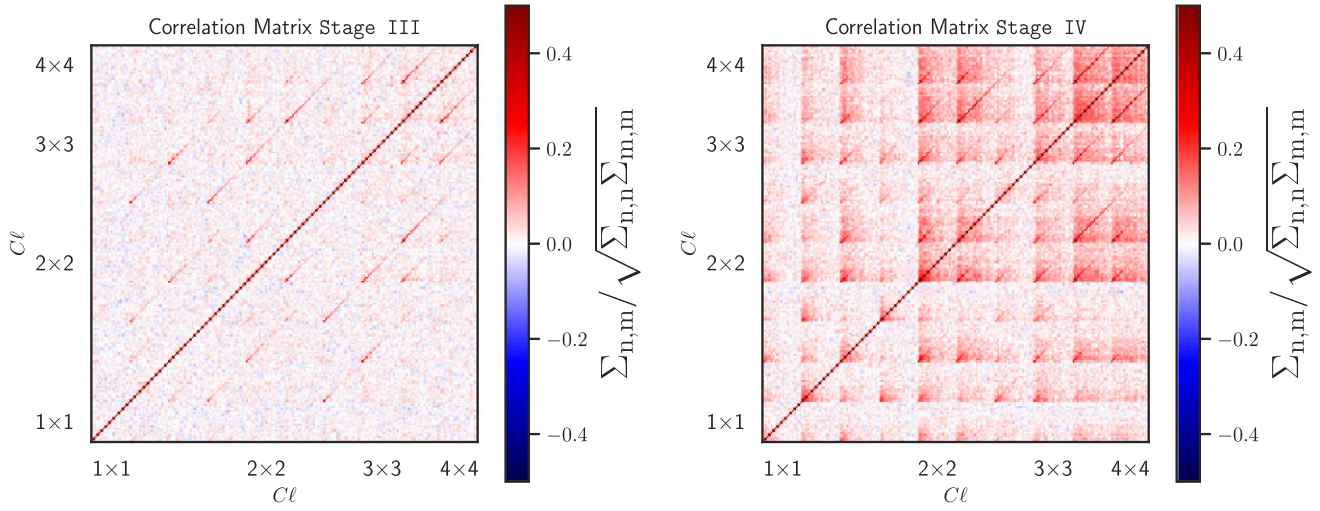

Figure 2. Correlation matrices for the stage III survey (left-hand panel) and the stage IV survey (right-hand panel). The ordering of the redshift tomographic bin combinations for the angular power spectra is 1 × 1, 1 × 2, 1 × 3, 1 × 4, 2 × 2, 2 × 3, 2 × 4, 3 × 3, 3 × 4, and 4 × 4, from left to right. For each angular power spectrum, all 20 bins ranging from $\ell = 100$ to $\ell = 1000$ are shown.

The spherical Kaiser–Squires mass mapping technique (Kaiser & Squires 1993; Wallis et al. 2022) is used to obtain the cosmic shear signal from the simulated mass maps. To forward-model a realistic weak lensing survey a shape noise signal must then be added to the cosmic shear signal and an appropriate survey mask must be applied. The survey masks are regularly chosen such that we obtain eight stage III surveys and two stage IV surveys from each full-sky map.

The shape noise signal is obtained in the same way as described in Zürcher et al. (2021). We randomly sample galaxy positions within the survey region until the target source density is reached. The intrinsic ellipticities of the galaxies are then drawn from a probability distribution that was fit to the observed galaxy ellipticities in Troxel et al. (2018) (see Zürcher et al. 2021). The ellipticity of each individual galaxy is rotated by a random phase. Using 5 and 20 shape noise realization per survey patch, we achieve the desired number of $N = 2000$ survey realization for the stage III and stage IV survey setup, respectively.

The tomographic angular power spectra realization $C_{\ell,\mathrm{i}}$ are then measured from the forward-modelled surveys using the `anafast` routine of the HEALPY software (Zonca et al. 2019) using 20 bins from $\ell_{\mathrm{min}} = 100$ to $\ell_{\mathrm{max}} = 1000$, the same as Sgier et al. (2019), where the index i runs over the number of survey realization N. The covariance matrix Σ is estimated according to

$$\hat{\Sigma} = \frac{1}{N-1}\sum_{\mathrm{i}=1}^{N}(C_{\ell,\mathrm{i}} - \bar{C}_\ell)(C_{\ell,\mathrm{i}} - \bar{C}_\ell)^T, \tag{9}$$

where $\bar{C}_\ell$ indicates the mean of the angular power spectra realization $C_{\ell,i}$. The estimated correlation matrices $C_{\mathrm{n,m}} \equiv \Sigma_{\mathrm{n,m}}/\sqrt{\Sigma_{\mathrm{n,n}}\Sigma_{\mathrm{m,m}}}$ are presented in Fig. 2.

3.3 Likelihood analysis

We use a Bayesian likelihood approach to evaluate the cosmological parameter constraints of different predictors. We assume a Gaussian error model and the likelihood is realized by:

$$\log\mathcal{L} = -\frac{1}{2}\sum_{ij}\left(C^i_{\ell,\mathrm{truth}} - C^i_{\ell,\mathrm{compare}}\right)^T \Sigma^{-1}\left(C^j_{\ell,\mathrm{truth}} - C^j_{\ell,\mathrm{compare}}\right) \tag{10}$$

Table 2. The fiducial values for the cosmological parameters and the flat priors for the cosmological parameters that are varied in the analysis.

Parameters	Fiducial values	Priors (stage III survey)	Priors (stage IV survey)
Ω_{m}	0.291	[0, 0.6]	[0.2, 0.4]
n_{s}	0.969	[0.3, 2.0]	[0.7, 1.2]
h	0.69	[0.1, 2.5]	[0.4, 0.9]
σ_8	0.826	[0.3, 1.4]	[0.7, 0.95]
w_0	− 1.0	[−3.5, 0.5]	[−2.5, 0.5]
Ω_{b}	0.0473		

Here $C_{\ell,\mathrm{truth}}$ stands for the value of the observable, computed using `PyCosmo` (Refregier et al. 2018; Tarsitano et al. 2021; Moser et al. 2022) with a chosen predictor and the fiducial cosmological parameters, measured by the Wilkinson Microwave Anisotropy Probe satellite (WMAP) 9 (Hinshaw et al. 2013), presented in Table 2. $C_{\ell,\mathrm{compare}}$ is predicted using another predictor for comparison. The cosmology for the observable is different from what is used for the covariance matrix. However, this effect is neglected assuming the covariance matrix parameter independent (Kodwani, Alonso & Ferreira 2018). Σ^{-1} is the unbiased estimate of the inverse covariance matrix (Hartlap Simon & Schneider 2007; Percival et al. 2014) represented as:

$$\Sigma^{-1} = \frac{N - N^{'} - 2}{N-1}\hat{\Sigma}^{-1}, \tag{11}$$

N is the number of realization generated from the simulations and $N^{'}$ is the total number of data bins, which is given by

$$N^{'} = N_{\mathrm{redshift}} \times N_\ell. \tag{12}$$

Here, we have $N = 2000$, $N_\ell = 20$, and $N_{\mathrm{redshift}} = 10$.

3.4 Parameter inference

The posterior is sampled efficiently using the Markov Chain Monte Carlo (MCMC) ensemble sampler, `emcee` (Foreman-Mackey et al. 2013). We vary four cosmological parameters $\{\Omega_{\mathrm{m}}, \sigma_8, n_{\mathrm{s}}, \text{and } h\}$ for the ΛCDM cosmological model and an additional parameter w_0

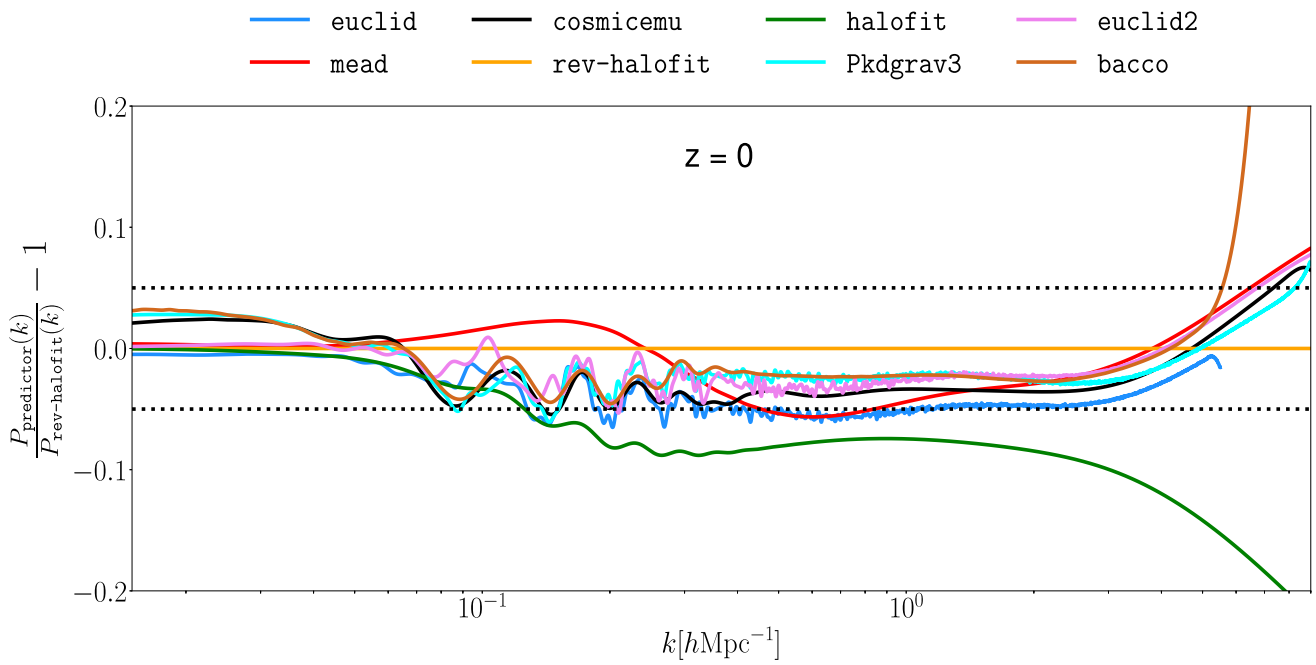

Figure 3. Comparison of dark-matter-only, non-linear $P(k)$ predictions for different predictors at redshift $z = 0$, subtracted and divided by `rev-halofit` as reference.

for the extended `wCDM` model, where we fix $w_a \equiv 0$. Table 2 shows the priors used for these parameters. We run the MCMC chains with 100 walkers per parameter and cut the burn in phase for each run as one- third of the chain length. Each individual chain has more than 100 000 samples. For the visualization of the marginalized posteriors, we use the public `Getdist` (Lewis 2019).

4 RESULTS

We present the results of our comparison of different predictors in this section, including the analysis of the matter power spectrum, the weak lensing power spectrum, and the cosmological parameter constraints based on the stage III and stage IV weak lensing surveys.

4.1 Power spectrum

We use the linear power spectrum predicted by `PyCosmo` and generated the following Eisenstein & Hu (1999) as the input for all predictors. Fig. 3 shows the comparison of dark-matter-only non-linear $P(k)$ predictions from different predictors at redshift $z = 0$, and the comparison for different redshifts ranging from $z = 0$ to $z = 5$ in Appendix A. The results are shown for k ranging from $k = 0.01$ to $9\,h\,\mathrm{Mpc}^{-1}$ using 10 000 bins. `BaccoEmulator` and `CosmicEmulator` are not valid for $z > 3$, so we do not present their comparison for the higher redshift at $z = 5$. Figs 3 and A1 indicate that:

(i) All the predictors except for `halofit` are within the 5 per cent level of accuracy compared to `rev-halofit` for $z < 2$ and $k < 7\,h\,\mathrm{Mpc}^{-1}$ (`BaccoEmulator` is valid for $z < 1.5$ and $k < 5\,h\,\mathrm{Mpc}^{-1}$, see the details in Fig. A1). Note that this is consistent with the comparison of `mead`, `rev-halofit` and `halofit` in Mead et al. (2015).

(ii) `halofit` shows stronger discrepancies compared with the other predictors at small scales for $k > 0.1\,h\,\mathrm{Mpc}^{-1}$ and this discrepancy can reach 20 per cent for $k \sim 10\,h\,\mathrm{Mpc}^{-1}$.

(iii) `mead` and `rev-halofit` show close agreement with the emulators at the 5 per cent level for $k < 9\,h\,\mathrm{Mpc}^{-1}$ and $z < 0.5$. However, at higher redshifts $1 < z < 5$, the discrepancies between `mead` and the emulators can reach 10 per cent for $k > 3\,h\,\mathrm{Mpc}^{-1}$, whereas `rev-halofit` provides a more consistent precision within 5 per cent.

(iv) All the emulators yield an agreement within the 2 − 3 per cent level compared with the PKDGRAV3 simulation for $k < 9\,h\,\mathrm{Mpc}^{-1}$ and $z < 1.5$. However, this is not valid at higher redshifts. The disagreement at higher redshifts between emulators and PKDGRAV3 might be due to the fact that emulators were built by interpolation within a certain parameter range, thus the accuracies could not be ensured beyond this range.

(v) For large scales with $k < 0.5\,h\,\mathrm{Mpc}^{-1}$, the different predictors show a better agreement at higher redshifts.

4.2 Weak lensing power spectrum

We compute the weak lensing shear power spectrum C_ℓ for the stage III and the stage IV survey with different predictors. Limited by the range of k_{max} of the emulators, the C_ℓs are computed using 20 ℓ-bins spaced linearly between $\ell_{\mathrm{min}} = 100$ and $\ell_{\mathrm{max}} = 1000$ (A further investigation on the impact of varying ℓ_{max} is presented in Appendix B2). The integrated redshift range is [0.08, 2.0] for the stage III survey and [0.08, 3.0] for the stage IV survey. This setting was chosen in order to avoid the instability of emulators for low redshifts, where we found that `EuclidEmulator` and `EuclidEmulator2` predict the C_ℓs with a discrepancy larger than 10 per cent at $z < 0.08$. This choice differs from the setting used for the generation of the covariance matrix. However, we find that this only changes the discrepancies between different predictors for C_ℓs by 0.1 per cent, since only 1 per cent of the low-redshift galaxies are missed for the stage III survey and 0.1 per cent of the galaxies for the stage IV survey. Using this redshift range, we have to exclude `CosmicEmulator` from the comparison for the

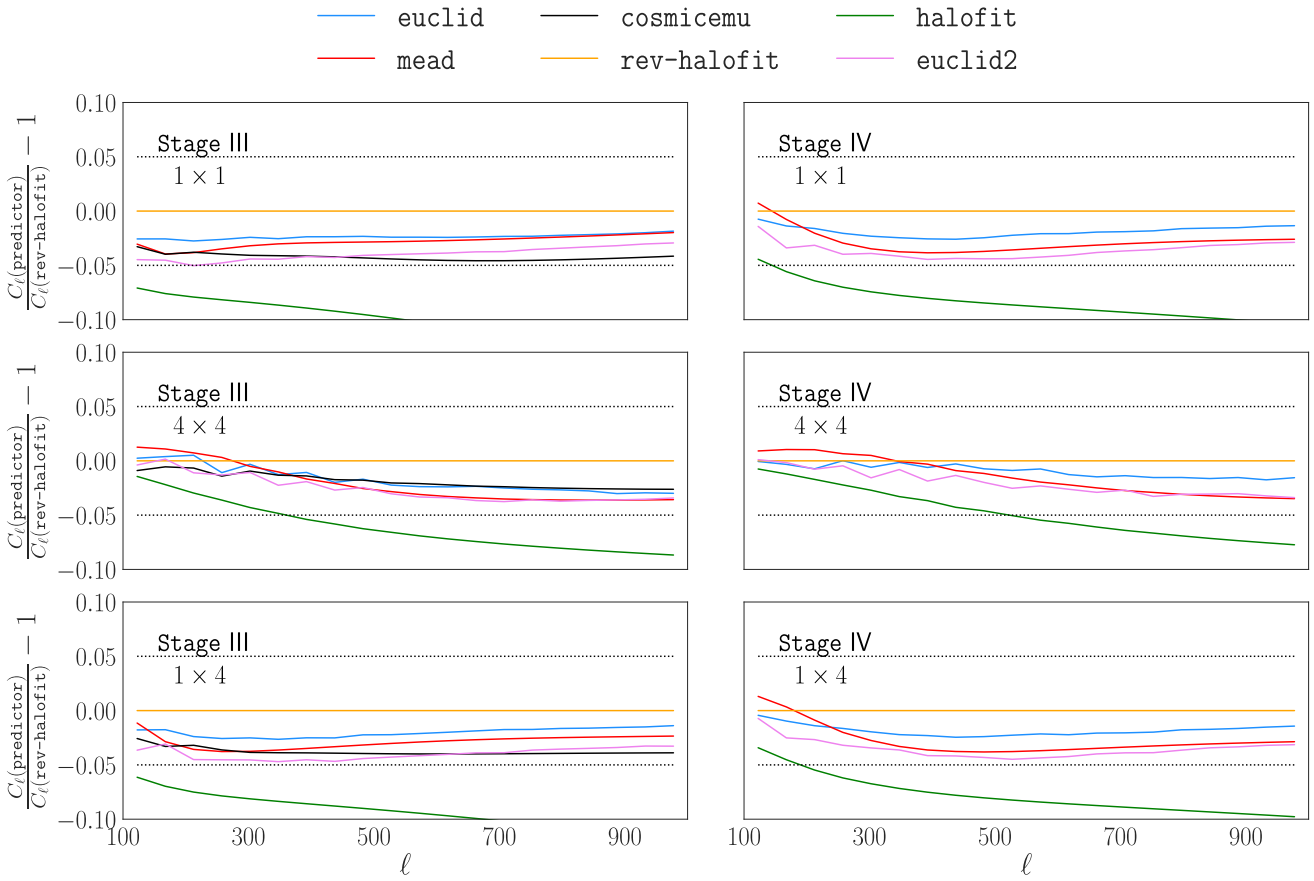

Figure 4. The comparison of weak lensing shear C_ℓs for different predictors. Each C_ℓ is multiplied by $\ell(\ell+1)/2\pi$. The upper two panels in each column show the auto-correlated C_ℓs for the first, and the fourth redshift bin and the bottom ones show the cross correlated C_ℓs between these two bins. The left-hand panels show the plots for the stage III survey and the right-hand side shows the stage IV survey results.

stage IV survey as it allows only up to $z = 2.0$ (`BaccoEmulator` is also excluded due to the redshift range up to $z = 1.5$). The comparison is shown in Fig. 4, with the left-hand panels showing the results for the stage III survey and the right-hand side showing the stage IV survey results. In the individual panels, we present $C_\ell\,\ell(\ell+1)/2\pi$ for each predictor and illustrate the comparison by subtracting and dividing `rev-halofit` as the reference. In Fig. 4 the first row shows the comparison of the auto-correlated C_ℓs for the redshift bins 1×1, the second row for 4×4, and the bottom row shows the cross correlated C_ℓs for 1×4. From Fig. 4, one can infer that:

(i) All the predictors, except for `halofit`, yield an agreement at the 5 per cent level, both for the auto and cross C_ℓ. This is consistent with our results for $P(k)$.
(ii) `mead` shows a good agreement with `CosmicEmulator`, `EuclidEmulator2`, and `EuclidEmulator`, whereas `rev-halofit` exhibits a larger discrepancy.
(iii) The comparison of C_ℓ for different predictors does not show a significant difference between the stage III and the stage IV survey.

4.3 Cosmological parameters constraints

The comparison of the weak lensing cosmological parameter constraints for different predictors is present in this section. As indicated in Section 3, we consider a stage III and a stage IV surveys. For each survey, we perform a comparison using the standard ΛCDM cosmological model and the extended wCDM model. A summary of the constraints on $\{S_8, \Omega_m, w_0\}$ is presented in Table 3, and the constraints on $\{S_8, \Omega_m, n_s, h, w_0\}$ in Table B1.

4.3.1 ΛCDM cosmology constraints

We present the two-dimensional 68 per cent and 95 per cent confidence level contours of the posterior distributions for the ΛCDM model in Figs 5 and 6 for the stage III and stage IV survey setup, respectively. The parameters $\{\Omega_m, \sigma_8, n_s, h\}$ are varied in the MCMC analysis. We additionally compute the constraints on S_8, and summarize the shifts in S_8 in Fig. 9, presenting the median values of the posteriors and the error bars indicating the 68 per cent confidence limits of the constraints. For two different predictors, the significance of disagreement is computed by dividing the difference of their means by their combined uncertainties. One can infer from the posterior distributions in Fig. 9 and Table B1 that the agreement on S_8 between different predictors is less than 0.6σ for the stage III survey ($0.2 - 0.3\sigma$ if `halofit` excluded), while being much larger for the stage IV survey. This is caused by the higher constraining power of the stage IV survey. More specifically, the agreements are generally on the $1.4 - 6.1\sigma$ level ($1.4 - 3.0\sigma$ if `halofit` excluded). `mead` shows good agreement with `CosmicEmulator`, `EuclidEmulator`, and `EuclidEmulator2` for the stage III survey while it only agrees well with `EuclidEmulator2` for the stage IV survey. The constraints on h do not show significant

Table 3. Numerical constraints on the cosmological parameters corresponding to the contours in Figs 5, 6, 7, and 8. For each predictor, the σs show the theoretical discrepancies for each parameter, compared to the reference one.

Survey cosmology	Predictor ref: `rev-halofit`	S_8	(σ)	Ω_m	(σ)	w_0	(σ)
Stage III	`rev-halofit`	$0.8147^{+0.0241}_{-0.0203}$		$0.288^{+0.0817}_{-0.0662}$			
	`mead`	$0.8035^{+0.0269}_{-0.0202}$	0.33	$0.2996^{+0.0848}_{-0.0698}$	0.11		
ΛCDM	`halofit`	$0.7946^{+0.0292}_{-0.0201}$	0.57	$0.2884^{+0.0783}_{-0.074}$	0.0		
	`euclid`	$0.8083^{+0.0256}_{-0.0201}$	0.2	$0.2987^{+0.0831}_{-0.0709}$	0.1		
	`cosmicemu`	$0.8047^{+0.0285}_{-0.018}$	0.29	$0.2916^{+0.0789}_{-0.0741}$	0.03		
	`euclid2`	$0.8031^{+0.0269}_{-0.0177}$	0.34	$0.2887^{+0.0835}_{-0.0679}$	0.01		
Stage III	`rev-halofit`	$0.8165^{+0.0433}_{-0.0661}$		$0.2846^{+0.092}_{-0.09}$		$-0.9242^{+0.4704}_{-2.294}$	
	`mead`	$0.7947^{+0.0497}_{-0.0588}$	0.26	$0.31^{+0.0824}_{-0.1022}$	0.18	$-1.139^{+0.647}_{-2.2626}$	0.09
wCDM	`halofit`	$0.7879^{+0.0517}_{-0.0612}$	0.34	$0.2968^{+0.0787}_{-0.1011}$	0.09	$-1.1333^{+0.6581}_{-2.3122}$	0.09
	`euclid`	$0.7977^{+0.0545}_{-0.0542}$	0.22	$0.3049^{+0.085}_{-0.1017}$	0.15	$-1.1886^{+0.7187}_{-2.1508}$	0.11
	`cosmicemu`	$0.7982^{+0.0504}_{-0.0572}$	0.22	$0.2931^{+0.0921}_{-0.0969}$	0.06	$-1.1408^{+0.6926}_{-2.3046}$	0.09
	`euclid2`	$0.8018^{+0.0461}_{-0.0627}$	0.18	$0.2896^{+0.0928}_{-0.0877}$	0.04	$-1.0254^{+0.5498}_{-2.2745}$	0.04
Stage IV	`rev-halofit`	$0.8135^{+0.0023}_{-0.0024}$		$0.2915^{+0.0077}_{-0.0084}$			
	`mead`	$0.8028^{+0.0027}_{-0.0026}$	2.96	$0.3008^{+0.0094}_{-0.0074}$	0.87		
ΛCDM	`halofit`	$0.7944^{+0.002}_{-0.0029}$	6.11	$0.2856^{+0.0097}_{-0.0064}$	0.46		
	`euclid`	$0.8094^{+0.0018}_{-0.003}$	1.37	$0.2917^{+0.0079}_{-0.0084}$	0.02		
	`euclid2`	$0.8058^{+0.0017}_{-0.0032}$	2.62	$0.2926^{+0.0079}_{-0.0084}$	0.1		
Stage IV	`rev-halofit`	$0.8127^{+0.0079}_{-0.0063}$		$0.2909^{+0.0095}_{-0.0086}$		$-1.0127^{+0.1171}_{-0.1046}$	
	`mead`	$0.7968^{+0.0067}_{-0.0069}$	1.73	$0.2979^{+0.0106}_{-0.0092}$	0.53	$-1.106^{+0.1107}_{-0.1163}$	0.61
wCDM	`halofit`	$0.7902^{+0.007}_{-0.0073}$	2.39	$0.2856^{+0.0093}_{-0.0096}$	0.42	$-1.0646^{+0.1069}_{-0.1197}$	0.35
	`euclid`	$0.8061^{+0.0073}_{-0.0072}$	0.68	$0.2908^{+0.0088}_{-0.0094}$	0.01	$-1.046^{+0.1142}_{-0.1288}$	0.22
	`euclid2`	$0.7996^{+0.0078}_{-0.0069}$	1.31	$0.2901^{+0.0099}_{-0.0094}$	0.06	$-1.0965^{+0.1288}_{-0.1255}$	0.51

discrepancies for both surveys, while n_s reveals discrepancies of several σs for different predictors for the stage IV survey.

4.3.2 *wCDM cosmology constraints*

We consider the constraining power of weak lensing surveys on dark energy parameters by adopting a time-dependent dynamical dark energy equation of state, the CPT-parametrization (Chevallier & Polarski 2001; Linder 2003), as an extension to the ΛCDM model. The equation-of-state parameter is given by

$$w(a) = w_0 + w_a(1-a), \tag{13}$$

where we use a fixed $w_a = 0$ and a free w_0. We present the two-dimensional marginal posterior distributions for the wCDM cosmology parameters in Figs 7 and 8, for the stage III and the stage IV surveys, respectively. Taking into account the dark energy model changes the shape and the contour size of the posterior distributions, decreasing the constraining power on the cosmological parameters. The discrepancies in S_8 between predictors are generally smaller compared with the ΛCDM model due to the decrease in constraining power: $0.18-0.34\sigma$ for the stage III survey and $0.7-2.4\sigma$ for the stage IV survey ($0.18-0.26\sigma$ and $0.7-1.7\sigma$ if `halofit` is excluded, respectively). `mead` shows relatively good agreement with `EuclidEmulator` and `EuclidEmulator2` for both the stage III and the stage IV surveys. `rev-halofit` agrees with all the predictors within 0.3σ for the stage III survey, and shows discrepancies at the $0.7-2.4\sigma$ level for the stage IV survey. Furthermore, we also consider the case with both free w_0 and w_a (With a flat prior $[-2,1]$). Compared with the case with a fixed w_a, this setting gives a tiny impact on the discrepancies between different predictors for $\{S_8, n_s, h\}$. However, it obtains weaker constraints on $\{\Omega_m, w_0\}$, resulting in the good agreements between different predictors. The discrepancies on w_a are within 0.5σ.

4.4 Systematic effects

In this study, we include dark-matter-only predictions, without any consideration of baryonic effects, which can have a strong impact on small scales (Jing et al. 2006; Rudd, Zentner & Kravtsov 2008), and the computation of the matter power spectrum (van Daalen et al. 2011; Casarini et al. 2012; Castro et al. 2018; Debackere, Schaye & Hoekstra 2020). Current studies of halo-model based fitting functions already include other systematics, i.e. massive neutrino and baryonic effects like AGN feedback and gas cooling. The inclusion of these systematics will significantly reduce the constraining power, and might alleviate the discrepancies between the predictors. The impact of taking into account the baryonic effects on cosmic shear can be found in Semboloni et al. (2011) and Martinelli et al. (2021), which indicates that including different baryonic models leads to discrepancies with $<0.5\sigma$ on cosmological parameter constraints for $\ell_{max} = 1500$, and more significant biases (A few σs) for higher $\ell_{max} \sim 5000$. However, it does not broaden significantly the constraints in both cases. In our scenario where ℓ_{max} is fixed to 1000, it can be foreseen that including baryons will involve a non-negligible impact on the agreements between different predictors.

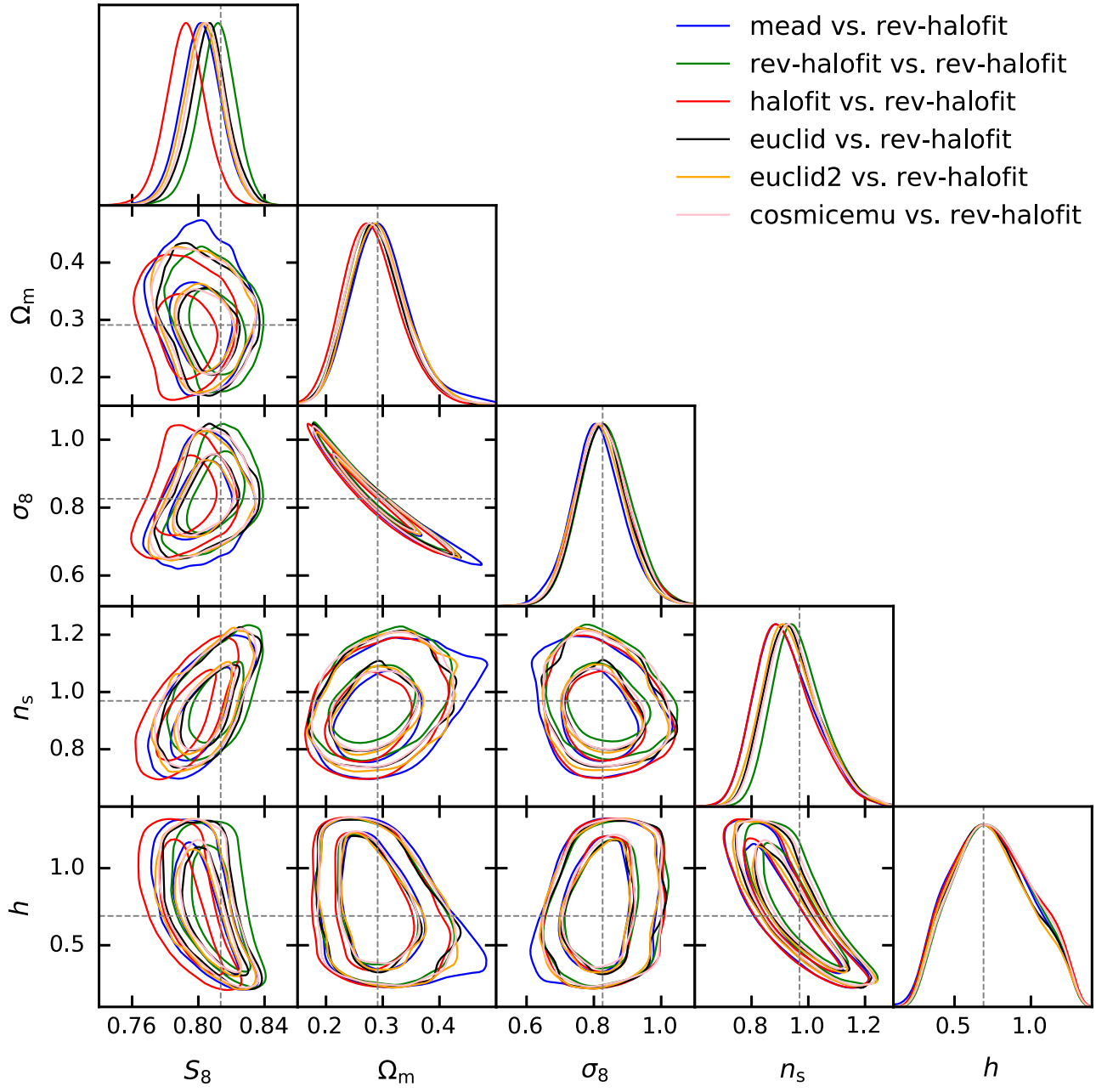

Figure 5. Cosmological parameter constraints for the stage III survey in the ΛCDM model. For each constraint, $C_{\ell,\mathrm{truth}}$ is predicted using the first predictor shown in the legend, and $C_{\ell,\mathrm{compare}}$ computed using the second predictor, as indicated in Section 3.3. For the stage III survey, we set $C_{\ell,\mathrm{truth}}$ with the halo-model based fitting functions (`rev-halofit`, `mead`, and `halofit`) and three emulators (`EuclidEmulator`, `EuclidEmulator2`, and `CosmicEmulator`), and compare with predictions from only the fitting functions (in this figure only `rev-halofit`).

In practice, there are also other sources of uncertainties in weak lensing experiments, such as photometric redshift uncertainty (Huterer et al. 2006; Choi et al. 2016; Hildebrandt et al. 2020), shear bias (Bernstein & Jarvis 2002; Hirata et al. 2004; Bernstein 2010; Melchior & Viola 2012; Refregier et al. 2012), and galaxy intrinsic alignment (Heavens et al. 2000; Hirata & Seljak 2004; Bridle & King 2007; Joachimi et al. 2011; Fluri et al. 2019). These systematics effects will contribute to the total error budget and broaden the constraints on cosmological parameters. In our analysis, we computed the impact of theoretical uncertainties and compared them to statistical errors. This is useful to allocate a given budget to this source of error, independently of the choices in the treatment of the other systematics. However, it is also useful to estimate the fraction of the theoretical statistical errors compared with these systematic errors, in order to study their contribution to the total error budget.

For this purpose, we estimate the impact of these systematics by considering other works which have carried out measurements and forecasts for stage III and stage IV surveys. For DES-like stage-III surveys, we can infer from Secco et al. (2022) and Amon et al. (2022), that the constraining power on S_8 will be decreased by ~ 20 per cent when considering the intrinsic alignment models, and less than ~ 5 per cent when considering the photometric redshift uncertainties and shear bias. For LSST-like stage-IV surveys, Krause,

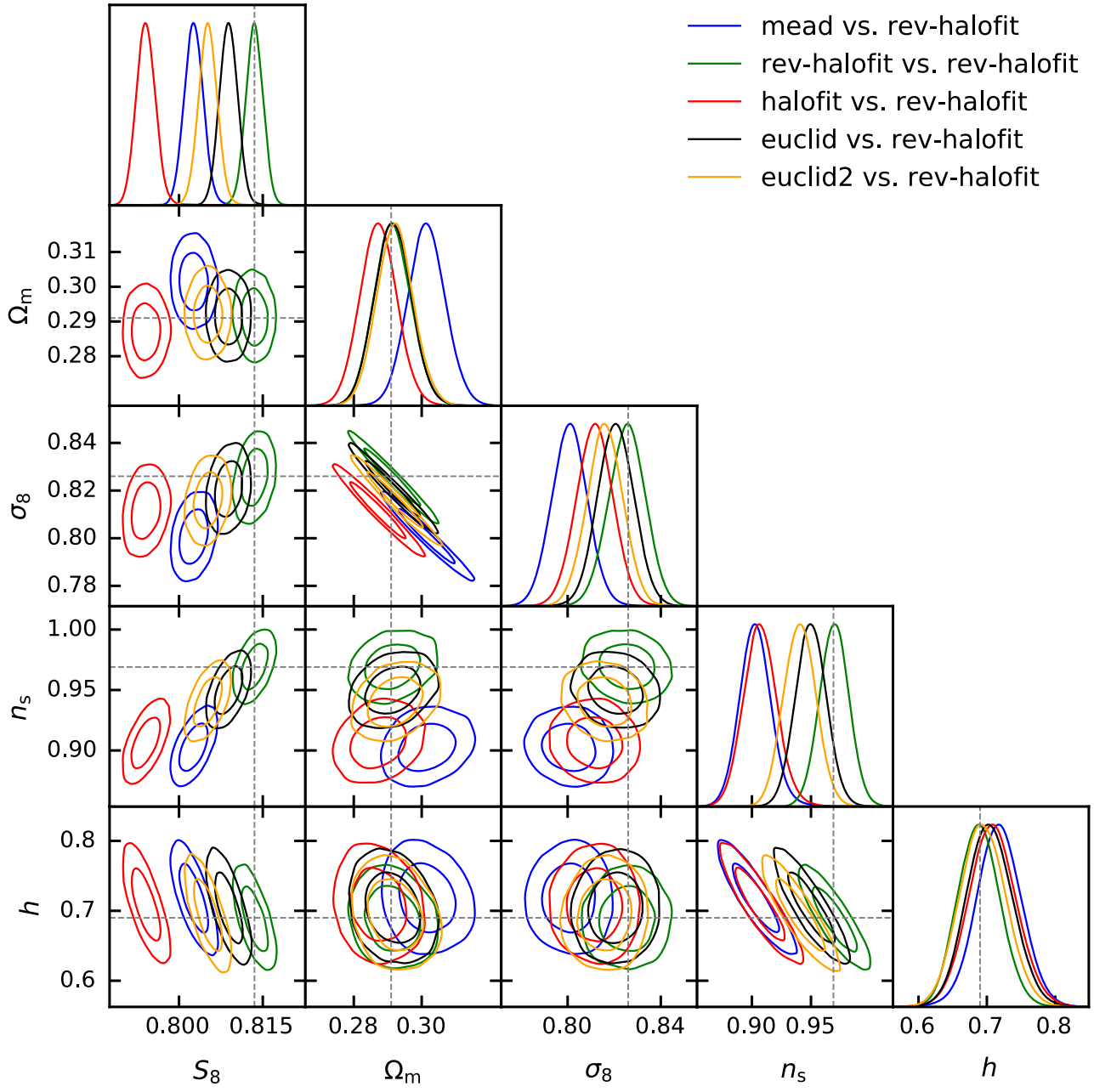

Figure 6. Cosmological parameter constraints of the stage IV survey in the ΛCDM model. Only two emulators, i.e. `EuclidEmulator` and `EuclidEmulator2`, are chosen for $C_{\ell,\mathrm{truth}}$, as `CosmicEmulator` does not provide a sufficient redshift range for the stage IV survey.

Eifler & Blazek (2016) shows that the constraints for Ω_m and σ_8 could be broadened when considering different systematics by: ~ 40 per cent (pessimistic LSST photo-z errors), ~ 50 per cent (optimistic LSST photo-z errors & non-linear intrinsic alignment (IA NLA) model), and ~ 100 per cent (pessimistic LSST photo-z errors & IA NLA model). In this case, the significance of the discrepancies between different predictors will be reduced by 25 per cent – 50 per cent, while still significant with the smallest between `mead` and `EuclidEmulator` larger than 0.6σ. In practice, the inclusion of all these systematics, as well as theoretical uncertainties will be needed to estimate the total error budget of specific weak lensing measurements.

5 CONCLUSIONS

The different halo-model based fitting functions and emulators have been widely used for the prediction of non-linear power spectrum to study the large-scale structure of the Universe. It is essential to understand their advantages, limitations, and theoretical uncertainties for different surveys and cosmologies. From our results, we conclude that:

(i) Compared with PKDGRAV3 simulations, the halo-model based fitting functions, except `halofit`, yield a $5-10$ per cent level accuracy for the matter power spectrum $P(k)$ for $k < 9\,h\,\mathrm{Mpc}^{-1}$ and $z < 2$, while emulators show better precision at the 2 per cent level.

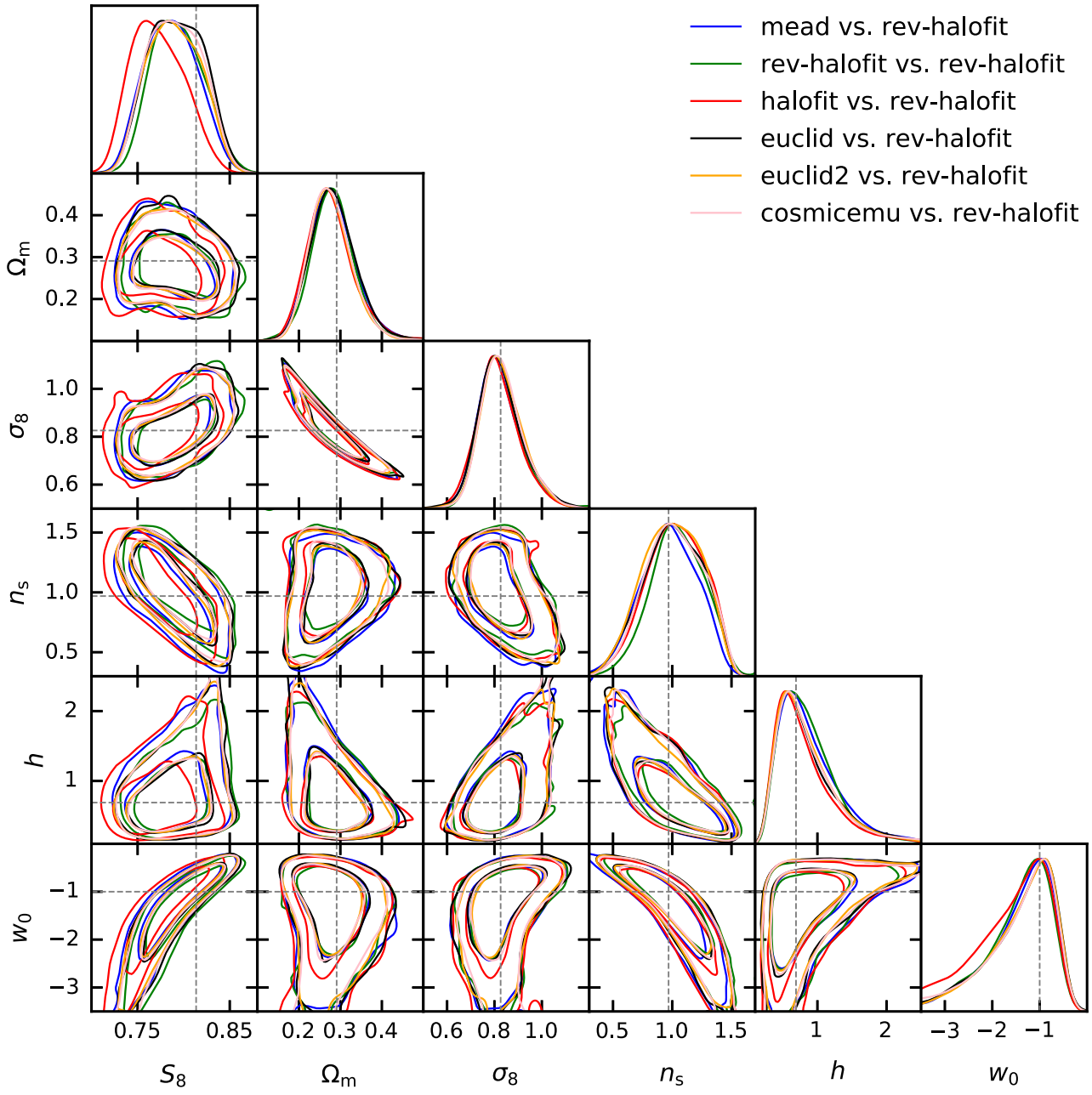

Figure 7. Cosmological parameter constraints of the stage III survey in the wCDM cosmological model. Including w_0 reduces significantly the constraining power, yielding much broader contours than the ΛCDM model.

For the weak lensing shear power spectrum C_ℓ, all the predictors, except for `halofit`, show a 5 per cent level mutual agreement.

(ii) For the stage III survey with a ΛCDM cosmology, the agreement on S_8 between different predictors are within 0.6σ, and within 0.2σ for other cosmological parameters (0.3σ and 0.2σ if we exclude `halofit`, respectively). This indicates the applicability of the studied predictors for the stage III surveys.

(iii) For the stage IV survey using a ΛCDM cosmology, the disagreements on S_8 are increased to several σs, with the largest discrepancy of 6.1σ between `rev-halofit` and `halofit`, and the best agreement between `mead` and `EuclidEmulator2`.

(iv) If w_0 is taken into account for the wCDM cosmology, we get weaker constraints on S_8, and the discrepancies between different predictors are reduced to $0.2-0.3\sigma$ and $0.7-2.4\sigma$ for the stage III and the stage IV surveys, respectively ($0.18-0.26\sigma$ and $0.7-1.7\sigma$ if we exclude `halofit`, respectively). If w_a is taken into account, we get very similar constraints on S_8 compared to the w_0-only case.

(v) The accuracy of the current fitting function models and emulators therefore appear sufficient for stage III surveys. However, for the future IV surveys, our results suggest that the fitting function models are currently not sufficiently accurate, and would need further improvements in the future. For emulators, it is required to explore wider ranges of cosmological parameters, k-modes, and redshifts, while pursuing consistent precision with reliable hydrodynamic N-body simulations.

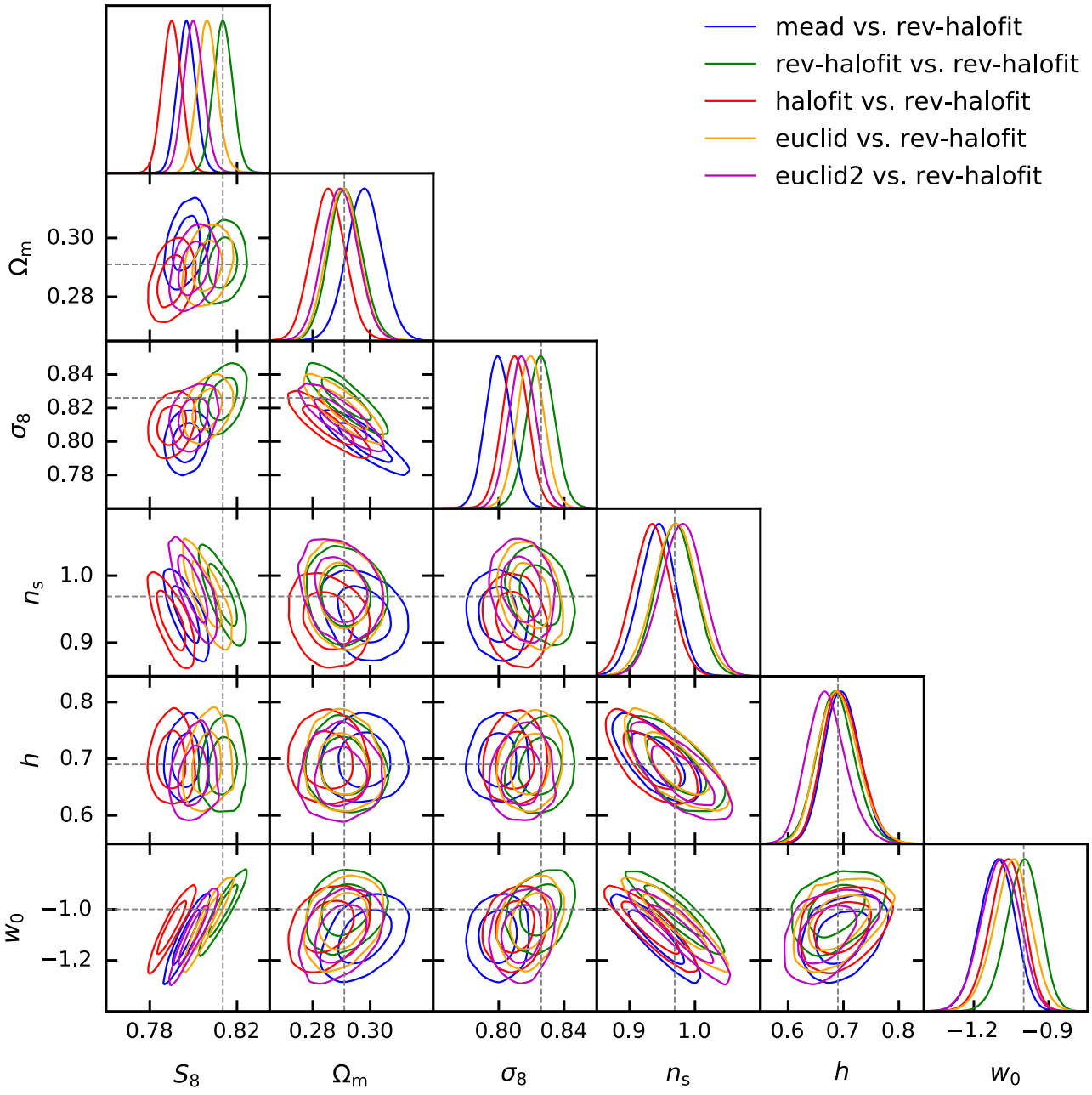

Figure 8. Cosmological parameter constraints of the stage IV survey in the wCDM cosmological model. The discrepancies between the predictors are alleviated, taking into account a simple wCDM cosmological model with a varying w_0.

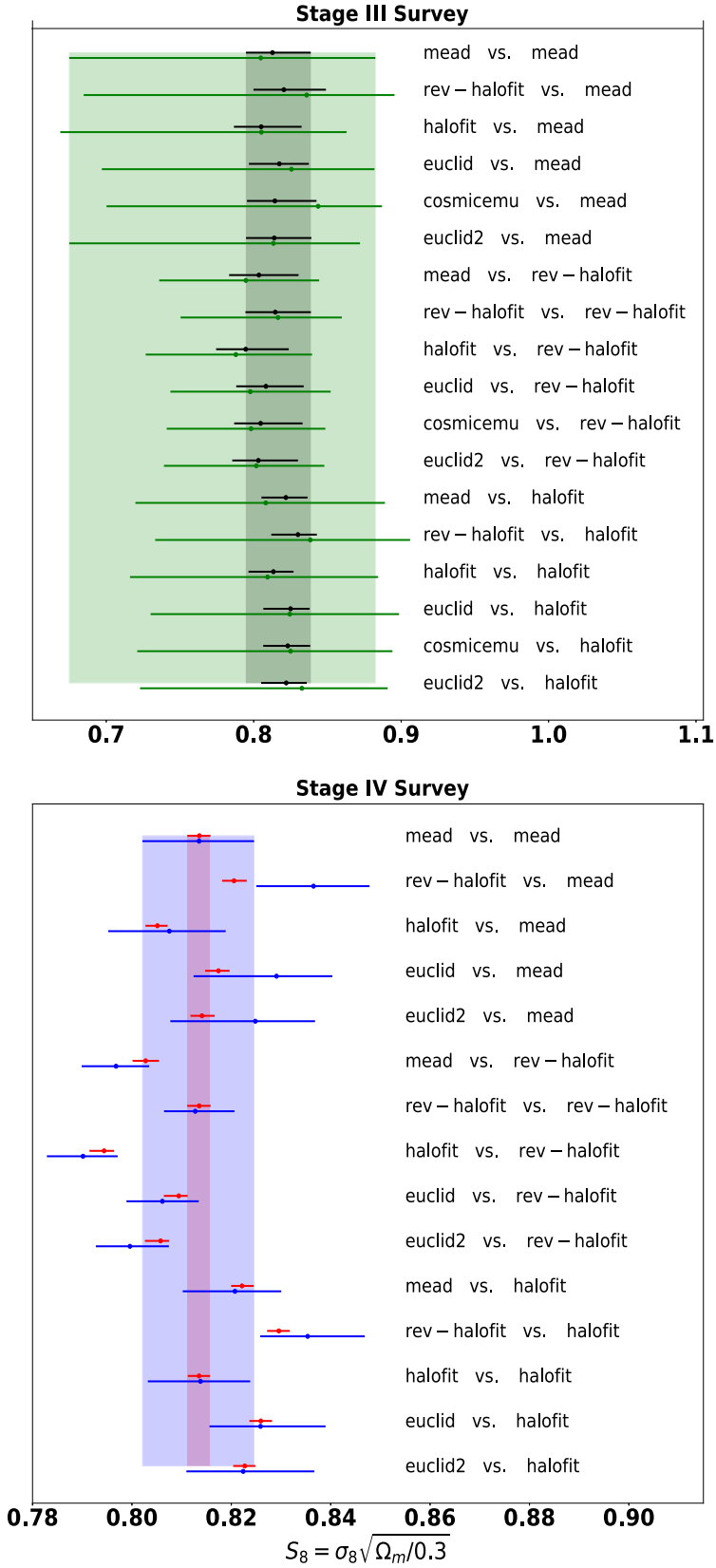

Figure 9. Deviations of the parameter constraints on S_8. The upper plot shows the result for the stage III survey, for the ΛCDM model (black) and the wCDM model (green), respectively. The lower plot shows the stage IV survey, for the ΛCDM (red) and wCDM (blue), respectively.

(vi) Taking into account other systematic effects such as baryonic effects, photometric redshift uncertainty, shear bias, and galaxy intrinsic alignment will broaden the parameters constraints by 40 per cent − 100 per cent from stage IV weak lensing surveys. This will tend to reduce the significance of the discrepancies between the different predictor. The theoretical uncertainties however remain non-negligible and need to be included in the total error budget for future surveys.

ACKNOWLEDGEMENTS

This work was supported in part by grant 200021_192243 from the Swiss National Science Foundation.

We thank Mischa Knabenhans from University of Zürich for the distribution of PKDGRAV3. We further thank Aurel Schneider from the University of Zürich for the useful discussions regarding this project and the covariance matrix for a stage IV survey. We would also like to thank Uwe Schmitt from ETH Zürich for his support with the GitLab server and development of `PyCosmo`.

The Collaborating Institutions are the Eidgenössische Technische Hochschule (ETH) Zürich, Ecole Polytechnique, the Laboratoire de Physique Nucléaire et des Hautes Energies of Sorbonne University.

DATA AVAILABILITY

Most of the analysis in this work is down on the Euler cluster[5] operated by ETH Zurich. Here follows the computational codes used in this study: `PyCosmo` (Refregier et al. 2018; Tarsitano et al. 2021; Moser et al. 2022) is used as the main tool where all the non-linear codes are implemented for the computation of auto (cross) power spectra, galaxy redshift distribution counts, and observable of cosmic shear. It is also extended to include interfaces with the emulators. `Anafast` is used for computation of power spectra from simulations, and all the the maps (masks, weight, shear, and mass) in pipeline are in `HealPix` format. We use `Emcee-3.0.2` (Foreman-Mackey et al. 2013) for the sampling of parameter space and `Getdist` (Lewis 2019) for the plotting of likelihood contours and `Uhammer` for the simplification of `Emcee` running. Some of the results in this paper have been derived using the `healpy` and `HEALPix packages` (Gorski et al. 1999). In this study, we made use of the functionalities provided by `numpy` (van der Walt, Colbert & Varoquaux 2011), `scipy` (Virtanen et al. 2020), and `matplotlib` (Hunter 2007).

REFERENCES

Abell P. A. et al., 2009, preprint (arXiv:0912.0201)
Aghanim N. et al., 2020, A&A, 641, A6
Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Amara A., Réfrégier A., 2007, MNRAS, 381, 1018
Amon A. et al., 2022, Phys. Rev. D, 105, 023514
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Angulo R. E., Zennaro M., Contreras S., Aricš G., Pellejero-Ibaśz M., Stścker J., 2020, MNRAS, 507, 5869
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, MNRAS, 506, 4070
Bartelmann M., Maturi M., 2016, Scholarpedia, 12, 32440
Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291
Baumann D., Nicolis A., Senatore L., Zaldarriaga M., 2012, J. Cosmol. Astropart. Phys., 2012, 051
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bernstein G. M., 2010, MNRAS, 406, 2793
Bernstein G., Jarvis M., 2002, AJ, 123, 583
Beutler F. et al., 2017, MNRAS, 464, 3409
Bird S., Viel M., Haehnelt M. G., 2012, MNRAS, 420, 2551
Blas D., Garny M., Konstandin T., 2014, J. Cosmol. Astropart. Phys., 2014, 010
Blas D., Garny M., Ivanov M. M., Sibiryakov S., 2016, J. Cosmol. Astropart. Phys., 2016, 052
Bridle S., King L., 2007, New J. Phys., 9, 444
Casarini L., Bonometto S. A., Borgani S., Dolag K., Murante G., Mezzetti M., Tornatore L., La Vacca G., 2012, A&A, 542, A126
Castro T., Quartin M., Giocoli C., Borgani S., Dolag K., 2018, MNRAS, 478, 1305
Cataneo M., Lombriser L., Heymans C., Mead A., Barreira A., Bose S., Li B., 2019, MNRAS, 488, 2121
Chevallier M., Polarski D., 2001, Int. J. Mod. Phys., 10, 213
Choi A. et al., 2016, MNRAS, 463, 3737
Chudaykin A., Ivanov M. M., Philcox O. H., Simonović M., 2020, Phys. Rev. D, 102, 063533
Collaboration E. et al., 2020, MNRAS, 505, 2840
Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519
Crocce M., Scoccimarro R., Bernardeau F., 2012, MNRAS, 427, 2537
d'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-MaÃrn H., 2020, J. Cosmol. Astropart. Phys., 2020, 005
D'Amico G., Senatore L., Zhang P., 2021, J. Cosmol. Astropart. Phys., 2021, 006
Dominik Z. et al., 2022, MNRAS, 511, 2075
Debackere S. N., Schaye J., Hoekstra H., 2020, MNRAS, 492, 2285
Eisenstein D. J., Hu W., 1999, ApJ, 511, 5
Fluri J., Kacprzak T., Lucchi A., Refregier A., Amara A., Hofmann T., Schneider A., 2019, Phys. Rev. D, 100, 063514
Foreman S., Senatore L., 2016, J. Cosmol. Astropart. Phys., 2016, 033
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Giannantonio T., Porciani C., Carron J., Amara A., Pillepich A., 2012, MNRAS, 422, 2854
Giblin B., Cataneo M., Moews B., Heymans C., 2019, MNRAS, 490, 4826
Giblin B. et al., 2021, A&A, 645, A105
Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, The HEALPix Primer, preprint (arXiv:astro-ph/9905275)
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hamilton A. J. S., Kumar P., Lu E., Matthews A., 1991, ApJ, 374, L1
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heavens A., Refregier A., Heymans C., 2000, MNRAS, 319, 649
Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, ApJ, 705, 156
Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104
Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2013, ApJ, 780, 111
Heitmann K. et al., 2016, ApJ, 820, 108
Hildebrandt H. et al., 2020, A&A, 633, A69
Hinshaw G. et al., 2013, ApJS, 208, 19
Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526
Hirata C. M. et al., 2004, MNRAS, 353, 529
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101
Jing Y., Zhang P., Lin W., Gao L., Springel V., 2006, ApJ, 640, L119
Joachimi B., Mandelbaum R., Abdalla F., Bridle S., 2011, A&A, 527, A26

[5] https://scicomp.ethz.ch/wiki/Euler.

Kaiser N., 1992, ApJ, 388, 272
Kaiser N., 1998, ApJ, 498, 26
Kaiser N., Squires G., 1993, ApJ, 404, 441
Kilbinger M. et al., 2017, MNRAS, 472, 2126
Kitching T. D., Alsing J., Heavens A. F., Jimenez R., McEwen J. D., Verde L., 2017, MNRAS, 469, 2737
Knabenhans M. et al., 2019, MNRAS, 484, 5509–5529
Knabenhans M., Brinckmann T., Stadel J., Schneider A., Teyssier R., 2021, MNRAS, 518, 1859
Kodwani D., Alonso D., Ferreira P., 2018, preprint (arXiv:1811.11584)
Krause E., Eifler T., Blazek J., 2016, MNRAS, 456, 207
Lawrence E. et al., 2017, ApJ, 847, 50
Lesgourgues J., 2011, The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparision with CAMB for LambdaCDM(arXiv:1104.2934)
Lewis A., 2019, GetDist: a Python package for analysing Monte Carlo samples, preprint (arXiv:1910.13970)
Limber D. N., 1953, ApJ, 117, 134
Linder E. V., 2003, Phys. Rev. Lett., 90, 091301
LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506
Ma C.-P., Fry J. N., 2000, ApJ, 543, 503
Mancini A. S., Piras D., Alsing J., Joachimi B., Hobson M. P., 2022, MNRAS, 511, 1771
Martinelli M. et al., 2021, A&A, 649, A100
Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, MNRAS, 454, 1958
Mead A. J., Heymans C., Lombriser L., Peacock J. A., Steele O. I., Winther H. A., 2016, MNRAS, 459, 1468
Melchior P., Viola M., 2012, MNRAS, 424, 2757
Mohammed I., Seljak U., 2014, MNRAS, 445, 3382
Moser B., Lorenz C., Schmitt U., Réfrégier A., Fluri J., Sgier R., Tarsitano F., Heisenberg L., 2022, Astron. Comput., 40, 100603
Nishimichi T., Bernardeau F., Taruya A., 2016, Phys. Lett. B, 762, 247
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P., 2020, Large-Scale Structure of the Universe by Phillip James Edwin Peebles. Princeton University Press
Percival W. J. et al., 2014, MNRAS, 439, 2531
Petri A., Haiman Z., May M., 2017, Phys. Rev. D, 95, 123503
Potter D., Stadel J., Teyssier R., 2017, Comput. Astrophy. Cosmol., 4, 2
Press W. H., Schechter P., 1974, ApJ, 187, 425
Refregier A., Kacprzak T., Amara A., Bridle S., Rowe B., 2012, MNRAS, 425, 1951
Refregier A., Gamper L., Amara A., Heisenberg L., 2018, Astronomy and computing, 25, 38
Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, ApJ, 672, 19
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 047
Scoccimarro R., Sheth R. K., Hui L., Jain B., 2001, ApJ, 546, 20
Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515
Seljak U., 2000, MNRAS, 318, 203
Seljak U. c. v., Vlah Z., 2015, Phys. Rev. D, 91, 123516
Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, MNRAS, 417, 2020
Sgier R. J., Réfrégier A., Amara A., Nicola A., 2019, J. Cosmol. Astropart. Phys., 2019, 044
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Smail I., Hogg D. W., Yan L., Cohen J. G., 1995, ApJ, 449, L105
Smith R. E. et al., 2003, MNRAS, 341, 1311
Springel V., 2005a, MNRAS, 364, 1105
Springel V., 2005b, MNRAS, 364, 1105
Springel V., Yoshida N., White S. D., 2001, New Astron., 6, 79
Springel V., Pakmor R., Zier O., Reinecke M., 2021, MNRAS, 506, 2871
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Tarsitano F. et al., 2021, Astronomy and Computing, 36, 100484
Tram T., Brandbyge J., Dakin J., Hannestad S., 2019, J. Cosmol. Astropart. Phys., 2019, 022
Troxel M. A. et al., 2018, Phys. Rev. D, 98, 043528
van Daalen M. P., Schaye J., Booth C., Dalla Vecchia C., 2011, MNRAS, 415, 3649
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wallis C. G., McEwen J. D., Kitching T. D., Leistedt B., Plouviez A., 2022, MNRAS, 509, 4480
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Zonca A., Singer L. P., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K. M., 2019, J. Open Source Softw., 4, 1298
Zürcher D., Fluri J., Sgier R., Kacprzak T., Refregier A., 2021, J. Cosmol. Astropart. Phys., 2021, 028

APPENDIX A: POWER SPECTRUM COMPARISON

In this section, we present the comparison of the non-linear power spectrum for all redshifts, as shown in Fig. A1.

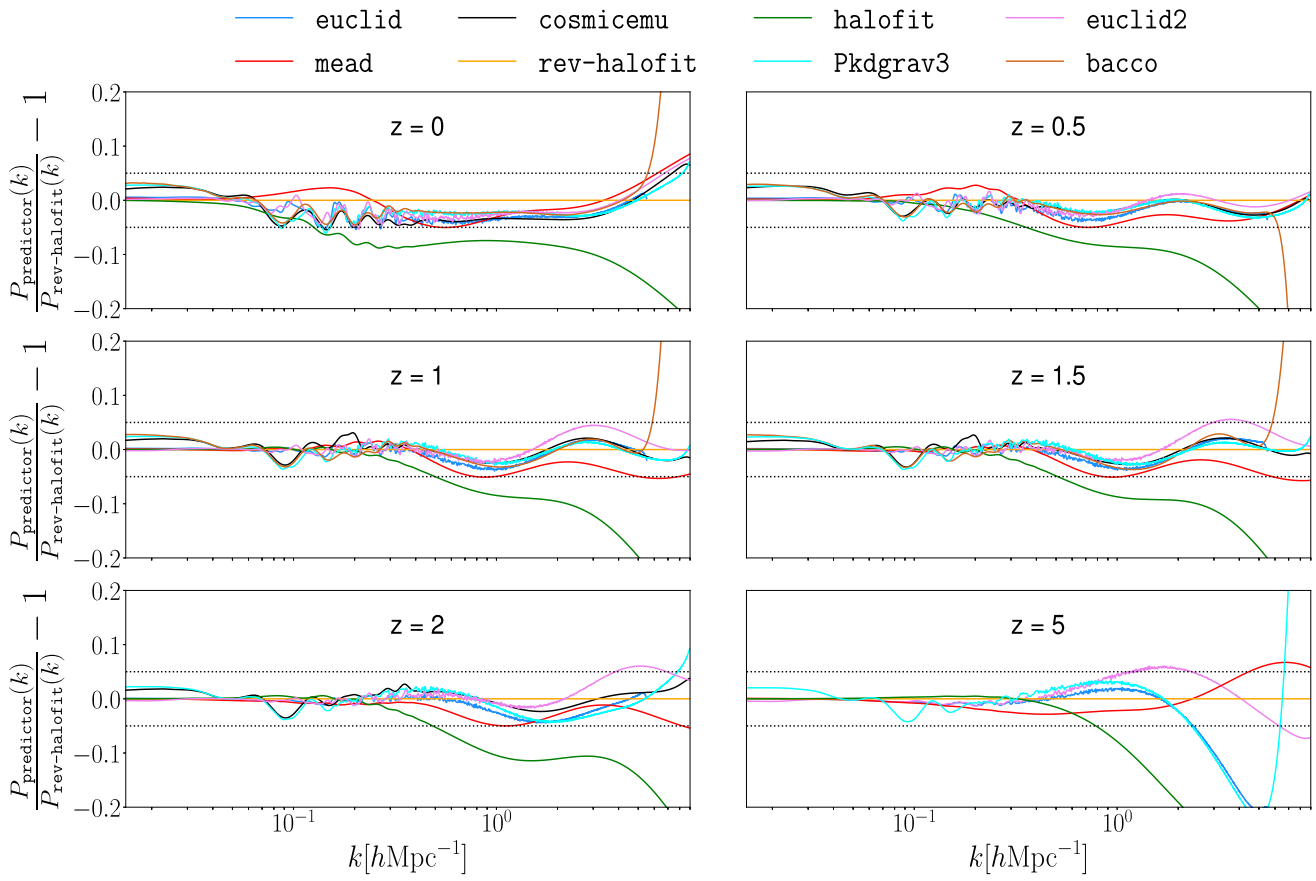

Figure A1. The comparison of the dark-matter-only non-linear $P(k)$ of different predictors at different redshifts ($z = 0$, 0.5, 1, 1.5, 2, and 5), subtracted and divided by `rev-halofit` as reference. `BaccoEmulator` and `CosmicEmulator` are not valid for $z > 3$, so we do not take them into comparison for $z = 5$.

APPENDIX B: COSMOLOGICAL PARAMETER CONSTRAINTS

B1 Summary of constraints

The summary of constraints on $\{S_8, \Omega_{\rm m}, n_{\rm s}, h, w_0\}$ is concluded in this section, shown in Table B1.

B2 Different $\ell_{\rm max}$

We investigate the variation of constraints on S_8 with different $\ell_{\rm max}$ (800 or 1000) and the results are summarized in Fig. B1. We only consider the stage IV survey with the Λ`CDM` model, since it gives the largest discrepancies between different predictors. When $\ell_{\rm max}$ is reduced from 1000 to 800 where we have less non-linear effect information, the marginalized 1D constraints on S_8 are broadened by 5 per cent. With this change, `rev-halofit` shows better agreements with `mead`, and larger discrepancies with `EuclidEmulator` and `EuclidEmulator2`.

B3 w`CDM` with free w_a

We present in Fig. B2 the cosmological parameter constraints of the stage IV survey in the `wCDM` cosmological model, with both varying w_0 and w_a. Compared with the case with a fixed w_a, this setting gives a tiny impact on the discrepancies between different predictors for $\{S_8, n_{\rm s}, h\}$. However, it obtains weaker constraints on $\{\Omega_{\rm m}, w_0\}$, resulting in the good agreements between different predictors. The discrepancies on w_a are within 0.5σ.

Table B1. Complete numerical constraints on the cosmological parameters corresponding to the contours in Figs 5, 6, 7, and 8. For each predictor, the σs show the theoretical discrepancies for each parameter, compared to the reference one.

Survey cosmology	Predictor ref: `rev-halofit`	S_8	(σ)	Ω_m	(σ)	n_s	(σ)	h	(σ)	w_0	(σ)
Stage III	`rev-halofit`	$0.8147^{+0.0241}_{-0.0203}$		$0.288^{+0.0817}_{-0.0662}$		$0.9741^{+0.2489}_{-0.1475}$		$0.6736^{+0.5642}_{-0.4172}$			
	`mead`	$0.8035^{+0.0269}_{-0.0202}$	0.33	$0.2996^{+0.0848}_{-0.0698}$	0.11	$0.9144^{+0.2425}_{-0.1562}$	0.21	$0.6859^{+0.5779}_{-0.4358}$	0.02		
ΛCDM	`halofit`	$0.7946^{+0.0292}_{-0.0201}$	0.57	$0.2884^{+0.0783}_{-0.074}$	0.0	$0.9196^{+0.2566}_{-0.1662}$	0.18	$0.6777^{+0.6198}_{-0.4339}$	0.01		
	`euclid`	$0.8083^{+0.0256}_{-0.0201}$	0.2	$0.2987^{+0.0831}_{-0.0709}$	0.1	$0.9547^{+0.2428}_{-0.1562}$	0.07	$0.6644^{+0.5612}_{-0.4159}$	0.01		
	`cosmicemu`	$0.8047^{+0.0285}_{-0.018}$	0.29	$0.2916^{+0.0789}_{-0.0741}$	0.03	$0.9332^{+0.2613}_{-0.1399}$	0.14	$0.7452^{+0.5542}_{-0.486}$	0.1		
	`euclid2`	$0.8031^{+0.0269}_{-0.0177}$	0.34	$0.2887^{+0.0835}_{-0.0679}$	0.01	$0.9184^{+0.2496}_{-0.1316}$	0.19	$0.7467^{+0.5503}_{-0.4853}$	0.1		
Stage III	`rev-halofit`	$0.8165^{+0.0433}_{-0.0661}$		$0.2846^{+0.092}_{-0.09}$		$0.9164^{+0.5799}_{-0.3511}$		$0.7868^{+0.9823}_{-0.5347}$		$-0.9242^{+0.4704}_{-2.294}$	
	`mead`	$0.7947^{+0.0497}_{-0.0588}$	0.26	$0.31^{+0.0824}_{-0.1022}$	0.18	$0.9768^{+0.5012}_{-0.4514}$	0.08	$0.6527^{+1.0402}_{-0.4233}$	0.11	$-1.139^{+0.647}_{-2.2626}$	0.09
wCDM	`halofit`	$0.7879^{+0.0517}_{-0.0612}$	0.34	$0.2968^{+0.0787}_{-0.1011}$	0.09	$0.9919^{+0.4913}_{-0.4914}$	0.1	$0.6192^{+1.1863}_{-0.3779}$	0.13	$-1.1333^{+0.6581}_{-2.3122}$	0.09
	`euclid`	$0.7977^{+0.0545}_{-0.0542}$	0.22	$0.3049^{+0.085}_{-0.1017}$	0.15	$1.032^{+0.4723}_{-0.4813}$	0.15	$0.6209^{+1.1393}_{-0.3873}$	0.13	$-1.1886^{+0.7187}_{-2.1508}$	0.11
	`cosmicemu`	$0.7982^{+0.0504}_{-0.0572}$	0.22	$0.2931^{+0.0921}_{-0.0969}$	0.06	$1.0031^{+0.5093}_{-0.4918}$	0.11	$0.6688^{+1.2915}_{-0.4301}$	0.08	$-1.1408^{+0.6926}_{-2.3046}$	0.09
	`euclid2`	$0.8018^{+0.0461}_{-0.0627}$	0.18	$0.2896^{+0.0928}_{-0.0877}$	0.04	$0.9272^{+0.5729}_{-0.3921}$	0.02	$0.7461^{+1.0126}_{-0.5134}$	0.04	$-1.0254^{+0.5498}_{-2.2745}$	0.04
Stage IV	`rev-halofit`	$0.8135^{+0.0023}_{-0.0024}$		$0.2915^{+0.0077}_{-0.0084}$		$0.9696^{+0.0178}_{-0.0192}$		$0.6889^{+0.0481}_{-0.0433}$			
	`mead`	$0.8028^{+0.0027}_{-0.0026}$	2.96	$0.3008^{+0.0094}_{-0.0074}$	0.87	$0.9021^{+0.0193}_{-0.0189}$	2.48	$0.7181^{+0.0495}_{-0.0441}$	0.45		
ΛCDM	`halofit`	$0.7944^{+0.002}_{-0.0029}$	6.11	$0.2856^{+0.0097}_{-0.0064}$	0.46	$0.9054^{+0.0203}_{-0.0197}$	2.3	$0.7134^{+0.0494}_{-0.05}$	0.35		
	`euclid`	$0.8094^{+0.0018}_{-0.003}$	1.37	$0.2917^{+0.0079}_{-0.0084}$	0.02	$0.9497^{+0.0198}_{-0.0193}$	0.72	$0.7058^{+0.0505}_{-0.0479}$	0.25		
	`euclid2`	$0.8058^{+0.0017}_{-0.0032}$	2.62	$0.2926^{+0.0079}_{-0.0084}$	0.1	$0.9402^{+0.0195}_{-0.0206}$	1.07	$0.6958^{+0.0519}_{-0.0475}$	0.1		
Stage IV	`rev-halofit`	$0.8127^{+0.0079}_{-0.0063}$		$0.2909^{+0.0095}_{-0.0086}$		$0.9741^{+0.045}_{-0.0555}$		$0.6884^{+0.0599}_{-0.052}$		$-1.0127^{+0.1171}_{-0.1046}$	
	`mead`	$0.7968^{+0.0067}_{-0.0069}$	1.73	$0.2979^{+0.0106}_{-0.0092}$	0.53	$0.9426^{+0.0431}_{-0.0466}$	0.45	$0.698^{+0.0563}_{-0.0449}$	0.13	$-1.106^{+0.1107}_{-0.1163}$	0.61
wCDM	`halofit`	$0.7902^{+0.007}_{-0.0073}$	2.39	$0.2856^{+0.0093}_{-0.0096}$	0.42	$0.9306^{+0.047}_{-0.0479}$	0.6	$0.6986^{+0.0606}_{-0.0507}$	0.13	$-1.0646^{+0.1069}_{-0.1197}$	0.35
	`euclid`	$0.8061^{+0.0073}_{-0.0072}$	0.68	$0.2908^{+0.0088}_{-0.0094}$	0.01	$0.9671^{+0.0574}_{-0.053}$	0.09	$0.6968^{+0.0609}_{-0.0604}$	0.1	$-1.046^{+0.1142}_{-0.1288}$	0.22
	`euclid2`	$0.7996^{+0.0078}_{-0.0069}$	1.31	$0.2901^{+0.0099}_{-0.0094}$	0.06	$0.9791^{+0.0515}_{-0.0588}$	0.07	$0.6711^{+0.0657}_{-0.0548}$	0.21	$-1.0965^{+0.1288}_{-0.1255}$	0.51

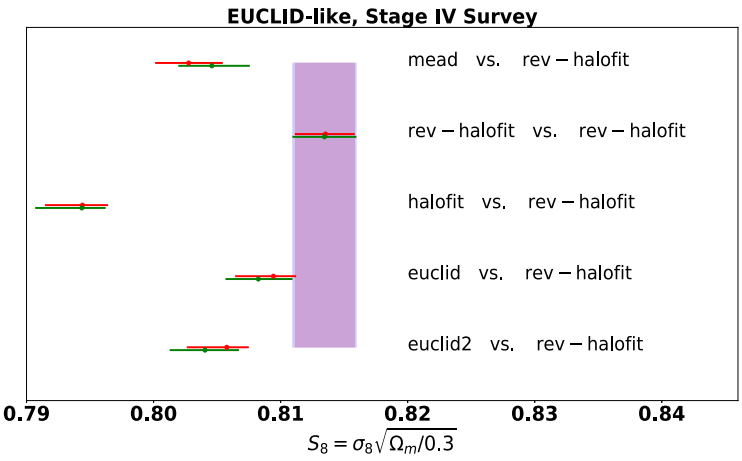

Figure B1. Deviations of the parameter constraints on S_8, for the stage IV survey, with the ΛCDM model, and different $\ell_{\rm max}$ [$\ell_{\rm max} = 1000$ (red) and $\ell_{\rm max} = 800$ (green)].

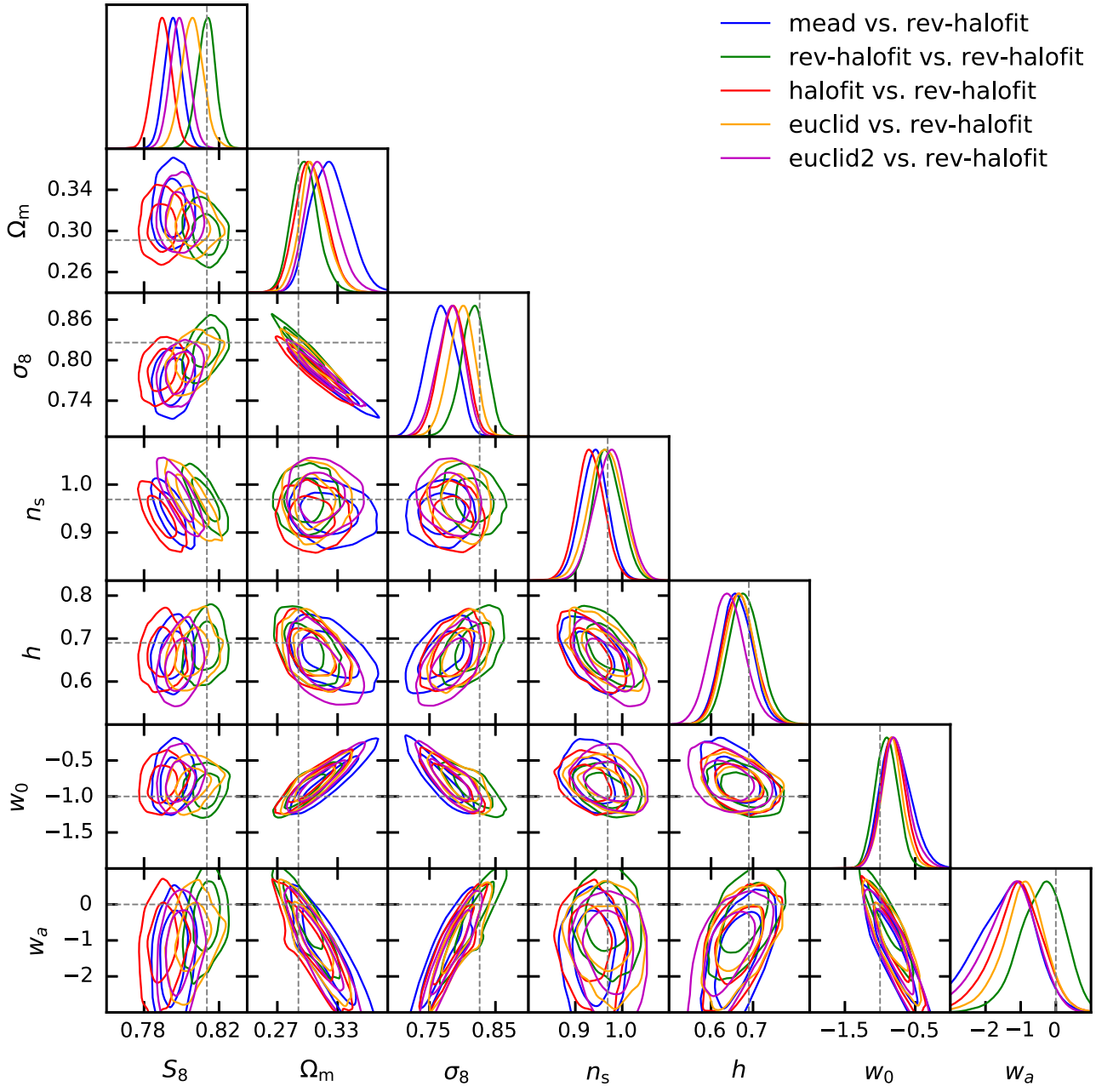

Figure B2. Cosmological parameter constraints of the stage IV survey in the wCDM cosmological model. The discrepancies between the predictors are alleviated, taking into account a simple wCDM cosmological model with both varying w_0 and w_a.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Résumé

Les oscillations acoustiques baryoniques (BAO) sont une sonde puissante permettant de mesurer l'expansion accélérée de l'univers et de fournir des contraintes sur les modèles d'énergie noire. Il peut être mesuré à l'aide de la fonction de corrélation à deux points des traceurs de matière, et le but de cette thèse est de mesurer le BAO à des redshifts $z > 2,1$ élevés en utilisant les forêts Lyman-α (Lyα). Cette thèse utilise des données d'observation spectroscopiques et des catalogues simulés (simulations) de deux grandes enquêtes cosmologiques, eBOSS (DR16) et DESI (EDR). Je présente l'analyse comparative Lyα de ces deux enquêtes et je les trouve cohérentes en termes de qualité et d'ajustement des données. J'ai étudié à la fois sur des simulations et sur des données, l'un des effets systématiques les plus importants de l'analyse Lyα, la présence de systèmes à haute densité de colonnes (HCD). J'ai proposé un modèle empirique et développé un modèle analytique, le modèle Voigt, pour caractériser leur impact sur les fonctions de corrélation Lyα. Le modèle Voigt est bien vérifié sur des simulations et fournit une mesure physique des paramètres de biais et RSD des HCD, ainsi qu'une bonne contrainte sur les paramètres Lyα.

Mots clés : cosmology, large-scale structure, bao

Résumé

The Baryon Acoustic Oscillations (BAO) is a powerful probe to measure the accelerated expansion of the universe and provide constraints on dark energy models. It can be measured using the two-point correlation function of matter tracers, and the goal of this thesis is to measure the BAO at high redshifts $z > 2.1$ using Lyman-α (Lyα) forests. This thesis makes use of spectroscopic observation data and simulated catalogs (mocks) from two large cosmological surveys, eBOSS (DR16) and DESI (EDR). I present the comparison Lyα analysis of these two surveys and found them consistent in terms of data quality and fits. I studied on both mocks and data, one of the most important systematic effects of Lyα analysis, the presence of High Column Density Systems (HCDs). I proposed an empirical model and further developed an analytical model, the Voigt model, to characterize their impact on Lyα correlation functions. The Voigt model is well verified on mocks and provides a physical measurement of the bias and RSD parameters of HCDs, and a good constraint on the Lyα parameters.

Mots clés : cosmology, large-scale structure, bao

Laboratoire de physique nucléaire et des hautes énergies
Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France