UAB
Universitat Autònoma de Barcelona

Measuring cosmic expansion and growth with the Lyman – α forest in DESI

Calum Gordon

UNIVERSITAT AUTÒNOMA DE BARCELONA
INSTITUT DE FÍSICA D'ALTES ENERGIES

Universitat Autònoma
de Barcelona

Measuring cosmic expansion and growth with the Lyman-α forest in DESI

Doctorate in Physics
Programa de Doctorat en Física

Calum Gordon

Thesis supervised by:

Andreu Font Ribera (Director) and Eduard Massó i Soler (Tutor)

A thesis submitted for the degree of Philosophae Doctor (PhD)

2025

ABSTRACT

Since the early 2000s, ground- and space-based cosmological surveys have produced increasingly detailed 3D maps of galaxies to study the energy content and expansion history of the Universe. Clustering measurements of galaxies in the late-time Universe reveal signatures of Baryon Acoustic Oscillations (BAO)—fluctuations in matter caused by sound waves propagating through the pre-recombination plasma. The maximum distance travelled by these waves at recombination defines the sound horizon (r_d), which manifests as a peak in the galaxy correlation function. By measuring this scale across redshifts, surveys such as the Dark Energy Spectroscopic Instrument (DESI) constrain the expansion history and probe the nature of dark energy (DESI Collaboration et al., 2025c).

Around the same time, the Lyman-α (Lyα) forest emerged as a tracer of the matter density field. These are a series of absorption lines in the spectra of quasars [1], formed when light from these objects at the Lyα wavelength ($\sim$ 1216 Å), is absorbed by neutral hydrogen in the Intergalactic Medium (IGM). Because quasars are relatively bright objects, the Lyα forest can be measured at relatively high redshift ($z \sim 2-4$), making BAO measurements from it extremely complementary to the lower redshift galaxy measurements. In chapter 5, I present the first clustering measurements in the Lyα forest from early DESI data, comparing them to earlier measurements from the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) - the previous benchmark. We also discuss any changes to analysis and data processing techniques that have improved upon eBOSS.

Beyond BAO, cosmological information resides in the full shape of the galaxy and Lyα correlation functions. These analyses can utilise Redshift-Space Distortions (RSD) — the warping of clustering in redshift space due to peculiar motions of tracers — or the Alcock-Paczyński (AP) effect, which measures anisotropies in clustering. The former can also be used to measure cosmic growth, as peculiar

---

[1] These are also visible in galaxies, but have primarily been observed up until now in quasars.

motions are a direct consequence of it. While full-shape measurements enhance our cosmological constraining power, they are also more sensitive to systematics. In chapter 6, I present work that addresses the effect of one such systematic - quasar redshift errors. Using synthetic datasets from the DESI survey, we create and apply a new model of this effect to our full-shape measurements, which we show to de-bias a key set of parameters used for cosmological inference.

After the DESI survey has finished, it will be succeeded by DESI-II. This will build on the DESI survey, by observing a larger number of galaxies and quasars, to even greater depth (in a partially overlapping footprint). This includes a high-redshift survey up to $z \sim 4$, which will increase the target density of quasars from 60 to 100 deg^{-2} (with respect to DESI), and observe two new tracer types: Lyman break galaxies (LBGs) and Lyman-α emitters (LAE). Both of these galaxy types will be observed to high density, $\sim 1000\,\text{deg}^{-2}$, and LBGs additionally contain visible Lyα forests which can be used for cosmology. In chapter 7, I detail a set of DESI-II forecasts, which emulate the conditions and target selection of the survey. We forecast a full set of BAO measurements from LBGs, LAEs and quasars, including cross-correlations between the tracers and their forests. Our cross-correlation of LBGs/LAEs with quasars Lyα forests for example, will take advantage of the high signal-to-noise of the latter, and the high target density of former.

CONTENTS

List of Figures

List of Tables

ACRONYMS

BAL	Broad Absorption Line *(p. 37)*
BBN	Big Bang Nucleosynthesis *(p. 10)*
BOSS	Baryon Oscillation Spectroscopic Survey *(p. 32)*
DESI	Dark Energy Spectroscopic Instrument *(p. 27)*
DLA	Damped Lyman-α Absorption *(p. 37)*
eBOSS	Extended Baryon Oscillation Spectroscopic Survey *(p. 32)*
EDR	Early Data Release *(p. 33)*
EDR+M2	Early Data Release plus the first two months of main survey *(p. 34)*
HCD	High Column Density Systems *(p. 48)*
IGM	Intergalactic Medium *(p. 24)*
Lyα	Lyman-α *(pp. 106, 108)*
RSD	Redshift-Space Distortions *(p. 22)*
SDSS	Sloan Digital Sky Survey *(p. 32)*

1

THE COSMOLOGICAL BACKGROUND

Cosmology is the study of the Universe: its origins, its expansion history and its evolution. Modern cosmology is built upon the principle that matter is homogeneously and isotropically distributed at large enough scales. The theory of General Relativity (GR; Einstein, 1915), developed by Einstein in the early 20th century, is a framework used to describe the behaviour of matter and radiation at large distances, linking the underlying geometry of the Universe to the energy densities of its constituent parts.

In this chapter, we will introduce the tools necessary to understand the cosmological background. These include the metric, which gives invariant distances between two points in a curved spacetime, and Einstein's field equations, which connect spacetime with the energy and momentum contained within it. We will then go on to derive a set of equations governing an expanding Universe, under a perfect fluid assumption. Finally, we will discuss distance measurements in an expanding Universe, as well as the concept of redshift. The formation of the complex structure that we observe today, caused by quantum perturbations in the early Universe, will be explored in the chapter following this one.

1.1 Spacetime

The theory of GR is founded upon several key postulates. The first is the Weak Equivalence Principle (WEP), the statement that particles accelerate at the same rate under the influence of a gravitational field, or that inertial and gravitational mass are the same. This was first experimentally observed by Galileo, and was later incorporated into Newton's theory of gravitation. Expanding upon this, Einstein noted that the laws of physics are the same whether observed in a reference frame in a constant gravitational field, or an accelerating reference frame. This expansion of WEP (known as the Einstein Equivalence Principle, or EEP) is consistent with

Special Relativity, and implies both local Lorentz invariance and local positional invariance. Noting then that the effect of gravity can be eliminated by switching to a free falling frame, Einstein had a famous realisation: that gravity was not a force acting within spacetime, but rather a consequence of spacetime itself.

To encapsulate this idea mathematically, we introduce the metric. The metric is a tensor that relates coordinate distance to proper distance, for any spacetime. The metric also encodes the effect of gravity, and allows one to derive the motion of particles as a consequence of spatial curvature. The invariant interval ds^2 is defined using the metric, as:

$$ds^2 = g_{\mu\nu} dx^{\mu} dx^{\nu}. \tag{1.1}$$

Here $g_{\mu\nu}$ is the metric tensor, and dx is an infinitesimal interval for $\mu, \nu = \{0, 1, 2, 3\}$, where 0 is the time-like coordinate and 1 to 3 are spatial coordinates [1]. For any coordinate system in any spacetime, the metric g preserves ds^2. For example, Special Relativity is described by the Minkowski metric:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1, \end{pmatrix} \tag{1.2}$$

where we are using the "mostly positive" convention, with a negative time component and positive spatial components. We can then write the invariant interval as:

$$ds^2 = -cdt^2 + \sum_i dx_i^2 \tag{1.3}$$

If $ds^2 < 0$, the interval is said to be time-like, where two events are separated by more time than space, implying it is possible to find a frame where the two events occur in the same place. If $ds^2 > 0$, the interval is space-like, and cannot be traversed by an object moving at $v \leq c$, where c is the speed of light. It also represents the interval between two events that occur at the same time, but cannot be causally connected. Finally, $ds^2 = 0$ is a "light-like" interval, that can only be travelled by massless particles travelling at c.

---

[1] Note that we are using Einstein summation convention to implicitly sum over repeated indices.

1.1.1 Geodesics and Christoffel symbols

In Newtonian physics, a particle travels in a straight line unless acted upon by external forces, like gravity. In curved spaces, a straight line is generalised to the geodesic, the shortest path between two points. In GR, spacetime curvature is a result of gravity, so the geodesic is the shortest path between two points for any particle in the absence of forces other than gravity.

Mathematically, we want to generalise $d^2x/dt^2 = 0$ in Newtonian physics, to other more complex spacetimes and coordinate systems. To begin, we transform between coordinate systems x and x', by writing:

$$\frac{dx^i}{dt} = \frac{\delta x^i}{\delta x'^j}\frac{dx'^j}{dt}. \tag{1.4}$$

Taking the time derivative of both sides of this equation results in the geodesic equation:

$$\frac{d^2x^i}{d\lambda^2} + \Gamma^i_{jk}\frac{dx^j}{d\lambda}\frac{dx^k}{d\lambda} = 0, \tag{1.5}$$

where we have replaced time with λ, representing any choice of evolution. We have also introduced the Christoffel symbol, which can be written in terms of the metric as[2]:

$$\Gamma^i_{jk} = \frac{g^{il}}{2}\left[\frac{\partial g_{jl}}{\partial x^k} + \frac{\partial g_{kl}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^l}\right]. \tag{1.6}$$

Using only equations 1.5 and 1.6, we could calculate how a massless or massive particle moves in a flat expanding Universe described by the FLRW metric.

1.1.2 Einstein's field equations

The geometry of the Universe is connected to the energy-momentum tensor (section 1.3) through Einstein's field equations:

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R + \Lambda g^{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.7}$$

The terms on the left hand side are completely defined in terms of the metric, where $R^{\mu\nu}$ is the Ricci tensor and R is the Ricci scalar, a contraction of the Ricci tensor and the metric ($R = g_{\mu\nu}R^{\mu\nu}$). The Ricci tensor itself is defined purely by the metric and its derivates:

$$R^{\mu\nu} = \Gamma^\alpha_{\mu\nu,\alpha} - \Gamma^\alpha_{\mu\alpha,\nu} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\beta\nu}\Gamma^\beta_{\mu\alpha}, \tag{1.8}$$

---

[2] This can also be written in terms of coordinates, but it's usually less convenient.

recalling the definition of the Christoffel symbol (equation 1.6), and $\Gamma^{\alpha}_{\mu\nu,\alpha} = \partial\Gamma^{\alpha}_{\mu\nu}/\delta x^{\alpha}$. Note that the cosmological constant (Λ) term can also be included on the right-hand side, where it is effectively the energy-momentum tensor of the vacuum, with form:

$$T^{\mu\nu}_{(\mathrm{vac})} = -\frac{\Lambda}{8\pi G} g^{\mu\nu}. \tag{1.9}$$

1.2 An expanding Universe

In 1927, Georges Lemaître found a linear relationship between the distance of other galaxies, and their recessional velocity. This experimentally supported the notion of an expanding Universe, which had been a solution derived from Einstein's field equations by Alexander Friedmann in 1922 (see section 1.4). In 1929, Edwin Hubble confirmed the measurements of Lemaître, cementing what is now known as the Hubble-Lemaître law:

$$v = H_0 x. \tag{1.10}$$

The constant of proportionality between recession velocity v and distance x is known as Hubble's constant. Note that this equation is only strictly true for small recession velocities (i.e. nearby galaxies), since the expansion rate of the Universe evolves with time.

The relative size of the Universe at any time t is described by the scale factor $a(t)$, which relates the proper (physical) distance of an object x to its coordinate, or comoving distance χ:

$$x = a(t)\,\chi(t)\,. \tag{1.11}$$

The Hubble parameter $H(t)$, which is a measure of the expansion rate of the Universe, is defined by the scale factor:

$$H(t) = \frac{1}{a}\frac{da}{dt}. \tag{1.12}$$

Note that Hubble's constant H_0 in equation 1.10 is the Hubble parameter at the present ($H(t_0)$).

It will also be important to define a new metric which accounts for the expansion of the Universe. In the simplest case, where the Universe is Euclidean (flat), this is the Friedmann–Lemaître-Robertson–Walker (FLRW) metric:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}. \tag{1.13}$$

Note that the only difference between this and the Minkowski metric (equation 1.2) is that we now multiply the spatial components g_{ij} by $a^2(t)$. We will explore how the metric changes when we introduce perturbations and curvature, in chapter 2.

1.2.1 Redshift

Light that is emitted from distant galaxies will be subject to the expansion of the Universe, as it travels towards us. This causes the wavelength of the light to stretch, in a phenomenon known as cosmological redshift. We define redshift in terms of wavelength as:

$$z = \frac{\lambda_{\text{observed}} - \lambda_{\text{emitted}}}{\lambda_{\text{emitted}}} \tag{1.14}$$

One can also write a direct relation between redshift and the expansion of the Universe from the FLRW metric:

$$1 + z = \frac{1}{a(t)}. \tag{1.15}$$

1.2.2 Comoving distance

Measuring distances in cosmology is complicated by the expansion of the Universe, and the curvature of spacetime. Accounting for the expansion of the Universe, the comoving (or coordinate) distance travelled by light from time t to now is:

$$\chi(t) = \int_{t_0}^{t} \frac{1}{a(t')} dt'. \tag{1.16}$$

Here we have written the small distance element dx as $dt/a(t)$ (setting $c = 1$). It is normal to write this relation in terms of redshift, using the relations of equations 1.12 and 1.15. Doing this gives us:

$$\chi(z) = \int_0^z \frac{dz'}{H(z')}. \tag{1.17}$$

These expressions can likewise be used to calculate the comoving separation between any two objects. Distances to an object can also be measured through the angle θ subtended by that object in the sky, given by:

$$d_\mathrm{A} = \frac{s}{\theta}, \tag{1.18}$$

where s is the physical size of the object. This is known as the angular diameter distance. In an expanding Universe, the comoving size of the object is equivalent to $s/a(t)$, and the comoving distance to the object (from equation 1.16) is $\chi(t)$. This means that the angular size of the object evolves as $s/a(t)\chi(t)$, which we can substitute into equation 1.18 to derive an expression for d_A in terms of redshift:

$$d_\mathrm{A}(z) = \frac{\chi(z)}{1+z}. \tag{1.19}$$

The angular diameter distance increases with time, until a turnover point (at $a \sim 0.5$, $z \sim 1$), where it starts to decrease. Remembering equation 1.18, this means that after the turnover point, objects start to appear larger with time, rather than smaller as one would naïvely expect. This occurs because the Universe, and therefore the physical separation between the observer and the object, were smaller in the past.

Equation 1.19 is only valid for a Universe of zero curvature. The general expression in the presence of curvature is:

$$d_\mathrm{A}(z) = a(z) d_\mathrm{M}(z), \tag{1.20}$$

where d_M is the transverse comoving distance given by:

$$d_\mathrm{M}(z) = \frac{1}{H_0\sqrt{|\Omega_\mathrm{k}|}} \begin{cases} \sinh(\sqrt{\Omega_\mathrm{k}} H_0 \chi(z)),\ \Omega_\mathrm{k} > 0 \\ \sin(\sqrt{-\Omega_\mathrm{k}} H_0 \chi(z)),\ \Omega_\mathrm{k} < 0, \end{cases} \tag{1.21}$$

where Ω_k is the curvature critical density parameter - the contribution to the energy content of the Universe from intrinsic curvature. In the following sections, we will go into detail about how the expansion of the Universe is related to its energy contents.

1.3 Constituents of the cosmos

At large scales, we can approximate the Universe as a perfect fluid since it is statistically uniform and isotropic. In such a fluid, there are no internal mechanisms (like conduction or viscosity) that cause net transfer of momentum or energy. We can represent this with the energy-momentum tensor given by:

$$T^{\mu\nu} = (\rho + p)u^{\mu}u^{\nu} + pg^{\mu\nu}, \tag{1.22}$$

where ρ is energy density, p is pressure and u^{μ} is the four-velocity. In the rest frame this is $u^{\mu} = (1, 0, 0, 0)$, which leaves us with energy-momentum:

$$T^{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}. \tag{1.23}$$

In GR, local energy and momentum conservation are no longer described by the continuity ($\delta\rho/\delta t = 0$) or Euler equations ($\delta p/\delta x = 0$), since these are coordinate dependent. Instead, we express it with the covariant derivative:

$$\nabla_{\mu}T^{\mu\nu} = \frac{\delta T^{\mu\nu}}{\delta x^{\mu}} + \Gamma^{\mu}_{\alpha\mu}T^{\alpha\nu} + \Gamma^{\nu}_{\alpha\mu}T^{\mu\alpha} = 0. \tag{1.24}$$

Using the FLRW metric, we find that the only non-zero part of this equation is the energy conservation part, for $\nu = 0$:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \tag{1.25}$$

From this we can derive a general expression for the evolution of the density of any constituent of the Universe:

$$\rho_{\mathrm{x}}(a) \propto a^{-3(1+w_{\mathrm{x}})}, \tag{1.26}$$

where w_{x} is the equation of state parameter for constituent x, defined as $w_{\mathrm{x}} = p_{\mathrm{x}}/\rho_{\mathrm{x}}$. In figure 1.1 we show the evolution of ρ for matter, radiation and dark energy. Note that equation 1.26 is only valid for a time-independent w, which is not necessarily true in general.

To measure the relative contributions of each constituent to the total energy content of the Universe, it is useful to define the density parameter:

$$\Omega_{\mathrm{x}}(a) = \frac{\rho_{\mathrm{x}}(a)}{\rho_{\mathrm{c}}(a)}. \tag{1.27}$$

where ρ_{c} is the critical density, or the energy density that gives a Euclidean Universe. We will see why this is the case in section 1.4, and for now simply define it as:

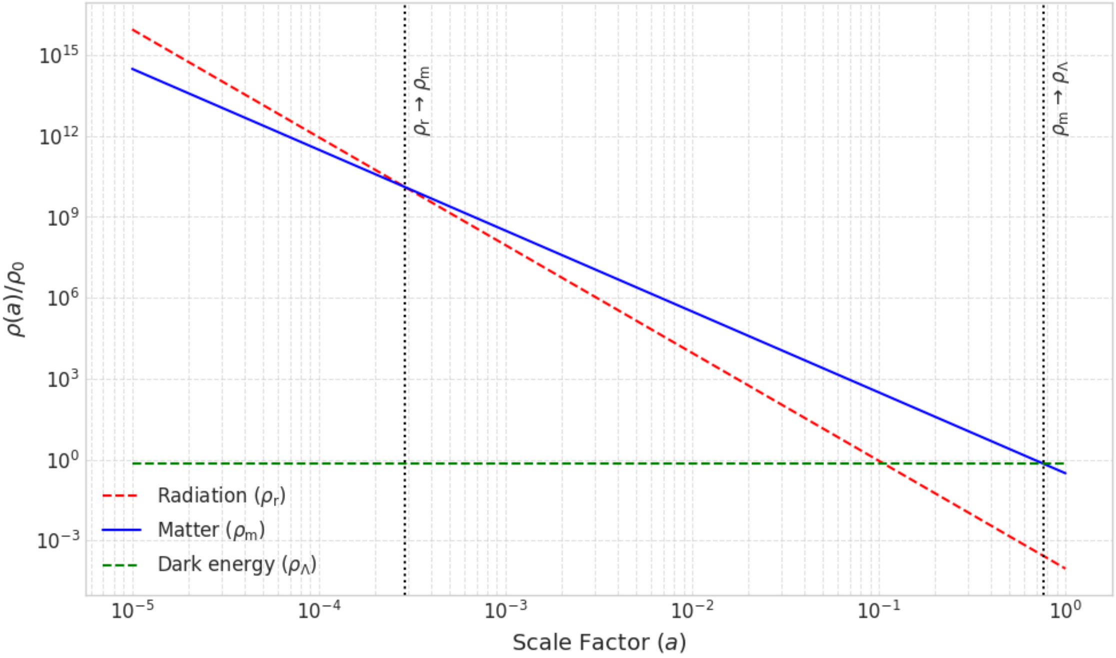

Figure 1.1: The evolution of density (ρ) for matter (dark + baryonic), radiation (relativistic neutrinos + photons) and dark energy, relative to the total density today (ρ_0). The dashed black vertical lines indicate the rough positions of the transfer from radiation to matter domination (left) and matter to dark energy domination (right).

$$\rho_c(a) = \frac{3H(a)^2}{8\pi G}, \tag{1.28}$$

where G is the Newtonian gravitational constant.

In the standard model of cosmology, known as ΛCDM, we distinguish between the following constituents to the energy density of the Universe:

- Non-relativistic matter: Denoted by Ω_m, this is the combination of baryonic matter (Ω_b) and cold dark matter (Ω_c). The latter was first introduced to explain the peculiar motions of galaxy clusters and stars within galaxies (i.e. galaxy rotation curve measurements) (Zwicky, 1933). These motions could only be explained by the presence of an additional, unobservable (hence "dark"), gravitationally-interacting substance. Furthermore, dark matter is also needed to explain the level of structure observed in the Universe today, where current measurements suggest it is much more abundant than baryonic matter ($\Omega_c \sim 5\Omega_b$). The observed amount of structure is also the reason dark matter is "cold"; "hot" dark matter particles (e.g. neutrinos) are too energetic to form bound objects like galaxies from an initially smooth

Universe. Finally, non-relativistic matter is pressureless, so $w = 0$ and energy density evolves with a^{-3}.

- Relativistic particles: Denoted by Ω_r, it is composed of photons and relativistic neutrinos, with $p = \rho/3$, $w = 1/3$. The energy density of radiation decays with time as a^{-4}, with an extra a term coming from the fact that the wavelength of these particles is stretched with the expansion of the Universe. In current times, the radiation density (Ω_r) is a negligible fraction of the total energy density, mostly consisting of free streaming photons that make up the cosmic microwave background (CMB)[3].
- Dark energy: The final, and most dominant energy contribution today, is dark energy represented by Ω_Λ. The discovery of the accelerating expansion of the Universe from observations of distant supernovae (Riess et al., 1998; Perlmutter et al., 1999) cemented the need for a new form of energy, with strong negative pressure. At its simplest form, this is the cosmological constant (Λ), originally introduced by Einstein to his field equations (equation 1.7) to give a static Universe solution [4]. In this form, dark energy is effectively the energy of the vacuum, with a negative equation of state ($w = -1$). According to current measurements, dark energy dominates the energy density in the Universe today.

1.4 Friedmann's equations

Einstein's field equations (equation 1.7) can be used, for a given metric and energy-momentum tensor, to derive a set of equations that govern cosmic expansion. This was done by Alexander Friedmann in 1922, assuming both Universal isotropy and homogeneity as defined by the FLRW metric, and the perfect fluid approximation given in section 1.3.

Let us first start by writing the FLRW in spherical coordinates, including a factor to allow for Universal curvature k:

$$ds^2 = -dt^2 + a^2(t)\left(\frac{dr^2}{1-kr^2} + r^2 d\phi^2 + r^2 \sin^2\phi\, d\theta^2\right). \tag{1.29}$$

Then, computing all of the necessary components of equation 1.7, for $T_{\mu\nu}$ defined by equation 1.23, we arrive at the first Friedmann equation:

---

[3] It is possible that the mass of one neutrino species is very close to 0, such that it would also be relativistic today.

[4] There is recent evidence that the dark energy is not a constant field, but is time-evolving DESI Collaboration et al., 2025c.

$$\frac{\dot{a}^2(t)}{a^2(t)} = \frac{8\pi G\rho}{3} - \frac{k}{a^2(t)} + \frac{\Lambda}{3}. \tag{1.30}$$

Now, recalling our definitions of the Hubble parameter (equation 1.12) and the density parameters (equation 1.27), we can re-write this as:

$$\frac{H^2(z)}{H_0^2} = \Omega_{\mathrm{m}}(1+z)^3 + \Omega_{\mathrm{r}}(1+z)^4 + \Omega_{\mathrm{k}}(1+z)^2 + \Omega_\Lambda, \tag{1.31}$$

where we have changed variables from a to z. We define $\Omega_{\mathrm{k}} = -k(1+z)^2/H^2(z)$, which is related to the sum of all other components by $\Omega_{\mathrm{k}} = 1 - \Omega_{\mathrm{tot}}$. We can infer Ω_{k} from measurements of expansion history, or through measurements on $H(z)$. If $\Omega_{\mathrm{k}} = 0$, we say the Universe is flat. If $\Omega_{\mathrm{k}} > 0$ or $\Omega_{\mathrm{k}} < 0$, we say the Universe is open or closed respectively. One can see in section 1.2.2, that our distance measurements are connected to redshift via the Hubble factor $H(z)$. This means that we can, given measurements of e.g. supernova redshifts and distances, derive values of the Hubble constant and energy densities. It is not unusual to simplify equation 1.31 at late-times, by assuming $\Omega_{\mathrm{r}} = \Omega_{\mathrm{k}} = 0$.

1.5 A brief history of the early Universe

In the moments after the Big Bang (up to $\sim 10^{-6}$ s), extremely high temperatures prevented the formation of hadrons, and the Universe is a dense quark-gluon plasma. As expansion continues, temperatures cool enough (~ 0.1 GeV) to form protons and neutrons[5]. Just before the temperature of the Universe reached 1 MeV, neutrinos decoupled from the rest of the plasma, as the rate of weak force interactions dropped below the expansion rate of the Universe. During this time, most of the plasma is in equilibrium through electromagnetic/weak interactions, and the mean free path of photons is roughly the size of an atom. In this period the Universe is radiation dominated, and the scale factor evolves as $a(t) \propto t^{1/2}$.

The Universe continues to expand, and at T~0.5 MeV, almost all of the electron-positron pairs are annihilated, dumping energy into the photon bath. This left behind a small number of excess electrons - enough to make Universe charge neutral. The weak interactions that govern proton-to-neutron conversion also freeze-out, resulting in protons being 6× more abundant than neutrons. At ~ 0.1 MeV, a process called Big Bang Nucleosynthesis (BBN) begins. This is marked

---

[5] At this stage anti-quarks also combine to form anti-hadrons, which are mostly annihilated, leaving a net amount of matter. If there there was a no matter excess, the baryons that we observe in the Universe today would not exist.

by the formation of deuterium (^{2}H), which started to occur when most photons dropped below the energy required to split it apart. This was immediately followed by the formation of heavier elements through nuclear reactions, predominantly Helium (^{4}He), but also trace amounts of Lithium and isotopes like ^{3}He and ^{3}H.

After this period of light element production, the Universe remained ionised for a long time. At $z \sim 3400$, or roughly 50 000 years after the Big Bang, the energy density of matter becomes greater than the energy density of radiation. This happens because, as we've shown, the density of radiation decreases faster ($\propto a^{-4}$, compared to a^{-3}). After this point, the scale factor evolves with time like $a(t) \propto t^{2/3}$. At $z \sim 1100$, or roughly 380 000 years after the Big Bang, the mean temperature of the Universe drops below the ionisation energy of Hydrogen (T$\sim$ 0.25 eV) [6], and neutral atoms begin to form. When this occurs, photons no longer scatter from free electrons (via Compton scattering), and begin to free stream. This marks the formation of what we observe today as the Cosmic Microwave Background (CMB), a snapshot of the early Universe. We will see in the next chapter, that although the CMB is an almost perfect blackbody at 2.7 K, there are small anisotropies caused by density fluctuations which are the seeds of structure in today's Universe.

---

[6] This also happens for Helium but in two stages, one of which begins earlier, at T$\sim$1 eV.

2

THE GROWTH OF STRUCTURE

In this chapter we will discuss inhomogeneities in the early Universe, and describe how these grew over time to produce the stucture we observe today. In the last chapter we worked with the approximation that the Universe was completely smooth and uniform. Here we will show how quantum fluctuations produced anisotropies at small scales, visible in the CMB and the clustering of galaxies.

2.1 Inflation

Structure is a result of small perturbations to the metric and the constituents parts (dark matter, baryons, radiation, dark energy[1]) of the Universe. Understanding how these perturbations evolve, requires solving a set of Boltzmann and Einstein equations for each constituent, and this in turn requires a set of initial conditions.

These initial conditions are closely tied to a fundamental issue in cosmology: the *horizon problem*. Observations of the CMB (Smoot et al., 1992) reveal that the Universe is extraordinarily uniform on large angular scales. Yet, under the standard Big Bang model, regions of the CMB that are widely separated could not have been in causal contact at the time of recombination. That is, they were separated by more than the comoving distance that light could have travelled since the Big Bang. This comoving distance, known as the comoving horizon, can be defined as:

$$\eta(a) = \int_0^t \frac{dt'}{a(t')} = \int_0^a d\ln a' \frac{1}{a'H(a')}, \tag{2.1}$$

where $1/a'H(a')$ is the comoving Hubble radius - the distance light can travel over one e-fold of expansion. In the standard radiation- and matter-dominated

---

[1] The nature and in fact, existence or not, of these perturbations depends on the type of dark energy considered - a cosmological constant for example does not perturb.

Universe, the comoving horizon at recombination η_* is ~280 h^{-1}Mpc (integrating equation 2.1 from $t = 0$ to t_*). Further, the comoving separation of two points in the CMB sky (for small angles) is $\chi(\theta) \approx \theta\,(\eta_0 - \eta_*)$, where $\eta_0 \sim 14200\,h^{-1}$Mpc is the comoving horizon today. From this, we derive that points separated by more than ~1.2° could not have been in causal contact in the standard picture.

A solution to this, is that the comoving Hubble radius, $(aH)^{-1}$, shrank in the early Universe. This behaviour could be realised during a phase of exponential expansion, known as inflation (Guth, 1981). In that epoch, quantum fluctuations that begin well inside the horizon are stretched to scales far beyond causal contact and effectively “freeze out". After inflation ends and the horizon grows again, these super-horizon modes re-enter, seeding the large-scale structure we see today. Inflation also solves other discrepancies in cosmology, like the flatness problem.

The simplest and most widely studied mechanism for inflation involves a scalar field, the inflaton ϕ, minimally coupled to gravity. Its energy density and pressure in a homogeneous FLRW background are given by

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi), \;\; p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi), \tag{2.2}$$

yielding an equation of state:

$$w = \frac{p}{\rho} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}. \tag{2.3}$$

Inflation occurs when the potential energy ($V(\phi)$) dominates over the kinetic term ($\dot{\phi}^2$), driving $w \approx -1$ and accelerating the expansion. This slow-roll regime requires that $V(\phi)$ be sufficiently flat, so the field evolves slowly toward the minimum of its potential. Note that w here is not related to dark energy, despite taking the same value during inflation.

2.2 Evolution of perturbations

Our goal is to understand primarily how small perturbations in the early Universe evolve over time, focusing particularly on the distribution of matter. In this section, we will mostly work in Fourier space, and define fluctuations in terms of wavenumber k. Fluctuations in Fourier space ($\tilde{\delta}(k)$) and configuration space ($\delta(r)$) are related by the Fourier transform:

$$\tilde{\delta}(\mathbf{k}) = \int \delta(\mathbf{r}) e^{-i\mathbf{kr}} d^3\mathbf{r}\,, \tag{2.4}$$

and inverse Fourier transform:

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \tilde{\delta}(\mathbf{k}) e^{i\mathbf{k}\mathbf{r}} d^3\mathbf{k} \tag{2.5}$$

where $\mathbf{r}$ is the separation vector between two points, and the prefactor $1/(2\pi)^3$ is included in the inverse transform by convention. Among other things, k-space is convenient when describing linear evolution because modes evolve independently.

Building on the work of the previous chapter, we now introduce perturbations to the metric using the conformal Newtonian gauge:

$$ds^2 = -(1 + 2\Phi(t, \mathbf{x}))dt^2 + a^2(t)(1 + 2\Psi(t, \mathbf{x}))\delta_{ij} dx^i dx^j \tag{2.6}$$

where Φ is the Newtonian gravitational potential, and Ψ is a spatial curvature perturbation.

The next step is to relate these metric perturbations to the perturbed Boltzmann equations describing the equation of state of matter. This can be done through Einstein's field equations (equation 1.7). This gives us an expression relating the potential Φ to dark matter and radiation densities:

$$k^2\Phi(\mathbf{k}, a) + 3\frac{a'}{a}\left(\Phi(\mathbf{k}, a)' + \frac{a'}{a}\Phi(\mathbf{k}, a)\right) = 4\pi G a^2[\rho_c\delta_c + 4\rho_r\Theta_{r,0}]. \tag{2.7}$$

where $\delta_c = \delta\rho_c/\rho_c$ is the dark matter overdensity, and $\Theta_{r,0}$ is the radiation temperature perturbation monopole. We have also changed to taking derivates with conformal time $d\eta = dt/a(t)$ (indicated by a' rather than $\dot{a}$). This equation is derived from the time-time component of Einstein's equation for scalar perturbations, with some simplifications. Namely, we neglect baryon perturbations using the tight-coupling approximation in Fourier space, and merge the neutrino/photon contributions. We have also set $\Psi = -\Phi$, which is true when photon (and neutrino) quadrupole moments are negligible, and there is no anisotropic stress component of $T^{\mu\nu}$.

To find solutions to equation 2.7, we make use of inflationary initial conditions, and consider both sub- and super-horizon modes, and how these evolve with time. It will also be useful in general to characterise perturbations as small fluctuations in curvature, or the gauge-invariant quantity $\mathcal{R} = \Psi + \frac{H}{\dot{\rho}}\delta\rho$. Outside of the horizon, $\mathcal{R}$ is conserved and potentials do not evolve significantly.

There are two important events that affect the evolution of perturbations, one is the matter-radiation equality, and the other is horizon crossing. For a given k-mode, these two events will have a different impact. For the large-scale (initially super-horizon; $k \ll aH$) modes which are relevant to large-scale structure studies today, the potential falls moving through matter-radiation equality (denoted by a_{eq}) as:

$$\Phi(\mathbf{k}) = \begin{cases} \frac{3}{5}\mathcal{R}(\mathbf{k}), & \text{matter domination} \\ \frac{2}{3}\mathcal{R}(\mathbf{k}), & \text{radiation domination.} \end{cases} \tag{2.8}$$

Modes which enter the horizon during radiation domination (before $k_{\mathrm{eq}} \sim 0.01\,\mathrm{Mpc}^{-1}$), have oscillating and decaying potentials equivalent to (a solution to equation 2.7):

$$\Phi(\mathbf{k}) \propto \mathcal{R}(\mathbf{k})\frac{\cos(k\eta/\sqrt{3})}{(\eta k)^2}\,, \tag{2.9}$$

for a plane wave perturbation. Physically, these are sound waves where radiation pressure counteracts gravity. In these potentials, matter overdensities can be found to grow as the logarithm of the scale factor, $\delta \propto \ln a$.

In the matter dominated era, we can ignore radiation terms in equation 2.7, giving us the solution $\Phi = constant$ [2]. In this regime, sub-horizon overdensities evolve according to the Meszaros equation Meszaros, 1974:

$$\frac{d^2\delta_{\mathrm{c}}}{dy^2} + \frac{2+3y}{2y(y+1)}\frac{d\delta_{\mathrm{c}}}{dy} - \frac{3}{2y(y+1)}\delta_{\mathrm{c}} = 0, \tag{2.10}$$

where $y = a/a_{\mathrm{eq}}$. The general solution to this equation contains a growing and decaying mode, which scale with a and $a^{-\frac{3}{2}}$ respectively.

2.2.1 Transfer function and growth factor

In general, as long as the initial perturbations $\mathcal{R}$ are adiabatic, we can write the matter density fluctuation $\delta_{\mathrm{m}}(k, a)$ at late times as:

$$\delta_{\mathrm{m}}(\mathbf{k}, a) = \frac{2}{5}\frac{k^2}{\Omega_{\mathrm{m}} H_0^2}\mathcal{R}(\mathbf{k})T(k)D(a). \tag{2.11}$$

where $T(k)$ is the transfer function, describing how different k modes are affected by processes like horizon crossing and matter-radiation equality crossing, and $D(a)$ is the growth factor, describing k-independent evolution with time. The transfer function is defined in terms of the Newtonian potential (Φ) as:

$$T(k) = \frac{\Phi(\mathbf{k}, a_{\mathrm{late}})}{\Phi(\mathbf{k}, a = 0)}\,, \tag{2.12}$$

---

[2] After matter domination, as dark energy begins to dominate ($a > 0.1$), Φ begins to decay with time, and growth begins to cease.

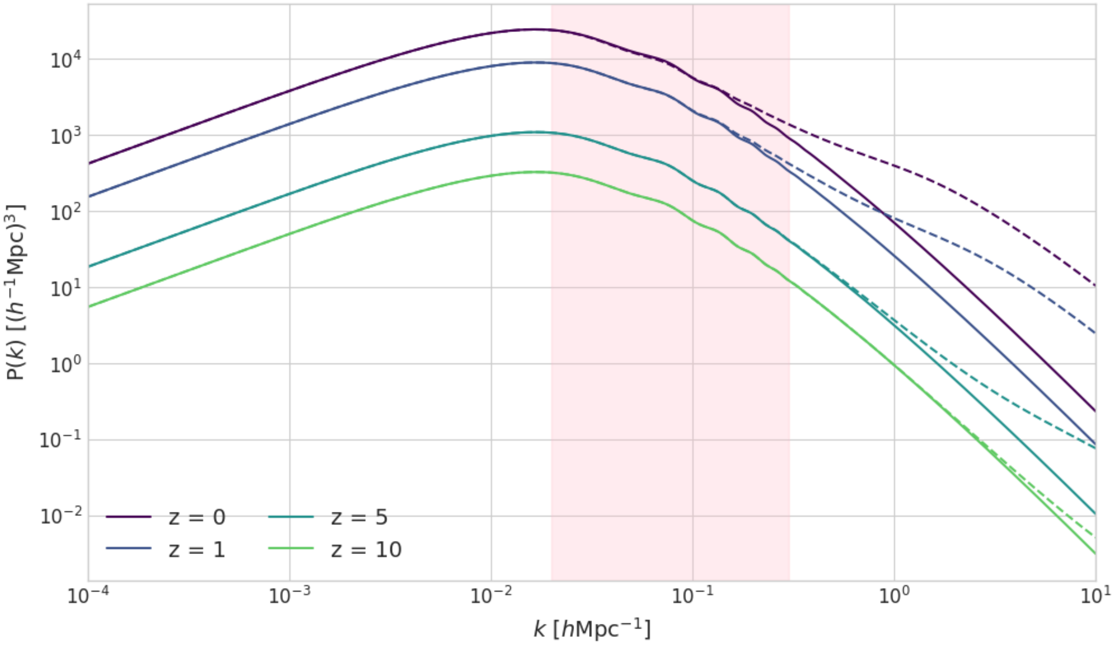

Figure 2.1: (Solid lines) the linear matter power spectrum (P_{L}) at various redshifts, where the red shaded region represents the approximate k-range of the BAO signature. (Dashed lines) the matter power spectrum including non-linearities.

where a_{late} is the scale factor far into matter domination. This expression is normalised such that the transfer function describes how the potential at k is suppressed relative to the largest modes.

In the matter dominated era, as a consequence of the fact that all modes grow at the same rate, the transfer function effectively becomes constant (hence we write $T(k)$ instead of $T(k,a)$). Then, the growth factor is the relevant function describing the potential evolution, defined as:

$$D(a) = a\frac{\Phi(\mathbf{k},a)}{\Phi(\mathbf{k},a_{\mathrm{late}})},\ a > a_{\mathrm{late}} \tag{2.13}$$

where this convention is set such that at late times, overdensities grow as $\delta_{\mathrm{m}} \propto D(a)$ ($D(a) \propto a$ in matter domination).

2.2.2 Linear matter power spectrum

The 2-point correlation function of linear-order density fluctuations in Fourier space is equivalent to: $\langle\tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}')\rangle = (2\pi)^3\delta^{\mathrm{D}}(\mathbf{k}+\mathbf{k}')P_{\mathrm{L}}(\mathbf{k})$, where $P_{\mathrm{L}}(\mathbf{k})$ is the linear matter power spectrum. In figure 2.1 we show $P_L(\mathbf{k})$ for several different redshifts, including deviations from linear fluctuations at high-k (dashed lines). This is a key measurement in modern cosmology, and we can relate it to the power spectrum of inflationary perturbations ($P_{\mathcal{R}}(\mathbf{k})$) using the functions defined above:

$$P_\mathrm{L}(k,a) = T(k)^2 D(a)^2 P_\mathcal{R}(k). \tag{2.14}$$

$P_\mathcal{R}(k)$ is the primordial power spectrum of the initial Gaussian perturbations $\mathcal{R}(k)$. It is parametrised as:

$$P_\mathcal{R}(k) = A_\mathrm{s}\left(\frac{k}{k_*}\right)^{n_\mathrm{s}-1}, \tag{2.15}$$

where A_s is the amplitude of the power spectrum, n_s is the spectral index and k_* is the pivot scale at which A_s is evaluated.

2.3 Baryon Acoustic Oscillations

In the period just before recombination, a dense hot soup of protons and electrons were tightly coupled to photons (via Compton scattering). As a result, the mean free path (λ_mfp) of these photons was negligible compared to the cosmic horizon. We can describe this epoch using the Boltzmann equations for photon and electron-baryon fluids, in the limit where $\lambda_\mathrm{mfp} \ll 1/H$.

To begin with we write a second order equation describing small perturbations Ψ and Φ (equation 2.6) in the coupled photon-baryon-electron soup, using the fluid approximation where only the monopole moment of the photon temperature perturbations is relevant:

$$\Theta_0(k,a)'' + \frac{a'R}{a(1+R)}\Theta_0(k,a)' + \mathbf{k}^2 c_\mathrm{s}^2 \Theta_0(k,a) = -\Phi(\mathbf{k},a)'' - \frac{a'R}{a(1+R)}\Phi(\mathbf{k},a)' - \frac{k^2}{3}\Psi(\mathbf{k},a). \tag{2.16}$$

Here $R = 3\rho_\mathrm{b}/4\rho_\gamma$[3], where ρ_γ is the photon energy density, and c_s is the speed of sound in the fluid, equivalent to:

$$c_\mathrm{s}(a) = \sqrt{\frac{1}{3(1+R(a))}}. \tag{2.17}$$

Equation 2.16 above is a driven, damped harmonic oscillator for the photon temperature monopole Θ_0, and describes acoustic oscillations in the photon-baryon-electron fluid. Oscillations continue to propagate in this fluid until the Universe has expanded enough such that electrons and protons combine (at the time of recombination; η_*), and photons which were tightly coupled by Compton scattering, began to free stream. Modes with different wavenumbers k oscillate at

---

[3] Note that R, and by extension c_s, is a function of time, since the relative energy densities of baryons and photons changes with the expansion of the Universe.

different rates, and their phases at the time of recombination become imprinted as peaks and troughs in the power spectra of both the CMB and matter distribution.

The characteristic Baryon Acoustic Oscillation (BAO) feature corresponds to the maximum distance that a sound wave could have propagated in the early universe before recombination — known as the sound horizon. The comoving distance of the sound horizon is given by:

$$r_{\mathrm{d}} = \int_0^{\eta_*} \frac{c_s(\eta)}{d\eta} = \int_{z_*}^{\infty} \frac{c_s(z)dz}{H(z)}, \tag{2.18}$$

where z_* corresponds to the redshift of recombination. This sound horizon also sets the physical scale of the BAO feature observed in the galaxy correlation function today. Studies of CMB anisotropies can be used to constraint the parameters that set the size of the sound horizon, which they predict to be $r_{\mathrm{d}} \approx 147\,\mathrm{Mpc}$.

3

Measuring large-scale structure

The expansion history, structure, and energy contents of the Universe can be measured through many different tracers. Observations of the CMB (Dicke et al., 1965; Penzias and Wilson, 1965; Smoot et al., 1992) provide a snapshot of the Universe at the time of recombination, encoding information about its early fluctuations. The angular position and height of the acoustic peaks in the CMB power spectrum are especially sensitive to parameters like the total matter density, the baryon-to-photon ratio, and spacial curvature.

Complementary to this, galaxy clustering measurements at lower redshifts have revealed how those early fluctuations evolved into the large-scale structure we see today (Zehavi et al., 2002). In particular, the imprint of BAO in the distribution of galaxies provides a standard ruler at multiple cosmic epochs. By comparing the observed angular and redshift separations of the BAO feature to the theoretically inferred sound horizon, one can map the expansion rate of the Universe and constrain dark energy (Cole et al., 2005; Eisenstein et al., 2005).

3.1 Clustering and redshift-space distortions

At large enough scales, galaxies and other structure (such as the Lyα forest) are biased-tracers of the underlying dark matter field. This means the tracer overdensity (δ_i) and dark matter overdensity (δ_m) are related by:

$$\delta_{\mathrm{i}}(\mathbf{k}) = b_i \, \delta_{\mathrm{m}}(\mathbf{k}) \, , \tag{3.1}$$

where b_i is the tracer bias. The power spectrum of fluctuations δ_{i} can further be related to the linear matter power spectrum (section 2.2.2) via the relation:

$$P_{\mathrm{ij}}(\mathbf{k}, z) = b_i^2(z)(1 + \beta_i(z)\mu^2(\mathbf{k}))^2 P_{\mathrm{L}}(k, z)\,, \tag{3.2}$$

The factor $(1 + \beta_i\mu^2)$ accounts for large-scale Redshift-Space Distortions (RSD) (Kaiser, 1987), and $\mu = k_{\parallel}/k$ is the cosine of the angle between $\mathbf{k}$ and the line-of-sight $k_{\parallel}$.

Redshift-space distortions are anisotropies in the clustering of objects in redshift-space due to peculiar motion, where the observed redshift of any object is given by:

$$1 + z_{\mathrm{obs}} = (1 + z_{\mathrm{cosmo}}) \left(1 - \frac{u_{\parallel}}{c}\right)^{-1}, \tag{3.3}$$

for cosmological redshift z_{cosmo} and peculiar velocity $u_{\parallel}$. RSD are classified in two distinct types:

- Linear RSD (Kaiser): These distortions are caused by large-scale infall of material into overdense regions of space. As a result the density field in redshift-space is compressed along the line-of-sight, which manifests as stronger clustering. This is a linear-order effect, modelled by the Kaiser formula (Kaiser, 1987): $P^s = b^2(1 + \beta\mu^2)^2 P$, where P^s is the redshift-space matter power spectrum (for a single tracer).
- Fingers of God (FoGs): These distortions are caused by random virial motions of objects at small-scales which stretch the density field along the line-of-sight, and consequently weaken clustering. Unlike the case above, this is a non-linear effect, and is normally treated phenomenologically.

The RSD parameter, β, is defined for point-tracers in linear theory as: $\beta_i = f\,/\,b_i$, where f is logarithmic growth-rate, related to the growth rate $D(a)$ (section 2.2.1) via:

$$f = \frac{d\ln D(a)}{d\ln a}. \tag{3.4}$$

This is also empirically related to the matter density via $f \sim \Omega_{\mathrm{m}}(a)^{0.55}$, a relation which holds true even in the presence of e.g. dynamical dark energy. The parameter σ_8 (sometimes defined in the derived parameter $S_8 = \sigma_8(\Omega_{\mathrm{m}}/0.3)^{0.5}$) quantifies the root-mean-square (RMS) amplitude of linear matter density fluctuations on scales of $8\,h^{-1}$ Mpc. It is computed by smoothing the linear matter power spectrum $P_m(k, z)$ with a spherical top-hat window function of radius $R = 8\,h^{-1}$ Mpc:

$$\sigma_8^2(z) = \int_0^{\infty} \frac{dk}{k}\, \frac{k^3 P_m(k, z)}{2\pi^2}\, |W(kR)|^2 \tag{3.5}$$

where $W(kR)$ is the Fourier transform of the top-hat window function. We can constrain the combination $f\sigma_8$ in linear theory from RSD measurements in e.g. galaxy clustering (Percival and White, 2009).

3.1.1 Measuring BAO

Baryon acoustic oscillations (BAO; see section 2.3) are a key feature in the large-scale clustering of matter. By measuring the BAO feature in the clustering of matter in the late Universe, such as in galaxy surveys or in the Lyα forest, one can compare the observed scale with its known value from the early Universe (from CMB measurements). This provides powerful constraints on the expansion history and geometry of the Universe.

The true physical sizes of the BAO peak in angular and velocity separations are related to the comoving size of the peak by:

$$\Delta\theta_{\mathrm{BAO}}(z) = \frac{r_{\mathrm{d}}}{d_{\mathrm{m}}(z)}, \quad \Delta v_{\mathrm{BAO}}(z) \approx \frac{r_{\mathrm{d}} H(z)}{1+z}, \tag{3.6}$$

where d_{m} is the comoving transverse distance (equation 1.21). From these measurements one can place constraints on dark energy and cosmic expansion rate, as was first done in Eisenstein et al., 2005. More recently, the Dark Energy Spectroscopic Instrument (DESI) survey has provided the best measurements of BAO from galaxies (DESI Collaboration et al., 2025c) and the Lyα forest (DESI Collaboration et al., 2025b). These have been combined to put precise constraints on cosmic expansion history, and currently favour an evolving dark energy equation-of-state (DESI Collaboration et al., 2025c).

We observe BAO as a peak in the 2-point (pt) correlation function, defined as:

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle. \tag{3.7}$$

This is also related to the power spectrum of fluctuations:

$$\xi_\ell(r,z) = \frac{i^\ell}{2\pi^2} \int_0^\infty dk\, k^2 j_\ell(kr) P_\ell(k,z)\,, \tag{3.8}$$

where j_l are the spherical Bessel functions, and:

$$\xi(r,\mu,z) = \sum_{\ell\ \mathrm{even}} L_\ell(\mu)\xi_\ell(r,z)\,. \tag{3.9}$$

Finally, L_ℓ are the Legendre polynomials, and P_ℓ are the power spectrum multipoles given by:

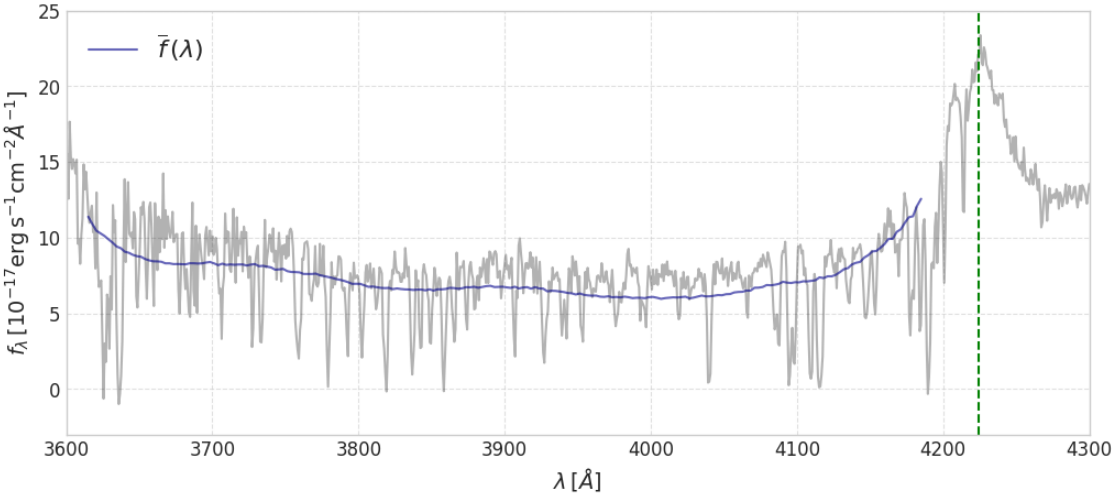

Figure 3.1: Flux in the Lyman-α forest of a quasar at $z = 2.5$, as a function of observed wavelength λ. Overlaid in blue is the mean expected flux, accounting for the average absorption level at each wavelength. Note that the ratio of flux and mean expected flux is $1 + \delta_\mathrm{F}$. The Lyα emission line is also highlighted (green dashed vertical) at $\lambda = \lambda_\alpha\,(1 + z_q)$.

$$P_\ell(k,z) = \frac{2\ell+1}{2}\int_{-1}^{+1} L_\ell(\mu)P_\ell(k,z,\mu)\,d\mu\,. \tag{3.10}$$

3.2 The Lyman-α forest

The Lyman-α (Lyα) forest (Lynds, 1971) is a series of absorption features in spectra of distant galaxies and quasars, produced when the light from these intersects neutral hydrogen (HI) gas in the Intergalactic Medium (IGM). Absorption occurs at the rest-frame wavelength of the Lyα transition, $\lambda_\alpha \approx 1216\,\text{Å}$, but appears across a wide range in the observer frame. This is because spectra are redshifted by cosmological expansion as they travel towards us, so that photons which were initially at $\lambda < \lambda_\alpha$ increase in wavelength, until they are readily absorbed by HI gas at λ_α.

The IGM is made up of sparse, smooth gas that traces the underlying dark matter distribution, making it a tracer of large-structure. The analogue of the matter overdensity field in the forest is the flux transmission field, defined as:

$$\delta_\mathrm{F}(\lambda) = \frac{F(\lambda)}{\overline{F}(\lambda)} - 1 \tag{3.11}$$

$$= \frac{f(\lambda)}{\overline{F}(\lambda)C(\lambda)} - 1\,, \tag{3.12}$$

where $F(\lambda)$ is the transmitted fraction of flux at wavelength λ in the observer frame, equivalent to $F = e^{-\tau_{\mathrm{HI}}}$ by definition, where τ_{HI} is the opacity of the IGM. $\overline{F}$ is the mean transmitted flux fraction at observed wavelength λ (or redshift z). In the second line we substitute $F(\lambda) = f(\lambda)/C(\lambda)$, where f is flux and C is the unabsorbed quasar continuum. Often in real analyses, $C(\lambda)$ and $F(\lambda)$ are combined into the mean expected flux $\overline{f}(\lambda) = F(\lambda)C(\lambda)$. In figure 3.1 we show the observed Lyα forest flux (grey) in a quasar at $z \sim 2.5$, overlaid (blue line) with the mean expected flux at each wavelength.

3.2.1 Optical depth

The general expression for the IGM optical depth along a line-of-sight, in terms of mean neutral hydrogen density n_{HI} and cross-section σ_{HI}, is:

$$\tau_\alpha = \int_{\mathrm{LOS}} n_{\mathrm{HI}}(x)\sigma_\alpha(\nu)\,dx, \tag{3.13}$$

Assuming neutral hydrogen is uniformly distributed in the IGM, we can substitute an expression for cross-section and replace the proper path length (dx) integral with redshift $dx = c\,dz\,/(1+z)H(z)$, which gives us the Gunn-Peterson optical depth (Gunn and Peterson, 1965):

$$\tau_{\mathrm{GP}}(z) = \frac{\pi e^2 f_{12}\lambda_\alpha}{m_e c H(z)} n_{\mathrm{HI}}(z). \tag{3.14}$$

Here m_e and e are the electron mass and charge respectively, and f_{12} is the oscillator strength of Lyα line. f_{12} and consequently the Lyα cross-section ($\sigma_\alpha \propto f_{12}$) are relatively large, meaning absorption is strong even if only a small fraction of the IGM is neutral hydrogen. The optical depth is also be related to the neutral hydrogen fraction X_{HI} via the expression (Barkana and Loeb, 2001):

$$\tau_{\mathrm{GP}}(z) \propto \frac{X_{\mathrm{HI}}(\Omega_{\mathrm{b}}h^2)(1+z)^{3/2}}{\Omega_{\mathrm{m}}^{1/2}}. \tag{3.15}$$

Before reionisation, when $\tau_{\mathrm{GP}} \gg 1$, the IGM was completely opaque to Lyα photons, causing complete absorption blue-wards of the Lyα line - the Gunn-Peterson trough. After reionisation, where $\tau_{\mathrm{GP}} \ll 1$ (but still non-zero), there is still enough residual neutral hydrogen to produce narrow Lyα absorption lines. Observations of Gunn-Peterson absorption have in fact suggested that the end of cosmic reionisation was around $z \sim 6$ (Fan et al., 2006).

The Lyα optical depth can be directly linked to density fluctuations in the underlying dark matter field:

$$\tau_\alpha = A(1+\delta_m)^\alpha\,, \tag{3.16}$$

where A is a constant dependent on redshift and the state of the IGM, and $\alpha \sim 1.6$ (Weinberg and et al., 1999). This is the Fluctuating Gunn-Peterson Approximation (FGPA), a fast semi-analytical method to modelling the Lyα forest, which has been proven to work well with high-resolution N-body simulations (Sorini et al., 2016). In later sections, when studying the impact of systematics on cosmology measured from Lyα correlation functions, we make use of synthetic datasets that use the FGPA approximation.

3.2.2 Cosmology with Lyα absorption

Like galaxies, the flux transmission field of the Lyα also linearly traces underlying dark matter at large enough scales, but with an additional term:

$$\delta_{\mathrm{F}}(\mathbf{k}) = b_\alpha \delta_{\mathrm{m}}(\mathbf{k}) + b_\eta \eta(\mathbf{k})\,. \tag{3.17}$$

Here, b_α and b_η are the density bias and velocity divergent bias of the Lyα forest respectively, where $\eta(\mathbf{k}) = f\mu^2\delta_m(\mathbf{k})$ is the radial peculiar velocity gradient. This additional term changes the Lyα RSD parameter, β_α: $\beta_\alpha = b_\eta f/b_\alpha$. It is important to highlight that for the Lyα forest, the combination $f\sigma_8$ is degenerate with the velocity divergence bias b_η, and cannot be measured except in combination with galaxies or quasars, since for those $\beta = f/b$.

The 3-dimensional Lyα power spectrum is similarly written (McDonald, 2003):

$$P_{\mathrm{F}}(\mathbf{k}, z) = b_\alpha(z)^2(1+\beta_\alpha(z)\mu^2)^2 D_{\mathrm{NL}}(\mathbf{k}, z) P_{\mathrm{L}}(k, z)\,, \tag{3.18}$$

where now we have included a factor D_{NL}, that models non-linear deviations from P_{L} at smaller scales. These additional complexities are due to the gas pressure, thermal effects and peculiar velocities within the IGM where absorption is produced. This function can be described by the empirical model of Arinyo-i-Prats et al., 2015:

$$D_{\mathrm{NL}}(k) = \exp\left[(q_1\Delta^2(k) + q_2\Delta^4(k))\left(1 - \left(\frac{k}{k_v}\right)^{a_v}\mu^{b_v}\right) - \left(\frac{k}{k_p}\right)^2\right], \tag{3.19}$$

where $\Delta(k) = (2\pi^2)^{-1}k^3 P_{\mathrm{L}}(k)$ is the dimensionless linear matter power spectrum. The parameters q_{x} account for non-linear growth, while parameters with subscript v account for thermal effects and velocities, and k_p accounts for gas pressure.

4

The Dark Energy Spectroscopic Instrument

The Dark Energy Spectroscopic Instrument (DESI) is a next-generation galaxy redshift survey designed to map the 3D structure of the universe with unprecedented precision. Its primary goal is to investigate the nature of dark energy, which DESI will achieve by observing the redshifts of more than 40 million galaxies and 3 million quasars (Schlafly et al., 2023), over 14000 deg^2 of sky.

Measuring the BAO feature (see section 3.1.1) is a key objective of DESI. By doing so at multiple redshifts between $z \sim 0.1 - 3.5$, DESI can put constraints on the expansion history of the universe. These then allow us to constrain models of dark energy, testing whether cosmic acceleration is driven by a cosmological constant, a dynamical dark energy field, or modifications to general relativity. In parallel, DESI will exploit RSD to measure the growth of cosmic structure (see section 3.1), offering a powerful means to test the validity of general relativity on cosmological scales.

4.1 Instrument and survey design

DESI is mounted on the 4-metre Mayall telescope at Kitt Peak (DESI Collaboration et al., 2022), and has 5000 robotically controlled spectroscopic fibres in its focal plane (Poppett et al., 2024). The focal plane is split into 10 "petals", each of which contains 500 fibres that are fed to an independent spectrograph which covers the wavelength range 3600 to 9800 Å. Each spectrograph consists of a blue (3600-5550Å), red (5550-6560Å) and near infrared (NIR; 6560-9800Å) arm, with resolutions varying from R = 2000-5000. 10 Guide, Focus and Alignment (GFA) cameras are also positioned on the periphery of each focal plane, to allow for accurate tracking and calibration of the instrument during observing runs. These

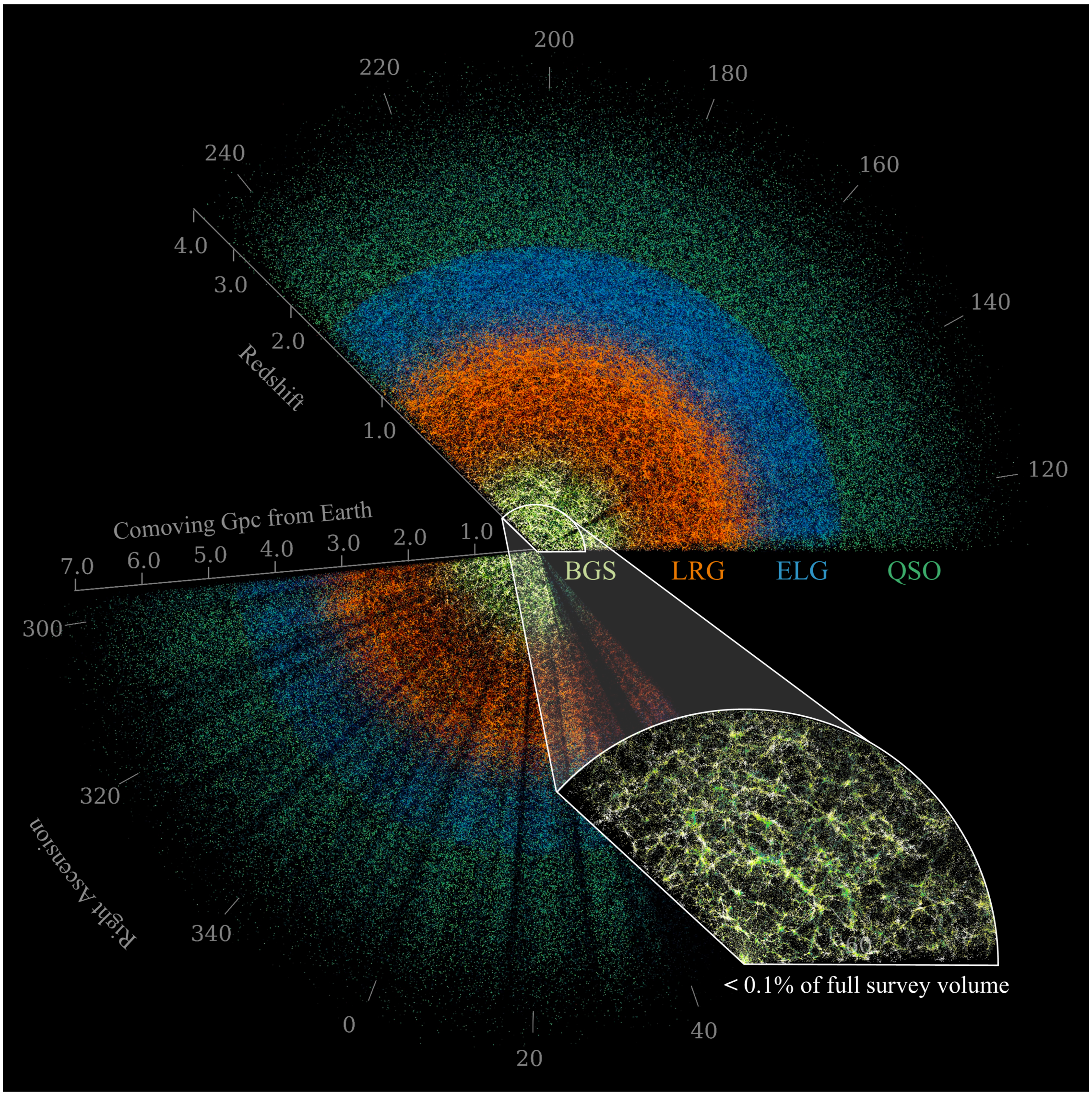

Figure 4.1: Observed galaxies and quasars from the first year of DESI data, known as data release 1 (DR1). Bright galaxy survey (BGS) targets are observed in "bright" time, where only the brightest (and closest) galaxies can be seen. Luminous red galaxies (LRG) are categorised by their red colour, and emission line galaxies (ELGs) have prominent spectral lines from active star formation. Quasars (QSO) are extremely bright, and thus make up the most distant observed sample. Note that the Lyα forest is not visible here, and that the galaxies in this image have varying declinations, which has been compressed for visualisation.

factors, combined with the 3 degree diameter, allow DESI to quickly collect high quality spectra.

The DESI pipeline (Guy et al., 2023) processes raw 2D CCD images from the ten spectrographs into calibrated 1D spectra. Raw CCD frames are converted to 1D spectra through a sequence of calibration steps, and sky emission lines are subtracted using designated "sky fibres" (fibres on empty sky). In parallel, flux calibration is performed using standard stars to convert measured counts into flux-calibrated spectra.

DESI uses the template fitting algorithm `Redrock` (Bailey et al., 2025) to classify and assign redshifts to each calibrated spectrum. `Redrock` contains a library of PCA-based templates for galaxies, stars and quasars, derived from Sloan Digital Sky Survey (SDSS) spectra. In practice, `Redrock` is very successful in correctly classifying and estimating redshifts for all target types, but to improve these things for quasars, two additional algorithms are used. The first is `QuasarNET`, a neural-network classifier specially trained on quasar spectra to recover broad-line redshifts missed by Redrock. Secondly, an afterburner is run on galaxy targets to identify the broad MgII emission line which is characteristic in quasars, helping to recover a number of mis-labelled objects.

Spectroscopic targets in the DESI survey are selected using imaging data from the DESI Legacy Imaging Surveys (Dey et al., 2019), a combination of three wide-area surveys: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). These surveys provide deep, uniform imaging in the optical g, r and z bands, which are complemented by mid-infrared photometry from the WISE $W1$ and $W2$ bands. Photometric selection algorithms are used to identify different cosmological tracers, including emission line galaxies (ELGs), luminous red galaxies (LRGs), and quasars, each optimised for different redshift ranges and scientific goals.

Once targets are selected, they are assigned to DESI's robotically controlled fibre positioning system. For each observation pass, individual fibres are automatically placed on their assigned targets with a precision of ~10 microns, ensuring optimal light capture into the spectrographs. This high-precision fibre placement is critical for achieving DESI's goals in redshift measurement accuracy and spectroscopic completeness across the full focal plane.

4.2 DESI Lyα

The Lyα forest is observed in DESI quasars with redshifts $\gtrsim 2$ (due to the finite spectrograph range), where each DESI Lyα quasar will receive a total of 4000 s effective exposure time. For details on the effective exposure time calculator, see Tie et al., 2020. The signal-to-noise and resolution of DESI Lyα forests is sufficient to perform precise percent-level BAO measurements, as well as the extraction of small-scale clustering information through the one-dimensional Lyα forest power spectrum.

The first data release of DESI, known as the early data release (EDR) was compiled from measurements made during the survey validation (SV) phase [1], and contained 68 750 Lyα quasars (60 deg^{-2}). The Lyα flux-transmission fields computed from each of these quasars is estimated and presented in Ramírez-Pérez et al., 2024, and the first clustering analysis of this data was done in Gordon et al., 2023, which is presented in chapter 5.

Following this, the 1st year of DESI observations was released, known as data release 1 (DR1) (DESI Collaboration et al., 2025a), shown in figure 4.1 (all tracers except the Lyα forest). This represented a significant increase in the number of Lyα quasars (to ~420 000), giving the first BAO measurement (Adame et al., 2025b) and the first cosmological analysis in combination with other tracers (Adame et al., 2025c). Importantly, DR1 also allowed for the first Lyα full-shape analysis—an approach that goes beyond BAO to exploit the full information content of the Lyα correlation functions (Cuceu et al., 2025a), including redshift-space distortions and broadband shape. Chapter 6 details a study of the impact of redshift errors on this full-shape analysis (Gordon et al., 2025).

---

[1] This was an early campaign conducted before the start of the main survey from December 2020 till April 2021, aiming to verify and refine all components of the survey pipeline

5

CLUSTERING MEASUREMENTS FROM THE LYα FOREST IN THE DESI EARLY DATA RELEASE

5.1 Introduction

The Lyman-α (Lyα) forest is a dense series of absorption lines found in the spectra of distant quasars, that we use to study the large-scale structure of the universe. These absorption features arise as a result of diffuse neutral hydrogen in the IGM (Lynds, 1971), which readily absorbs light at the Lyα wavelength $\lambda_\alpha = 1215.67$Å, and can be used as a biased tracer of the underlying matter density (Weinberg, 2003). Due to the low rest-frame wavelength of the Lyα transition, its absorption is only detected by optical spectrographs for quasars with redshift $z \geq 2$. The Lyman$-\alpha$ forest allows us to measure the expansion of the Universe at redshifts larger than those accessible by spectroscopic galaxy surveys (z <1.5; Alam et al., 2021).

The environment that gives rise to the Lyα forest has been well studied using hydro-dynamical simulations (Cen et al., 1994; Miralda-Escudé et al., 1996), where a combination of photo-ionisation heating and adiabatic cooling leads to a tight relation between temperature and gas density $T \propto (\rho/\overline{\rho})^{\gamma-1}$ (Hui and Gnedin, 1997). The transmitted flux fraction is related to the optical depth of the Lyα forest by $F = e^{-\tau}$. At sufficiently large scales ($> 10\,h^{-1}$Mpc), the power spectrum of transmitted flux $\delta_{\mathrm{F}} = F/\overline{F} - 1$, is found to trace linearly the power spectrum of $\delta(\mathrm{x}) = \rho(\mathrm{x})/\overline{\rho} - 1$ (Slosar et al., 2011). However, due to non-linear growth, thermal broadening, and Jeans smoothing, this relation becomes more complex and non-linear at smaller scales (McDonald, 2003; Arinyo-i-Prats et al., 2015; Givans et al., 2022).

The first studies of clustering in the Lyman$-\alpha$ forest looked at one-dimensional (1D) correlations along the line-of-sight towards a handful of quasars (Croft et al., 1998; McDonald et al., 2000). These 1D measurements were later repeated using larger data sets from the Sloan Digital Sky Survey (SDSS) McDonald et al., 2006, from the Baryon Oscillation Spectroscopic Survey (BOSS) (Palanque-Delabrouille et al., 2013), and the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Chabanier et al., 2019).

This paper, on the other hand, focuses on three-dimensional (3D) correlations in the Lyman$-\alpha$ forest. Using early data from the BOSS survey, Slosar et al., 2011 presented the first measurement of 3D correlations in the Lyman$-\alpha$ forest. These correlations were also computed from larger datasets to measure the scale of Baryon Acoustic Oscillations (BAO) around $z = 2.3$, in BOSS and eBOSS (Busca, N. G. et al., 2013; Kirkby et al., 2013; Slosar et al., 2013; Delubac et al., 2015; Bautista et al., 2017; de Sainte Agathe et al., 2019; Bourboux et al., 2020). The cross-correlation of quasars and the Lyman$-\alpha$ forest was also measured in BOSS (Font-Ribera et al., 2013), and used to improve the BAO measurements from the Lyα datasets (Font-Ribera et al., 2014b; du Mas des Bourboux et al., 2017; Blomqvist, Michael et al., 2019; Bourboux et al., 2020). The final eBOSS analysis (Bourboux et al., 2020), using data from SDSS DR16, included 210 005 Lyα forest from quasars at z >2.1 that were used in the autocorrelation, and 341 468 at z >1.77 that were also used in the cross-correlation. More recently, the full shape of the 3D Lyα correlation has been used to measure the Alcock-Paczynski effect in eBOSS data (Cuceu et al., 2022a), which combined with Lyα BAO gives the best measurement of the expansion rate of the Universe above $z = 1$.

We use early data from the Dark Energy Spectroscopic Instrument (DESI) (Levi et al., 2013; DESI Collaboration et al., 2016a) to measure the autocorrelation of the Lyman$-\alpha$ forest and its cross-correlation with quasars, and compare our results with those from the eBOSS collaboration (Bourboux et al., 2020). We also show in section 5.5 that we detect the BAO peak to high-confidence in our dataset. Given the early stage and relatively low statistical power of our data, we choose not to present measurements of the BAO scale parameters. Throughout the latter half of the paper, we make use of the best fit ΛCDM cosmological parameters from Planck 2018 (Planck Collaboration et al., 2020)[1], where $\Omega_{\rm m} = 0.3153$, to compute correlations in terms of co-moving separation. This allows us to combine a wide range of redshifts into a single measurement while preserving the BAO feature.

The structure of the paper follows roughly the steps required to perform the

---

[1] See table 2 of Planck Collaboration et al., 2020, where we use *TT,TE,EE+lowE+lensing* results.

end-to-end analysis. We start by describing the DESI instrument and the early data release, and outline how we measure the Lyα transmission catalogue from high-redshift quasar spectra in section 5.2. We then compute the auto- and cross-correlation and covariance matrices in section 5.3. In section 5.4, we discuss the physical model of Lyman$-\alpha$ forest correlations and contaminants, and in section 5.5 we present the results of fitting this model to our data and discuss the consistency of these results with the final eBOSS analysis (Bourboux et al., 2020). Finally, we summarise our results in section 5.6.

The Lyman$-\alpha$ forest dataset used here is described in detail in Ramírez-Pérez et al., 2024. Our analysis is also accompanied by a paper analysis on early DESI mock datasets (Herrera-Alcantar et al., 2025b), which utilises the same 13-parameter model to describe the auto- and cross-correlation. We further highlight two papers where systematic effects, including quasar redshift errors (Bault et al., 2025) and instrumental effects (Guy et al., 2025), are studied on the same early DESI data set.

5.2 Data

DESI is a multi-object spectrograph installed at the Mayall 4-meter telescope at Kitt Peak National Observatory (DESI Collaboration et al., 2022). DESI is equipped with 5000 robotically controlled fibres across 10 independent spectrographs, that have an approximately 3-degree field of view in each pointing (DESI Collaboration et al., 2016b; Silber et al., 2023; Miller et al., 2024). Each measured spectrum covers an observed wavelength range between 3600 and 9800Å with a pixel width of 0.8Å, which is split between 3 independent spectrograph arms ranging between 3600 – 5930, 5600 – 7720, and 7470 – 9800Å respectively (DESI Collaboration et al., 2022).

In May 2021, DESI started a 5-year campaign which will observe 3 million quasar (Chaussidon et al., 2023) and 36 million galaxy targets (Hahn et al., 2023; Raichoor et al., 2023; Zhou et al., 2023), over an area of 14 000 deg^2. These targets are selected from the DESI Legacy Imaging Survey (Zou et al., 2017; Dey et al., 2019; Schlegel et al. 2023, 2023). These observations are made thanks to extensive supporting software packages: one to select targets for spectroscopic observation (Myers et al., 2023), one to assign fibres to targets (Poppett et al., 2024), one to optimise survey operations (Schlafly et al., 2023), an exposure-time calculator (Kirkby et al. 2023, 2023), and a spectroscopic pipeline to reduce the data and calibrate the spectra (Guy et al., 2023).

We analyse data from the Early Data Release (EDR) of DESI (DESI Collaboration et al., 2023a), which is publicly available[2] and contains 68 750 quasars. EDR contains data from a pre-survey designed to optimise target selection algorithms (Chaussidon et al., 2023; DESI Collaboration et al., 2023b), refine the observational procedure, and generally evaluate the quality of spectra and the redshift accuracy (Alexander et al., 2022; Brodzeller et al., 2023). To increase the statistical power of the measurements, we use data from the Early Data Release plus the first two months of main survey (EDR+M2), containing 249 941 quasars.

The sample footprint is shown in figure 5.1, where we highlight the each survey phase and the eBOSS Data Release 16 (DR16) survey footprint (Bourboux et al., 2020). Even though we analyse all sources from EDR+M2 together, the typical signal-to-noise (SNR) of M2 and EDR spectra varies. Targets observed in M2 are limited to one observing pass[3], which will increase to 4 by the end of DESI observations. Quasars identified as Lyα quasars (z >2) receive more passes in general than other targets to reach the required SNR per pixel across the Lyα forest. On the other hand, roughly half of EDR spectra present 4 passes and half 1.3 (Chaussidon et al., 2023).

5.2.1 Quasar spectra

The DESI pipeline uses a robust spectroscopic reduction pipeline (Guy et al., 2023), in combination with a template-fitting algorithm `Redrock` (RR; Bailey et al., 2025) to give object classifications and estimate their redshifts. Templates used by RR are formed by a linear combination of bases resulting from the principal component analysis (PCA) decomposition of Sloan Digital Sky Survey spectra (Bailey et al., 2025) (see Brodzeller et al., 2023 for comparison when using a new set of quasar templates).

To quantify the ability of the pipeline to classify quasars we use two terms, purity and completeness. Purity is the percentage of classified quasars that are true quasars, which is >99% in our pipeline. Completeness is the percentage of true quasars identified as quasars, which we find to be ~86%. To increase this, we run two algorithms on the RR output. The first attempts to identify the MgII broad emission line – a feature only found in quasars – in the spectra of sources classified as galaxies (Chaussidon et al., 2023), and is especially effective at 0.5< z <1.5. The other is `QuasarNet` (QN; Busca and Balland, 2018; Farr et al., 2020a), a convolutional neural network trained using a large number of visually

---

[2] https://data.desi.lbl.gov/public/edr/

[3] Equivalent to nominal 1000s of effective exposure time, which is exposure time accounting for observing conditions and fibre effects.

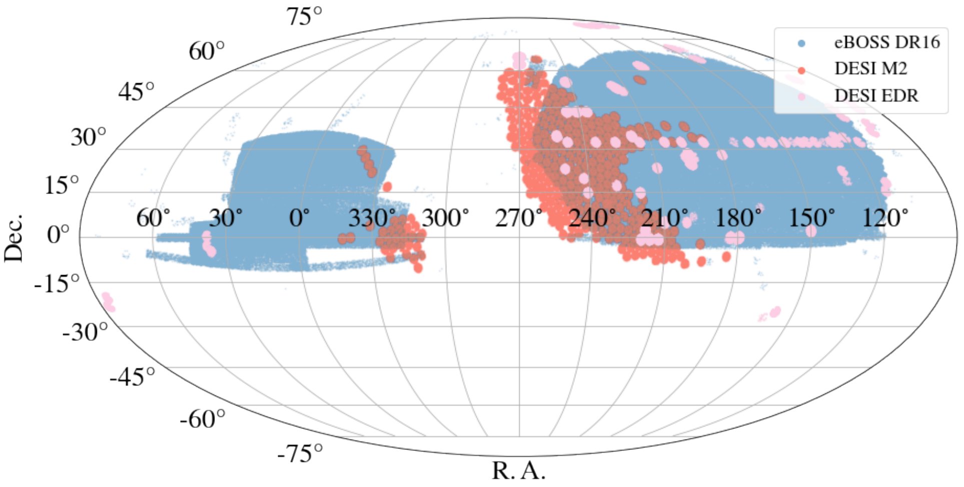

Figure 5.1: Footprint of quasar targets in DESI Early Data Release (purple), the first two months of main survey (red), and eBOSS DR16 (blue) (Bourboux et al., 2020). The EDR and M2 surveys cover an area of 250.1 deg^2 and 1290.9 deg^2.

inspected BOSS spectra, to identify faint quasars missed by both RR and the first afterburner, and is more useful at $z \gtrsim 2$. These algorithms increase the spectral completeness from ~86 to ~94% (Alexander et al., 2022), and only cause a ~1-2% drop in spectral purity at $z < 1$ (Chaussidon et al., 2023).

`Redrock` is very reliable in estimating the redshifts of high-quality spectra (Alexander et al., 2022). The number of catastrophic redshifts, where the difference between visually-inspected and RR redshifts is $\Delta z > 0.01$, is only ~1.5%. For sources presenting a disagreement between RR and QN redshifts, the pipeline computes the final redshift by re-running RR using the QN redshift as prior and only quasar templates. This reduces the fraction of catastrophic redshifts from 1.5 to 1% (Alexander et al., 2022).

We obtain an overall density of 210 quasars deg^{-2} (60 for z >2.1) (Chaussidon et al., 2023), which surpasses the DESI requirements of 170 quasars deg^{-2} (50 at z >2.1; DESI Collaboration et al., 2016a). Throughout this work, we will only analyse spectroscopically identified quasars that were also selected as quasar targets (i.e. we ignore the small fraction of quasars found in other target classes; Myers et al., 2023), to keep the purity of the catalogue as high as possible. These have a density of around 54 deg^{-2} at z >2.1. In figure 5.2 we show the redshift distributions of

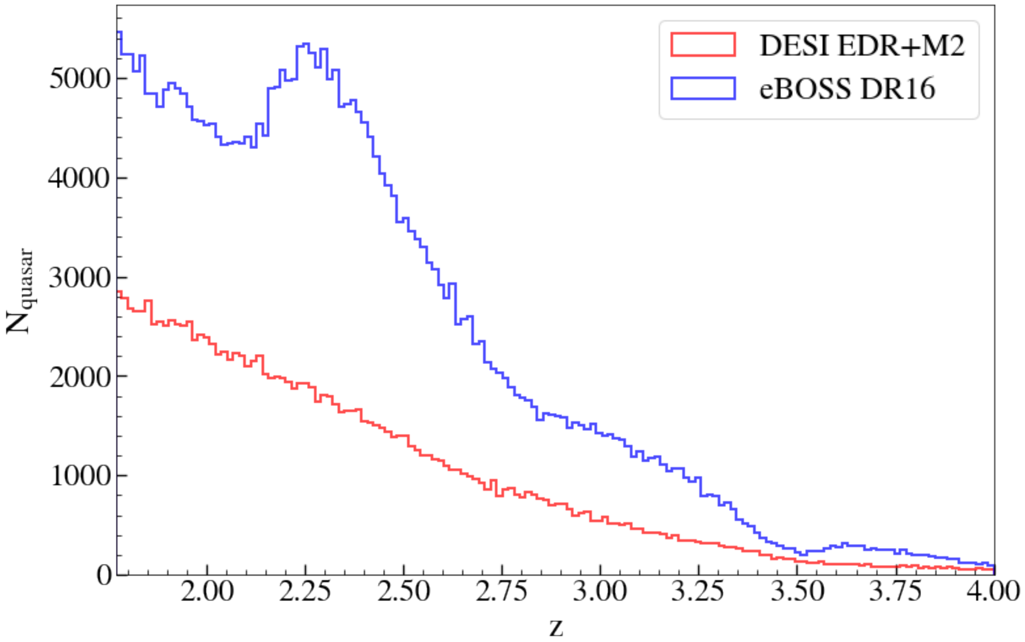

Figure 5.2: Redshift distribution of high-redshift quasars in DESI EDR+M2 and eBOSS DR16 (Bourboux et al., 2020), with 147 899 and 341 468 in each sample respectively. Only quasars between $z = 1.77$ and $z = 3.75$ feature in our analysis.

our EDR+M2 Lyα quasars and include for comparison the catalogue used in the Lyα analysis of eBOSS DR16 (Bourboux et al., 2020, hereby `dMdB20`).

Figure 5.3 shows one such quasar with $z = 2.44$. Purple labels indicate broad, distinctive emission lines often observed in high-redshift quasars, including the MgII line. We also highlight the Lyα forest in blue, which exhibits a high density of absorption lines. Due to the rest-frame wavelength range of the Lyman$-\alpha$ forest and the observed redshift distribution of quasars, most of our data falls within the blue arm of the spectrographs.

5.2.2 Pixel masking

The Lyα catalogue used to obtain the results in this paper is described in Ramírez-Pérez et al., 2024. We present here a summary of how we mask contaminants in the data, and refer the reader to Ramírez-Pérez et al., 2024 for a detailed description of the process.

In this section and while computing the flux transmission field, we use Lyα pixels of width 0.8Å in the observed frame - the resolution provided by the spectroscopic pipeline (Guy et al., 2023). When computing the correlation functions

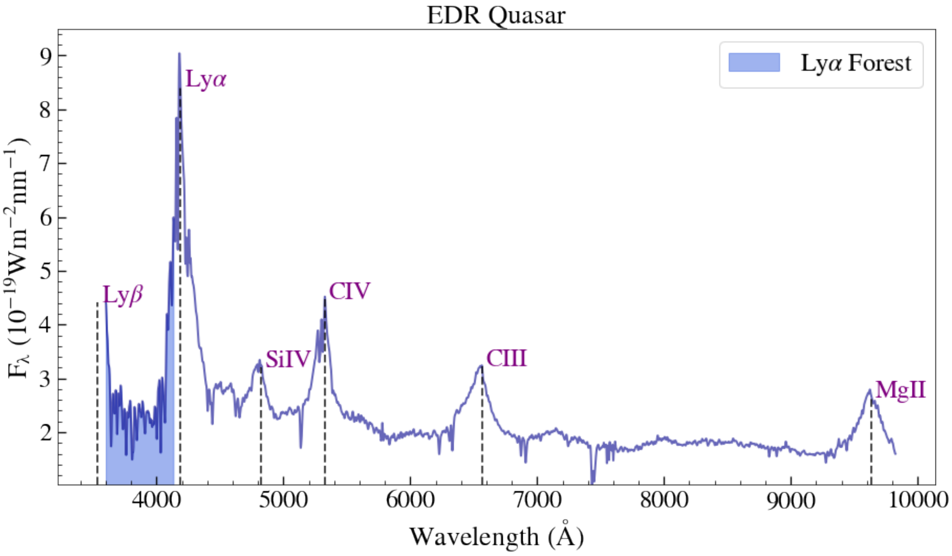

Figure 5.3: An example of a DESI EDR spectra at $z = 2.44$. Black dashed lines indicate the central wavelength of the strongest quasar emission lines, including the Lyα used in our analysis. The shaded blue region shows the Lyα forest between the Lyα and Lyβ lines, the latter only partially observed in this specific example. Also of note is the MgII line, used in the pipeline to re-classify galaxies as quasars and the region between the CIV and CIII] lines, used for the re-calibration step described in section 5.2.3.

(section 5.3) and the metal matrices (section 5.4.3.4), we combine these into 2.4Å analysis pixels to reduce the computation time, without affecting the measured correlations.

The two primary contaminants of our Lyα forests are Broad Absorption Line (BAL) and Damped Lyman-α Absorption (DLA). BALs are broad spectral troughs that arise from high-velocity gas outflows from quasars. They often fall into the Lyα forest region and are difficult to model. These have been identified in the past using convolutional neural networks (Guo and Martini, 2019), and a similar approach has been used in DESI (Filbert et al., 2024). Identified BALs are then masked in the affected spectra depending on the width of the trough, following the procedure described in Ennesser et al., 2022. Masking avoids discarding whole Lyα forests where there are often valuable pixels, as was done in previous Lyman$-\alpha$ forest BAO studies, like `dMdB20`.

DLAs are large absorption systems with column densities of neutral hydrogen $N_{\text{HI}} > 10^{20.3}\,\text{cm}^{-2}$. Like the Lyman$-\alpha$ forest, DLAs are a tracer of the underlying matter distribution but present a higher bias than the Lyα forest (Font-Ribera et al., 2012; Pérez-Ràfols et al., 2018, 2023). Their wide absorption profiles, however, increase the noise in the clustering measurements. Following `dMdB20`, we mask the regions where more than 20% of flux is absorbed DLAs and correct their damping wings in the unmasked regions using a Voigt profile. We use a DLA catalogue created using two different techniques, a convolutional neural network (CNN) and a Gaussian process (GP) finder, used in previous Lyα analyses (Parks et al., 2018; Ho et al., 2021) and re-developed for DESI (Wang et al., 2022; J Zou. et al., 2025). We decide to mask only those DLAs that are detected by both algorithms with a confidence higher than 50%, resulting in 7% (or ~8500) of the forests in our analysis containing at least one DLA.

To further remove noise from our data, we perform two additional steps. The first is to mask areas affected by sharp spectral features, like the bright [OI] 5577.3 sky line and the calcium H and calcium K lines from the interstellar medium. We mask the wavelength ranges between [5570.5,5586.7]Å, [3967.3,3971.0]Å and [3933.0,3935.8]Å respectively to avoid these (Ramírez-Pérez et al., 2024). Other features caused by systematic flux calibration errors are corrected using the CIII] spectral region (between CIV and CIII] in figure 5.3), see Ramírez-Pérez et al., 2024 for more details.

5.2.3 The flux transmission field

Here we describe the measurement of the overdensity field used in our correlation functions. We only discuss the most important steps; for a complete description of these processes and the DESI EDR Lyα catalogue, see Ramírez-Pérez et al., 2024.

Fluctuations around the mean transmitted flux fraction, known as the flux transmission field, ($\delta_\mathrm{q}(\lambda)$) are computed as a function of observed wavelength:

$$\delta_\mathrm{q}(\lambda) = \frac{f_\mathrm{q}(\lambda)}{\overline{F}(z)C_\mathrm{q}(\lambda)} - 1 \, , \tag{5.1}$$

where $f_\mathrm{q}(\lambda)$ is the observed flux of quasar q, $\overline{F}(z)$ is the mean transmission of the IGM at redshift z ($z(\lambda) = \lambda/\lambda_\alpha - 1$) and $C_\mathrm{q}(\lambda)$ is the unabsorbed quasar continuum. The product of the two terms in the denominator is equivalent to the mean expected flux for quasar q at observed wavelength λ.

We compute the transmission field of each forest following the methodology of Ramírez-Pérez et al., 2024, where the denominator of equation 5.1 is approximated for each forest as the product of a global rest-frame quantity $\overline{C}(\lambda^\mathrm{rf})$ - the weighted stack of f for all quasars - and a polynomial in $\log\lambda$ to account for quasar continuum diversity. For each individual line-of-sight, the polynomial term has two parameters a_q and b_q, controlling the quasar continuum amplitude and shape respectively.

The flux variance (σ_q^2) in each pixel, later used to calculate the pixel weights (equation 5.3), is the sum of two terms:

$$\sigma_{\mathrm{q}}^2 = \left[\overline{F}C_{\mathrm{q}}(\lambda)\right]^2 \sigma_{\mathrm{LSS}}^2(\lambda) + \eta(\lambda)\,\sigma_{\mathrm{pip,q}}^2(\lambda)\,, \tag{5.2}$$

where $\sigma_{\mathrm{LSS}}^2(\lambda)$ is the contribution from cosmic variance, and $\sigma_{\mathrm{pip,q}}^2(\lambda)$ is the pipeline noise contribution, normalised with the expected flux. The function η further corrects for any inaccuracies in the pipeline estimation. In the `dMdB20` analysis, $\eta(\lambda)$ ranged from 1.05 and 1.2 between 3600 and 5500 Å respectively. Over the same range in observed wavelength of the EDR+M2 dataset, this correction ranges from 1 to 1.01 (Ramírez-Pérez et al., 2024), demonstrating that the DESI pipeline gives a better noise estimate.

We use the publicly available code `Picca`[4] to compute our flux transmission field, computing all of the relevant terms described above. While our computation of the flux transmission field largely mirrors that of `dMdB20`, there are some key changes introduced by Ramírez-Pérez et al., 2024 which help to increase the relative SNR of our correlation functions. We will describe these in the following section.

5.2.3.1 The Lyα catalogue

In the observed wavelength regime we restrict our forests between $\lambda^{\mathrm{min}} = 3600$ Å and $\lambda^{\mathrm{max}} = 5772$Å. The lower limit is motivated by the limit of the DESI spectrograph, which for our Lyα catalogue translates to only including quasars above z~2. We set the upper limit to maximise the signal-to-noise of our correlation functions. In the rest-frame, our wavelength range is also reduced from the full extent of the Lyα forest to between $\lambda^{\mathrm{rf,min}}$ =1040Å and $\lambda^{\mathrm{rf,max}}$ =1205Å, to avoid the wings of the Lyα and Lyβ emission lines, which can vary from quasar to quasar and thus add error to the continuum fitting. Compared to `dMdB20`, we have a higher upper wavelength limit in both the observed and rest-frame cases. In Ramírez-Pérez et al., 2024, they found that $\lambda^{\mathrm{rf,max}}$ =1205Å was the limit corresponding to the lowest possible variance in the Lyα correlation function monopole, achieving a balance between the number of Lyα forest pixels and avoiding the wings of the Lyα emission line.

The limits above result in 109 900 Lyα forests, with a maximum Lyα absorption pixel redshift of $z = 3.75$. In practice we also place an upper limit on the quasar redshift of $z = 3.75$ to be consistent with our mock catalogues (Herrera-Alcantar et al., 2025b), but this is done when computing the correlation functions and removes

---

[4] https://github.com/igmhub/picca/

only a negligible number of forests. For the Lyα-quasar cross-correlation, we use a tracer catalogue of quasars with a minimum redshift of $z > 1.77$, containing 147 899 quasars. This limit corresponds to a maximum possible separation of 300 h^{-1}Mpc in the cross-correlation. In practice, we only measure correlations up to 200 h^{-1}Mpc but scales up to 300 h^{-1}Mpc are needed to accurately estimate the distortion (section 5.4.4). We place final restrictions on our sample by rejecting forests with fewer than 50 valid Lyman$-\alpha$ forest pixels, and forests with a failed continuum fit (Ramírez-Pérez et al., 2024). The resulting dataset contains 88 509 Lyα forests.

5.3 Measuring the correlation functions

In this section we present the measurement of the correlation functions. We first describe the weights entering in the estimator of the correlation function in section 5.3.1, and discuss a small correction applied to the forest to model the distortion introduced by the continuum fitting (section 5.3.2). We then proceed to measure the autocorrelation of the Lyman$-\alpha$ forest (section 5.3.3) and its cross-correlation with quasars (section 5.3.4), as well as the uncertainty in these measurements (section 5.3.5).

5.3.1 Lyα weights

Because of varying quasar brightness and exposure time, the signal-to-noise in our Lyα forest is quite diverse. Therefore we weight the Lyα pixels when measuring correlations. These weights are used to compute the autocorrelation in equation 5.9 and the cross-correlation in equation 5.10. An optimal analysis would weight all **pairs** of pixels according to their inverse covariance matrix (Slosar et al., 2013; Font-Ribera et al., 2018). However, this makes the analysis significantly more complex, so we follow previous analyses of eBOSS data (`dMdB20`) and weight pixels individually.

Instead of directly using the inverse of the pixel variance (equation 5.2), we use the modified Lyα weights introduced in Ramírez-Pérez et al., 2024:

$$w_{\rm i} = \frac{[(1+z_i)/(1+z_{\rm fid})]^{\gamma_\alpha - 1}}{\sigma^2_{\rm LSS}(\lambda)\sigma^2_{\rm mod} + \eta(\lambda)\sigma^2_{\rm pip,q}/\left[\overline{F}C_q(\lambda)\right]^2}\,, \tag{5.3}$$

where $z_{\rm fid}$=2.25 (although this factor cancels in equation 5.9). The extra parameter $\sigma^2_{\rm mod}$ was introduced in Ramírez-Pérez et al., 2024 to modulate the relative importance of instrumental noise and intrinsic Lyα fluctuations in the weights. The value of $\sigma^2_{\mathrm{mod}} = 7.5$ was found to be optimal for this dataset (Ramírez-Pérez et al., 2024), and it reduces the uncertainties on the auto- and cross-correlation by 20% and 10% respectively.

Finally, the numerator of equation 5.3 takes into account the redshift evolution of the Lyα bias (γ_α) and of the growth factor, and up-weights higher redshift pixels to favour absorption with higher signal. Following `dMdB20`, we use a value of $\gamma_\alpha = 2.9$, motivated by the observation that the amplitude of the Lyα power spectrum with redshift evolves as $(1+z)^{3.8}$ (McDonald et al., 2006) . This correction ranges from $\simeq 0.85$ to 2.3 between $z = 2$ and 4.

5.3.2 Continuum distortion

A notable side effect of fitting $(a_{\mathrm{q}}, b_{\mathrm{q}})$ using data from entire forests is that the measured $\delta_q(\lambda)$ will be distorted. In other words, our measured $\delta(\lambda_i)$ in pixel i will be a linear combination of all true underlying $\delta(\lambda_j)$, such that $\delta'_i = \eta_{ij}\delta_j$. The solution to this, described in detail in Bautista et al., 2017; du Mas des Bourboux et al., 2017; Adame et al., 2025b, is to apply a linear projection operator to our measured δ' that explicitly removes the mean and slope of each quasar forest from the field:

$$\hat{\delta}_{\mathrm{i}} = \sum_{\mathrm{i}} \eta_{\mathrm{ij}} \delta_{\mathrm{j}} \, , \tag{5.4}$$

where the projection matrix η_{ij} is given by

$$\eta^{\mathrm{q}}_{\mathrm{ij}} = \delta^K_{\mathrm{ij}} - \frac{w_{\mathrm{j}}}{\sum\limits_{\mathrm{k}} w_{\mathrm{k}}} - \frac{w_{\mathrm{j}}\kappa_{\mathrm{i}}\kappa_{\mathrm{j}}}{\sum\limits_{\mathrm{k}} w_{\mathrm{k}}\kappa^2_{\mathrm{k}}} \qquad \kappa_{\mathrm{k}} = \log\lambda_{\mathrm{k}} - \overline{\log\lambda_{\mathrm{q}}} \, . \tag{5.5}$$

δ^K_{ij} is the Kronecker delta and weights w are given in equation 5.3. The reason for doing this is that the projected field $\tilde{\delta}$ will have the same distortion as our projected true field. Thus, we measure the correlations of our projected field $\tilde{\delta}$ and, as we will show in section 5.4.4, project the model correlation function in the same way.

The transformation in equation 5.4 means that the average δ_{q} at wavelength λ is biased towards 0. Following `dMdB20`, we subtract $\overline{\delta(\lambda)}$ when calculating the cross-correlation (section 5.3.4) such that:

$$\tilde{\delta}_{\mathrm{q}}(\lambda) = \hat{\delta}_{\mathrm{q}}(\lambda) - \overline{\delta(\lambda)} \, . \tag{5.6}$$

This correction guarantees that the cross-correlation (section 5.3.4) tends to 0 at large separations regardless of the quasar redshift distribution.

5.3.3 Autocorrelation

To measure the correlation functions, redshift and angular (Δz,$\Delta\theta$) separations are first transformed to co-moving separations along ($r_{\|}$) and perpendicular ($r_{\perp}$) to the line of sight.

We calculate ($r_{\|}, r_{\perp}$) as:

$$r_{\|} = [d_{\mathrm{c}}(z_{\mathrm{i}}) - d_{\mathrm{c}}(z_{\mathrm{j}})]\cos\left(\frac{\Delta\theta}{2}\right) , \tag{5.7}$$

and

$$r_{\perp} = [d_{\mathrm{m}}(z_{\mathrm{i}}) + d_{\mathrm{m}}(z_{\mathrm{j}})]\sin\left(\frac{\Delta\theta}{2}\right) , \tag{5.8}$$

where (i,j) represent pixel-pixel or pixel-quasar pairs and $\Delta\theta$ the angle between them. The redshift of a pixel i is calculated assuming Lyα absorption, such that $z_{\mathrm{i}} = \lambda/\lambda_{\alpha} - 1$, where $\lambda_{\alpha} = 1215.67$Å. The co-moving distance is given by $d_{\mathrm{c}} = c/100\int_0^z \frac{dz'}{E(z')}$ in units of h^{-1}Mpc, where $E(z) = \sqrt{\Omega_{\mathrm{m}}(1+z)^3 + \Omega_{\Lambda}}$. d_m is the transverse co-moving distance (also known as angular co-moving distance). In the case of a flat Universe (Ω_{k}=0), which we assume here, $d_{\mathrm{m}} = d_{\mathrm{c}}$.

For the autocorrelation measurement we use a weighted covariance estimator following previous analyses (Slosar et al., 2011; Bourboux et al., 2020):

$$\xi_{\mathrm{A}} = \frac{\sum_{\mathrm{i,j}\in\mathrm{A}} w_{\mathrm{i}}w_{\mathrm{j}}\delta_{\mathrm{i}}\delta_{\mathrm{j}}}{\sum_{\mathrm{i,j}\in\mathrm{A}} w_{\mathrm{i}}w_{\mathrm{j}}} , \tag{5.9}$$

where A is a bin in ($r_{\|}, r_{\perp}$) space with width 4 h^{-1}Mpc and the weights w_{i} are described in section 5.3.1. We sum over all pixels (i,j) across different lines of sight, but do not include pixels in the same line of sight because of continuum fitting errors that affect the entire forest, leading to spurious correlations which bias our measurement. In each bin A, our model correlation ξ^{mod} is evaluated at the weighted (by the number of pixel pairs) mean separation ($r_{\|}, r_{\perp}$) of the Lyα pixels (equation 5.3) in our data.

We measure the autocorrelation from $[0, 200]$ h^{-1}Mpc using 50 bins of 4 h^{-1}Mpc along and perpendicular to the line of sight, giving us 2500 bins in total. The given limits on separation mean that the Lyα autocorrelation has 3.75×10^{11} pixel pairs. In figure 5.4 we present our Lyα autocorrelation measurement as a function of (μ, r) where $\mu = r_{\|}/r$ and $r^2 = r_{\|}^2 + r_{\perp}^2$, for EDR+M2 and eBOSS DR16. We show the autocorrelation in 4 wedges, computed by averaging the 2D correlation in different selections of μ, ranging from closest ($\mu \in [0.95, 1]$) to furthest ($\mu \in [0, 0.5]$) from the line of sight. In general, our DESI EDR+M2 measurement is consistent with `dMdB20`, and across all bins ($r_{\|}, r_{\perp}$), the uncertainties are only ~1.9 times larger.

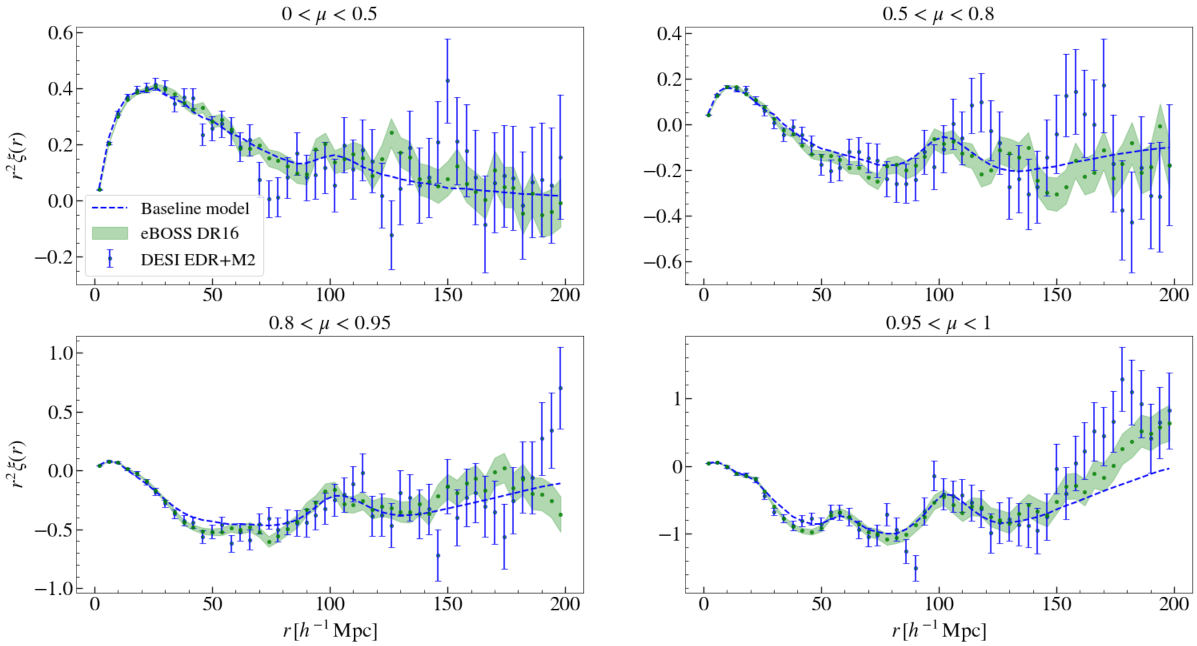

Figure 5.4: The DESI EDR+M2 (blue points) and eBOSS DR16 (Bourboux et al., 2020) (shaded green) Lyα autocorrelation compressed into weighted averages of $\mu = r_{\|}/r$, where $r = \sqrt{r_{\|}^2 + r_{\perp}^2}$. We also include the best-fitting model to EDR+M2 described in section 5.4 (blue dashed). We have multiplied the correlation by r^2 to visualise the BAO peak, which is visible in both data sets. From these plots, we can see the consistency between our measurements and those in eBOSS DR16 - a validation of the quality of DESI data at an early survey stage. Note that the presence of three other bumps in the line of sight plot (bottom right) at 20, 60 and 111 h^{-1}Mpc is due to correlations between the Lyα and SiIII(1207), SiII(1190)/SiII(1193) and SiII(1260) lines respectively. We model these contributions to the correlation in section 5.4.3.4.

It should be noted that there is overlap between the two datasets, since DESI is re-observing quasars observed in `dMdB20`. We present a constraint on the BAO peak amplitude from our best-fit model in section 5.5. For $\mu \in [0.95.1]$, we can also see peaks caused by other atomic transitions from elements like Si, referred to as metal absorption or contamination in the rest of the paper. For example, we see the SiIII(1207) peak at $r \sim 20\,h^{-1}$Mpc and the SiII(1190)/SiII(1193) peak at $r \sim 60\,h^{-1}$Mpc. At $r \sim 100\,h^{-1}$Mpc we observe the BAO peak, which appears more prominently in the $\mu \in [0.95, 1]$ plot, due to its overlap with the SiII(1260)xLyα peak at 111 h^{-1}Mpc. In section 5.4.3.4, we discuss the origin of the metal peaks and how we model them.

5.3.4 Cross-correlation

As with the autocorrelation, we define an estimator for the cross-correlation (Font-Ribera et al., 2012; Bourboux et al., 2020):

$$\xi_{\rm A} = \frac{\sum_{\rm i,j\in A} w_{\rm i} w_{\rm j} \delta_{\rm i}}{\sum_{\rm i,j\in A} w_{\rm i} w_{\rm j}} \, , \tag{5.10}$$

for a pixel of Lyα absorption i and quasar j. In this case, the weights $w_{\rm j}$ are corrected for the quasar bias evolution and are given by:

$$w_{\rm j} = \left(\frac{1 + z_{\rm j}}{1 + z_{\rm fid}} \right)^{\gamma_{\rm q}-1} \, , \tag{5.11}$$

where $\gamma_{\rm q} = 1.44 \pm 0.08$ (du Mas des Bourboux et al., 2019). The $(1 + z_{\rm fid})$ term, with $z_{\rm fid}$=2.25, cancels again in equation 5.10. The sum in equation 5.10 applies over all quasar-pixel pairs except pixels from their background quasar. Again, the correlation is binned in terms of line of sight and transverse separation $(r_{||}, r_{\perp})$, with $r_{\perp} \in [0, 200]\, h^{-1}$Mpc but now $r_{||} \in [-200, 200], h^{-1}$Mpc. The cross-correlation is not symmetric under permutation of the two tracers, primarily because peaks from metal contamination appear at either $r_{||} > 0$ or $r_{||} < 0$, and because of systematic redshift errors (equation 5.22) and because the bias of each tracer evolves differently with redshift. Therefore we choose to define -ve separations for the case when the quasar is behind the Lyα pixel ($z_{\rm q} > z_{\alpha}$) and likewise +ve separations when the quasar is between the observer and the Lyα pixel ($z_{\alpha} > z_{\rm q}$). This asymmetry allows us to study systematics like quasar redshift errors (Bault et al., 2025). The given range of separation in both directions and the $4\, h^{-1}$Mpc bin width now result in N = 100×50 = 5000 bins.

In figure 5.5 we again show the cross-correlation as a function of r and μ, this time averaging over $\mu \in [-1, 1]$, since the cross-correlation has -ve values of $r_{||}$. Here, the correlation is reversed with respect to the matter correlation function. This behaviour arises from the fact that the bias of the Lyα forest is negative. By convention we trace the flux-transmission field, which is higher in more under-dense regions - and the quasar bias is positive, giving a negative product. The uncertainties in our DESI EDR+M2 cross-correlation measurement across all $(r_{||}, r_{\perp})$ bins are ~1.7 times larger than in eBOSS DR16. In the following sections, we discuss the covariance of the measurements of both the auto- and cross-correlations.

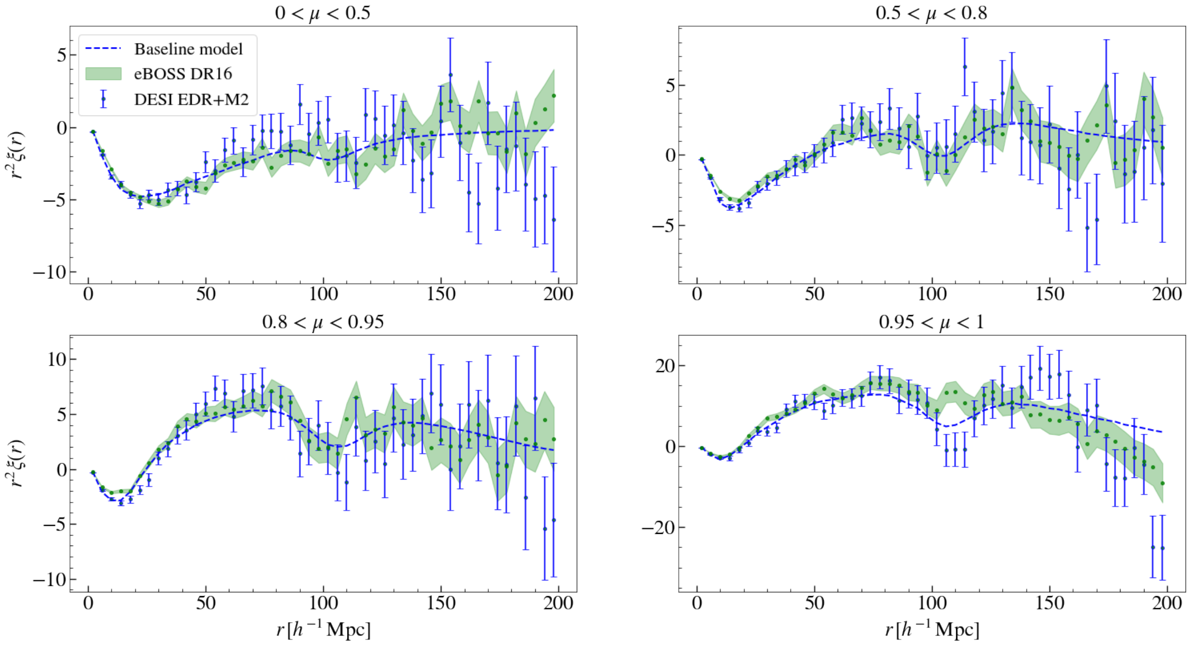

Figure 5.5: The 3D EDR+M2 (blue points) and eBOSS DR16 (Bourboux et al., 2020) (green shaded regions) Lyα-quasar cross-correlation, and the baseline fit (blue dashed) to EDR+M2 described in section 5.4. Because we have negative values of $r_{||}$ (when $z_\mathrm{q} > z_{Ly\alpha}$), we have negative values of $\mu = r_{||}/r$ and therefore average over $\mu \in [-1, 1]$. The cross-correlation is expectedly noisier than the autocorrelation, but still we see a good level of consistency between eBOSS and DESI at this early stage.

5.3.5 Covariance matrix

We estimate the covariances of the auto- and cross-correlations using a sub-sampling method, following Delubac et al., 2015. In this method, we define our sub-samples in 440 `HEALPix` (Gorski et al., 2005) pixels using `Healpy` (Zonca et al., 2019), equivalent to an *nside*=16, each of resolution 3.7 deg or solid angle 3.7×3.7 = 13.4 deg^2. We then calculate the weighted covariance C_AB between samples as:

$$C_\mathrm{AB} = \frac{1}{W_\mathrm{A} W_\mathrm{B}} \sum_{\mathrm{i,j}\in \mathrm{A}} \sum_{\mathrm{k,l}\in B} W^\mathrm{s}_\mathrm{A} W^\mathrm{s}_\mathrm{B} \left[\hat{\xi}^\mathrm{s}_\mathrm{A} \hat{\xi}^\mathrm{s}_\mathrm{B} - \hat{\xi}_\mathrm{A} \hat{\xi}_\mathrm{B}\right] \,, \tag{5.12}$$

where $\hat{\xi}^\mathrm{s}_\mathrm{A}$ is the measured correlation and W^s_A is the summed weight in sub-sample s, and $W_\mathrm{A} = \sum_\mathrm{s} W^\mathrm{s}_\mathrm{A}$. We assign each Ly$\alpha$ pixel pair, or quasar-Lyα pair, to a given `HEALPix` pixel based on the object with the highest right ascension. Our effective redshift, z_eff =2.376, is calculated by taking the average redshift of each bin of the correlation function weighted by the number of pixel pairs. The area of one

`HEALPix` pixel at this redshift is ~255×255 (h^{-1}Mpc)^2. The sub-sampling method of estimating covariance has also been validated in studies with mock catalogues in `dMdB20`.

Our auto- and cross-covariance matrices have N = 2500×2500 = 6.25 million and N = 5000×5000 = 25 million entries, with the diagonal entries dominating.

The normalised covariance, or correlation matrix, is defined as:

$$Corr_{\rm AB} = \frac{C_{\rm AB}}{\sqrt{Var_{\rm A} Var_{\rm B}}} \,, \tag{5.13}$$

where $Var_{\rm A}$=$C_{\rm AA}$ is the variance. The off-diagonal elements of equation 5.12 are noisy, and we need to smooth them to make our covariance matrices invertible. Following `dMdB20`, we model the off-diagonal elements of the correlation matrices as a function of line of sight and transverse separation, such that $Corr_{\rm AB}$ = $Corr(\Delta r_{\parallel},\Delta r_{\perp}) = Corr(|r_{\parallel}^{\rm A} - r_{\parallel}^{\rm B}|,|r_{\perp}^{\rm A} - r_{\perp}^{\rm B}|)$. In figure 7 of `dMdB20` we see that the correlation matrix in this model decreases rapidly with ($\Delta r_{\parallel}$,$\Delta r_{\perp}$).

5.4 Correlation function models

In this section, we discuss the theoretical model for our correlation functions, which includes the effect of all the major contaminants. For the most part, our model follows that in `dMdB20`, except the instrumental systematic effects model, which we adapt to account for the differences between the DESI and eBOSS instruments. The final fit, which we describe in section 5.5, consists of 13 free parameters shown in table 5.1.

5.4.1 Power spectra

For our correlation model, we start with an isotropic linear power-spectrum template, decomposed into a peak and a smooth component. We then add the Lyα/quasar Kaiser terms, some non-linear corrections, and contaminants, and convert the resulting anisotropic power spectrum into a correlation function. Finally, we multiply by the distortion matrix and recombine the peak and smooth components to obtain the final correlation model.

The auto- and cross-correlation functions are derived from the tracer biased power-spectrum, given by:

$$P_{\rm ij}(k,\mu_k,z) = b_{\rm i}(z)b_{\rm j}(z)(1+\beta_i\mu_{\rm k}^2)(1+\beta_{\rm j}\mu_{\rm k}^2)P_{\rm QL}(k,\mu_{\rm k},z)F_{\rm NL,ij}(k,\mu_{\rm k})G(k,\mu_{\rm k}) \,, \tag{5.14}$$

where the vector (k,μ_k) is defined such that $k^2 = k_{\parallel}^2 + k_{\perp}^2$ and $\mu_k = k_{\parallel}/k$, G is a term that accounts for binning on an $(r_{\parallel},r_{\perp})$ grid (equation 5.18) and $P_{\rm QL}$ is a quasi linear power spectrum that is multiplied by a non-linear correction term $F_{\mathrm{NL,ij}}(k,\mu_k)$ (section 5.4.2). The parameters b_i and β_i are the linear bias and RSD of tracer i.

To fit a unique function across each $(r_{||}, r_{\perp})$ bin where the effective redshift varies from that of the full dataset ($z_{\mathrm{eff}} = 2.376$), we need to evolve the Lyα and quasar biases using:

$$b_{\alpha}(z) = b_{\alpha}(z_{\mathrm{eff}})\left(\frac{1+z}{1+z_{\mathrm{eff}}}\right)^{\gamma_{\alpha}} , \tag{5.15}$$

$$b_{\mathrm{q}}(z) = b_{\mathrm{q}}(z_{\mathrm{eff}})\left(\frac{1+z}{1+z_{\mathrm{eff}}}\right)^{\gamma_{\mathrm{q}}} , \tag{5.16}$$

where γ_{α} =2.9 (McDonald et al., 2006) and γ_{q} =1.44 (du Mas des Bourboux et al., 2019).

For the autocorrelation (i=j) the only tracer is Lyα absorption and for the cross (i≠j) the tracers are Lyα absorption and quasars. Following past BAO analyses (`dMdB20`), we take the linear matter power spectrum from Planck 2018[5] (Planck Collaboration et al., 2020) and separate it into a smooth component and peak component with non-linear broadening corrections:

$$P_{\mathrm{QL}}(\mathbf{k},z) = P_{\mathrm{peak}}(k,z)\exp\left[-\frac{k_{||}^2\Sigma_{||}^2 + k_{\perp}^2\Sigma_{\perp}^2}{2}\right] + P_{\mathrm{smooth}}(k,z) , \tag{5.17}$$

where P_{peak} contains BAO and P_{smooth} does not. We make this decomposition using the side band technique from Kirkby et al., 2013 such that $P_{\mathrm{peak}} = P_{\mathrm{lin}} - P_{\mathrm{smooth}}$, and allows us to isolate the model for the BAO peak from the rest of the correlation and fit its position and amplitude. For the non-linear broadening correction of the BAO peak, we use $\Sigma_{||} = 6.36\,h^{-1}$Mpc and $\Sigma_{\perp} = 3.24\,h^{-1}$Mpc as in `dMdB20` (Eisenstein et al., 2007; Bourboux et al., 2020), assuming that our small difference in effective redshift has negligible effect.

The last term in equation 5.14, G, corrects for the fact that we bin correlations in a grid in $(r_{||}, r_{\perp})$ and is the product of the Fourier transform of two top-hat functions:

$$G(k_{||},k_{\perp}) = \mathrm{sinc}\left(\frac{k_{||}\Delta A_{||}}{2}\right)\mathrm{sinc}\left(\frac{k_{\perp}\Delta A_{\perp}}{2}\right) , \tag{5.18}$$

where $\Delta A_{||} = 4\,h^{-1}$Mpc and $\Delta A_{\perp} = 4\,h^{-1}$Mpc are the bin widths in the line-of-sight and transverse directions respectively. To compute equation 5.18, we make the approximation that our correlation is distributed homogeneously within each bin. In reality, it scales with $r_{\perp}$ in the transverse direction, but Bautista et al., 2017 have shown that the approximation made here produces an accurate correlation function.

---

[5] derived using `CAMB` https://github.com/cmbant/CAMB

5.4.2 Non-linear corrections

We multiply the quasi-linear power spectrum $P_{\mathrm{QL,ij}}(k, \mu_k)$ of equation 5.14 by a non-linear correction $F_{\mathrm{NL,ij}}(k, \mu_k)$. In the autocorrelation, the dominant effects encompassed by this correction are non-linear structure growth, peculiar motion, and broadening due to the thermal motion of particles, modelled in Arinyo-i-Prats et al., 2015.

Statistical redshift errors and galaxy peculiar velocities of quasars smear the cross-correlation along the line of sight $r_{\parallel}$, which following Percival and White, 2009 we model as Lorentzian damping:

$$F_{\mathrm{NL}}^{\mathrm{cross}}(k_{\parallel}) = \frac{1}{\sqrt{1 + (k_{\parallel}\sigma_{\mathrm{v}})^2}} \,, \tag{5.19}$$

where σ_{v} is one of the free parameters in our model, given in table 5.1. Statistical errors also affect the autocorrelation by adding random shifts to quasar spectral templates that in turn introduce systematic error in the mean continuum estimate $\overline{F}(z)C_q(\lambda)$. The impact of statistical redshift errors on BAO measurements from both the auto- and cross-correlation is described in the next chapter.

5.4.3 Modelling contaminants

Both the auto- and cross-correlation functions receive small contributions from other effects, which for the autocorrelation we model as:

$$\xi_{\mathrm{auto}} = \xi_{\mathrm{Ly}\alpha\times\mathrm{Ly}\alpha} + \sum_{\mathrm{m}} \xi_{\mathrm{Ly}\alpha\times\mathrm{m}} + \sum_{\mathrm{m_i,m_j}} \xi_{\mathrm{m_i}\times\mathrm{m_j}} + \xi_{\mathrm{inst}} \,, \tag{5.20}$$

where $\xi_{Ly\alpha\times Ly\alpha}$ is transformed from $P_{\mathrm{Ly}\alpha}(k, \mu_k)$ (equation 5.14) via a Fast Hankel transform. The Lyα bias parameter is also modelled to account for contamination by High Column Density Systems (HCD). $\xi_{Ly\alpha\times m}$ is the cross-correlation between $Ly\alpha$ absorption and metal absorption and $\xi_{m_i\times m_j}$ is the metal-metal autocorrelation discussed in section 5.4.3.4. Finally, ξ_{inst} is the contribution from DESI instrumental systematics, including sky subtraction residuals and spectrophotometric calibrations, described in full in Guy et al., 2025 and modelled in section 5.4.3.2. Likewise, for the cross-correlation, we write:

$$\xi_{\mathrm{cross}} = \xi_{\mathrm{Ly}\alpha\times\mathrm{q}} + \sum_{\mathrm{m}} \xi_{\mathrm{q}\times\mathrm{m}} + \xi^{\mathrm{TP}} \,, \tag{5.21}$$

where $\xi_{\mathrm{Ly}\alpha\times\mathrm{q}}$ is transformed from Fourier space in the same way as the autocorrelation, $\xi_{\mathrm{q}\times\mathrm{m}}$ is the quasar-metal line cross-correlation and ξ^{TP} is the contribution from the effect of quasar radiation given in section 5.4.3.1.

Finally we introduce another free parameter to our model to account for any systematic shift in line-of-sight separation between quasars and Lyα absorption:

$$\Delta \mathrm{r}_{\parallel,\mathrm{q}} = r_{\parallel,\mathrm{True}} - r_{\parallel,\mathrm{Measured}} \ . \tag{5.22}$$

This effect arises due to systematic quasar redshift errors and thus produces an asymmetry over positive and negative $\mathrm{r}_{\parallel}$ which can be studied using the cross-correlation of quasars and Lyα absorption, as is done in Bault et al., 2025.

5.4.3.1 Quasar radiation effects

At small scales, quasar radiation can strongly affect its surrounding gas or a nearby Lyα forest. This radiation dominates over the UV background (du Mas des Bourboux et al., 2017), so the ionisation fraction, and therefore the number of passing Lyα photons, increases. This is modelled following Font-Ribera et al., 2013; Bourboux et al., 2020, assuming isotropic quasar emission:

$$\xi^{\mathrm{TP}} = \xi_0^{\mathrm{TP}} \left(\frac{1\,h^{-1}\mathrm{Mpc}}{r} \right)^2 \exp\left(\frac{-r}{\lambda_{\mathrm{UV}}} \right) \ , \tag{5.23}$$

where we fit for the scale parameter ξ_0^{TP}, and λ_{UV} is the UV photon mean free path set to 300 h^{-1}Mpc following Rudie et al., 2013.

5.4.3.2 Instrumental systematics

The focal plane of DESI is divided in 10 "petals", with 500 fibres each that are fed to one of the 10 spectrographs. The calibration of DESI spectra (sky subtraction using sky fibres and flux calibration using standard stars) are performed separately for each of the 10 petals (Guy et al., 2023). This introduces a possible source of systematic correlations for pairs of spectra obtained with the same spectrograph. The impact of correlated sky residuals and flux calibration errors in the Lyα correlations in DESI is described in detail in Guy et al., 2025. Poisson fluctuations at the observed wavelength of the sky lines within each spectrograph, and flux calibration residuals, produce an excess correlation at $r_{\parallel} = 0$. We model the contributions of both of these, which contribute roughly equally, in the autocorrelation with an empirical functional form:

$$\xi_{\mathrm{inst}} = \begin{cases} \mathrm{A_{inst}} \left(\frac{r_\perp}{80} - 1 \right)^2 & r_\perp < 80\,h^{-1}\mathrm{Mpc},\ r_{\parallel} = 0 \\ 0 & r_\perp > 80\,h^{-1}\mathrm{Mpc} \end{cases} \tag{5.24}$$

where $\mathrm{A_{inst}}$ is fit and given in section 5.5. The limit of 80 h^{-1}Mpc roughly corresponds to the angular size of a DESI petal on the sky at the effective redshift

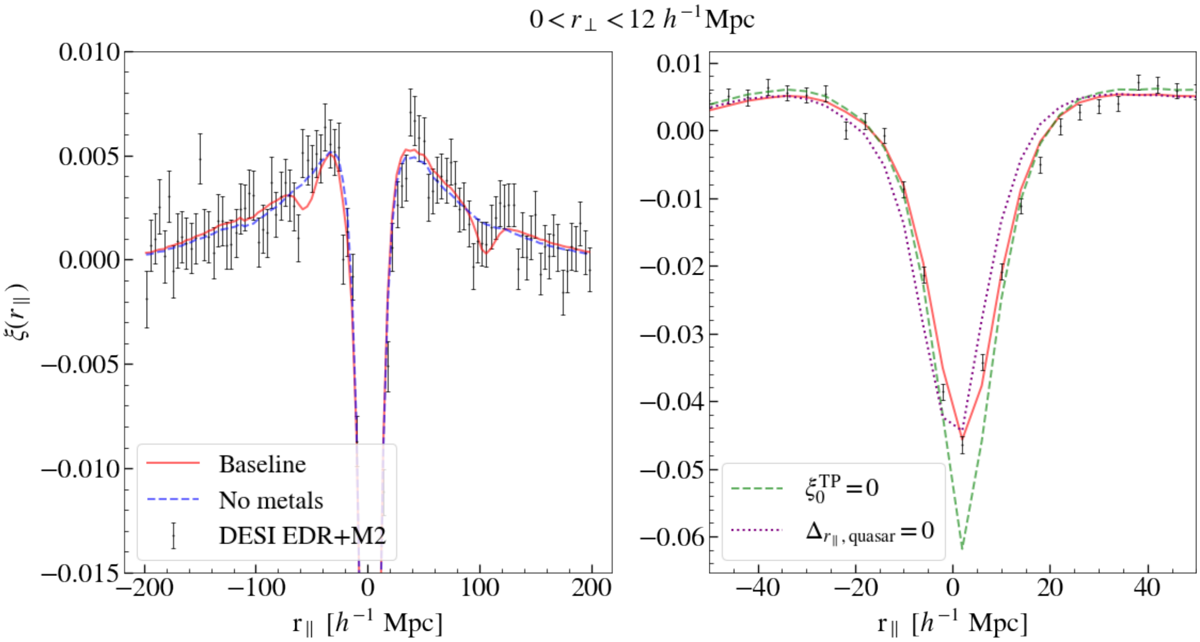

Figure 5.6: The Lyα-quasar cross-correlation as a function of $r_\parallel$, for $r_\perp \in [0, 12]\ h^{-1}$Mpc. On the right, we over-plot our baseline model (red solid) reported in table 5.1 and a model with no metal line component (blue dashed). On the left, we show two additional lines to highlight the effect of setting $\Delta r_{\parallel,q}$ to 0 and ignoring quasar radiation effects (green dashed). We can see the asymmetry of the cross-correlation through metals and systematic redshift errors (parametrised by $\Delta r_{\parallel,q}$) in the data here.

of our data. Note that we do not see this effect in the cross-correlation since it requires two correlating subtracted sky lines in two Lyα forests. In `dMdB20` the sky subtraction residuals were modelled with a Gaussian with two free parameters instead of one. The differences in the DESI and eBOSS instruments - such as eBOSS having only two spectrographs with 500 fibres each - lead to a difference in the shape and amplitude of the instrumental effects model. In section 5.5, we compare the amplitude of both models at $r_\parallel = r_\perp = 0$.

5.4.3.3 Absorption by high column density systems

The Lyα forest contains absorption from neutral hydrogen in the diffuse IGM and systems of neutral hydrogen with significantly higher column density[6]. In section 5.2.2, we defined DLAs as having column density $N_{HI} > 10^{20.3}\mathrm{cm}^{-2}$ and discussed how we identify and account for these using pixel masks and a Voigt profile to model the wings. Assuming we identify and treat all of the DLAs, the noise in our correlation measurements should be significantly reduced. We still expect to be contaminated by some objects of column density $N_{HI} < 10^{20.3}\mathrm{cm}^{-2}$

(sometimes referred to as sub-DLAs) which have a similar absorption profile, but lesser effect on our data. Ideally, we would also mask HCDs below the DLA limit, but these are difficult to identify with our current pipeline and resolution. Following Font-Ribera et al., 2012; Rogers et al., 2018 we model the contamination from these systems as a modification of the large-scale bias parameters:

$$\tilde{b}_\alpha = b_\alpha + b_{\text{HCD}} F_{\text{HCD}}(k_\parallel) \,, \tag{5.25}$$

$$\tilde{b}_\alpha \tilde{\beta}_\alpha = b_\alpha \beta_\alpha + b_{HCD} \beta_{\text{HCD}} F_{\text{HCD}}(k_\parallel) \,, \tag{5.26}$$

where b_{HCD} and β_{HCD} are the bias and RSD parameter of the HCD systems and $F_{\text{HCD}}(k_\parallel)$ is a factor depending on the distribution of HCDs, modelled following `dMdB20` as:

$$F_{\text{HCD}}(k_\parallel) = \exp(-L_{\text{HCD}} k_\parallel) \,. \tag{5.27}$$

Here L_{HCD} is the typical HCD length scale, and is set to $10\,h^{-1}$Mpc following `dMdB20` because of its degeneracies with other parameters in the fit. Note that it was verified in Cuceu et al., 2020, 2022b that varying L_{HCD} between ~7-13 h^{-1}Mpc does not significantly affect BAO measurements.

5.4.3.4 Absorption by metal lines

As stated earlier and shown in equations 5.20 and 5.21, our measured correlations are not only the result of Lyα absorption in the IGM or HCDs but also heavier elements in the IGM due to galactic gas outflows (Yang et al., 2022). Therefore, we introduce a cross power spectrum $\text{P}_{\text{mn}}(\mathbf{k}, z)$ for each pair of absorbers (m,n) and their configuration-space counterparts $\xi^{\text{m}\times\text{n}}$.

Because we measure our flux-transmission field and correlations assuming all absorption in the Lyα is due to the Lyα transition, we assign to each absorption pixel a redshift $z = \lambda/\lambda_\alpha - 1$. However, because of the metal lines from heavier elements, an excess correlation appears at the separation between the true and assigned redshifts z_{metal} and z_α:

$$r_\parallel = d_{\text{metal}} - d_{\text{Ly}\alpha} = c \int_{z_\alpha}^{z_{\text{metal}}} \frac{dz'}{H(\bar{z})} \approx \frac{c(1+z)}{H(z)} \frac{\lambda_{\text{metal}} - \lambda_\alpha}{\bar{\lambda}} \,, \tag{5.28}$$

where $\bar{\lambda}$ and $\bar{z}$ are the mean values of the Lyα and metal absorption. The excess correlation is more clearly observed for vanishing physical separations between two absorbers, when $r_\perp \sim 0$. We represent the correlation shift for each pair of

---

[6] *High column density systems* is a term used in past eBOSS publications to refer to neutral hydrogen with column density $N_{HI} > 10^{17.2}\text{cm}^{-2}$, of which DLAs are the highest column density subset and the most impactful contaminants on our data.

absorbers (m,n) using the metal matrix formalism of Blomqvist et al., 2018. In this method, the shifted correlation function $\xi_1^{\mathrm{m}\times\mathrm{n}}$ for each absorber pair (m,n) is given in terms of the un-shifted correlation function $\xi_0^{\mathrm{m}\times\mathrm{n}}$ as:

$$\xi_1^{\mathrm{m}\times\mathrm{n}}(\mathrm{A}) = \sum_{\mathrm{B}} M_{\mathrm{AB}}, \xi_0^{\mathrm{m}\times\mathrm{n}}(r_{\parallel}(\mathrm{B}), r_{\perp}(\mathrm{B})), \tag{5.29}$$

where (m,n)∈B is the pixel separation using the rest wavelength of each absorber m and n, (m,n)∈A refers to the pixel separation assuming both pixels have rest wavelength $\lambda_{\mathrm{Ly}\alpha}$, and:

$$M_{\mathrm{AB}} = \frac{\sum_{(\mathrm{m,n})\in\mathrm{A},(\mathrm{m,n})\in\mathrm{B}} w_{\mathrm{m}} w_{\mathrm{n}}}{\sum_{(\mathrm{m,n})\in\mathrm{A}} w_{\mathrm{m}} w_{\mathrm{n}}} \,. \tag{5.30}$$

For the cross-correlation, index m is a quasar with redshift z_{q} used to calculate the projected separation. To compute M_{AB} in a reasonable time, we only use 1% of pixel pairs. We have verified that this does not affect the measured metal biases.

Different rest-frame wavelengths peaks at different line of sight separations. Given the range of separations that we measure, we model 4 dominant metal lines - SiII(1190), SiII(1193), SiIII(1207), and SiII(1260) (Bautista et al., 2017). Each metal line has it's own bias and RSD parameters ($b_{\mathrm{metal}}, \beta_{\mathrm{metal}}$) and corresponding power spectrum P_{mn} as mentioned above. We cannot determine these two parameters separately since the correlations $\xi_1^{\mathrm{m}\times\mathrm{n}}$ only have a significant impact at small $r_{\perp}$, and therefore we set β_{m} =0.5 following Bautista et al., 2017. The values of b_{m} for each metal line are constrained in the fits discussed in section 5.5. The CIV metal bias has also been typically fit in previous Lyα forest correlation studies (Blomqvist et al., 2018; Gontcho A Gontcho et al., 2018), but we decide not to model it in this study because it was weakly detected and had a negligible effect on the `dMdB20` analysis, and we report the same for EDR+M2 data.

5.4.4 Modelling the continuum distortion

Due to distortions created in the quasar continuum fitting process, we apply two corrections to our measured delta fields in section 5.2.3. Doing this, each measured delta $\tilde{\delta}_{\mathrm{q}}(\lambda)$ becomes a linear combination of all other pixels in the same forest, distorting the correlation functions measured from these deltas. To correct for the distortion we relate the measured ($\hat{\xi}$) and true (ξ^{t}) correlations following Bautista et al., 2017; du Mas des Bourboux et al., 2017; Bourboux et al., 2020 with:

$$\hat{\xi}_{\mathrm{A}} = \sum_{\mathrm{A}'} D_{\mathrm{AA}'} \xi^{\mathrm{t}}_{\mathrm{A}'} \,, \tag{5.31}$$

where

$$D_{\mathrm{AA'},\mathrm{auto}} = \frac{\sum\limits_{(\mathrm{i},\mathrm{j})\in \mathrm{A}} w_\mathrm{i} w_\mathrm{j} \left(\sum\limits_{(\mathrm{i'},\mathrm{j'}\in \mathrm{A'})} \eta_{\mathrm{ii'}} \eta_{\mathrm{jj'}} \right)}{\sum\limits_{(\mathrm{i},\mathrm{j})\in \mathrm{A}} w_\mathrm{i} w_\mathrm{j}} , \tag{5.32}$$

for the autocorrelation, where η is the projection matrix given in equation 5.5. We also have

$$D_{\mathrm{AA'},\mathrm{cross}} = \frac{\sum\limits_{(\mathrm{i},\mathrm{k})\in \mathrm{A}} w_\mathrm{i} \sum\limits_{(\mathrm{j},\mathrm{k})\in \mathrm{A'}} \eta_{\mathrm{ij}}}{\sum\limits_{(\mathrm{i},\mathrm{k})\in \mathrm{A}} w_\mathrm{i}} , \tag{5.33}$$

for the cross-correlation, where i, j are pixels in the same forest and k is a quasar. Like the metal matrices, we compute the distortion matrices with only 1% of total pixel pairs to reduce computing hours. In our analysis, we model the effect of distortion out to a separation of 300 h^{-1}Mpc, since distortions at this scale are known to affect the correlation function at scales relevant to our analysis. The two expressions above equate to multiplying our physical model of the correlation by the distortion matrix before any comparison with data. The effects of distortion have also been validated in previous studies using mocks (Bautista et al., 2017).

5.5 Fits to the data

We now present the best-fitting solution of the model introduced in section 5.4 to the Lyα autocorrelation and Lyα quasar cross-correlation described in section 5.3. For this study, we have established a baseline model with 13 free parameters based on Kaiser models for biased tracers with additional contaminants, described in detail in section 5.4.

In the anisotropic correlation model of equation 5.14, it is evident that the cross-correlation we measure is only sensitive to the product of the Lyα and quasar biases. As a result, we present only the auto and auto+cross combined fits. In the latter, we break the bias degeneracy to measure the quasar bias $\mathrm{b_q}$ distinctly. Then, using the fact that for point tracers β_q=f/b_q, where $f = 0.97$ (Planck Collaboration et al., 2020) is the growth-rate for our fiducial cosmology at $z = 2.376$, we infer a value of β_q. The other advantage of combining both correlations is improving constraints on key parameters. For example, our constraints on $\mathrm{A_{BAO}}$ improve by 1.6x going from autocorrelation fit alone to the combined fit.

In table 5.1, mean values and 68% credible regions of the posterior distributions for each free parameter in the auto and combined auto+cross correlations are given[7].

---

[7] computed using the plotting tool `Getdist` (Lewis, 2019)

To perform each fit, we use the Python package `Vega`[8], which employs the nested sampler `Polychord` (Handley et al., 2015) to sample each parameter. We find that the auto-only results are very consistent with the auto+cross combined. In each column we put a dash symbol in entries where parameters are only relevant to the combined correlation and not to the autocorrelation alone.

The second part of table 5.1 gives the characteristics of each fit, including the number of bins in each measurement, the number of free parameters, the minimum χ^2 value, and the probability (p-value). Note that to obtain the latter two values, we use a minimiser instead of the nested sampler used to obtain parameter means and confidence regions. The minimiser makes use of the Python package `Iminuit`[9] We perform the fit for $r \in [10, 180]\, h^{-1}$Mpc due to the increasing complexity of modelling the correlation function at small scales (non-linearities). The $4\, h^{-1}$Mpc bin width of the correlation functions translates to 1590 and 4770 bins for the auto and combined fits.

To measure the full correlation function we re-combine the peak and smooth parts of the correlation (outlined in section 5.4.1):

$$\xi(r_{\parallel}, r_{\perp}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\mathrm{smooth}}(r_{\parallel}, r_{\perp}) + \mathrm{A_{BAO}}\xi_{\mathrm{peak}}(\alpha_{\parallel} r_{\parallel}, \alpha_{\perp} r_{\perp}) \,, \tag{5.34}$$

where $\alpha_{\parallel}$ and $\alpha_{\perp}$ are the BAO parameters (see `dMdB20` for recent constraints on these). Here $\mathrm{A_{BAO}}$ is the amplitude of BAO, which we detect at 3.8σ confidence (table 5.1). We fix both $\alpha_{\parallel}$ and $\alpha_{\perp}$ to 1, or the fiducial cosmology, and focus on preparing for a robust measurement with DESI year 1 data.

The best-fitting model is shown on four autocorrelation wedges in figure 5.4 and the same four cross-correlation wedges in figure 5.5. In figure 5.7, we also show the autocorrelation along the line of sight, averaged over $0 < r_{\perp} < 12\, h^{-1}$Mpc and overlaid with the baseline model from the minimiser fit. We also include a fit that does not model the effect of metal contamination, and a fit that does not model metal contamination or BAO. We multiply by r to see that the bumps caused by the metal lines SiIII(1207) and SiII(1190)/SiII(1193) at $r \sim 20\, h^{-1}$Mpc and $r \sim 60\, h^{-1}$Mpc respectively are well-fit by our model, and poorly-fit by the no metal model. The SiII(1260) line gives a produces a peak at $r \sim 111\, h^{-1}$Mpc, which along the line of sight dominates over the BAO peak with which it overlaps.

5.5.1 Fit probability

The combined fit presented in table 5.1 has a χ^2 probability of 0.03. Because this is a relatively low value, we discuss several caveats here.

---

[8] https://github.com/andreicuceu/vega

[9] https://github.com/scikit-hep/iminuit

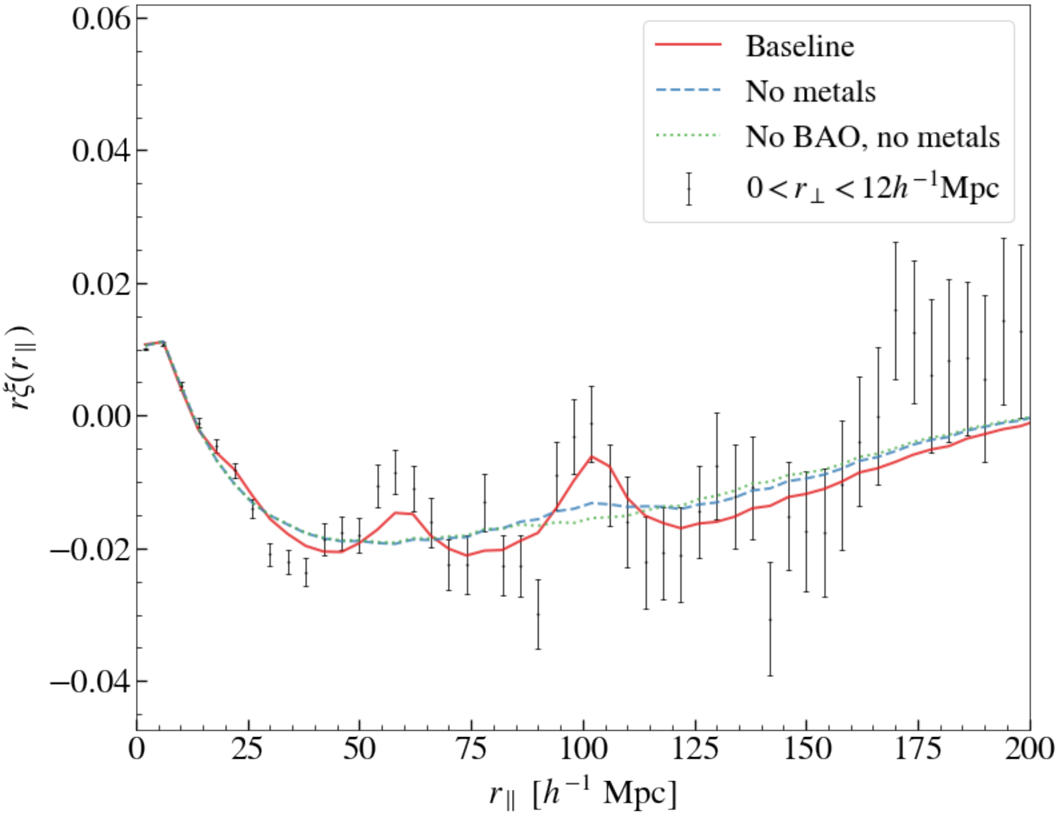

Figure 5.7: The best fitting solution (red solid) of our model to the DESI EDR+M2 Lyα autocorrelation (black points) function along the line of sight, where we average over 0< $r_\perp$ < 12 h^{-1}Mpc. We also include a fit which does not model metal contamination (blue dashed) and a fit which does not model metal contamination or BAO (green dotted). We multiply the y-axis by r for visualisation. The bumps in correlation caused by SiIII(1207) and SiII(1190)/SiII(1193) at $r_\parallel \sim 20\,h^{-1}$Mpc and $r_\parallel \sim 60\,h^{-1}$Mpc are visible. The most prominent peak is driven by SiII(1260) at $r_\parallel \sim 111\,h^{-1}$Mpc, rather than the overlapping BAO peak at $r_\parallel = 100\,h^{-1}$Mpc.

Parameter	Auto	Combined
b_α	-0.129±0.010	-0.134±0.009
β_α	$1.49^{+0.14}_{-0.19}$	$1.41^{+0.12}_{-0.15}$
b_{HCD}	-0.045±0.009	-0.039±0.009
$10^3 b_{\eta,\mathrm{SiII(1190)}}$	-3.3±1.0	-2.2±0.8
$10^3 b_{\eta,\mathrm{SiII(1193)}}$	$-1.7^{+1.1}_{-0.8}$	$-0.9^{+0.8}_{-0.3}$
$10^3 b_{\eta,\mathrm{SiII(1260)}}$	-2.9±1.1	-2.6±0.9
$10^3 b_{\eta,\mathrm{SiIII(1207)}}$	-3.9±1.0	-3.4±0.9
b_{q}	-	3.41±0.16
$\Delta r_{\parallel,\mathrm{q}}$ (h^{-1}Mpc)	-	-2.21±0.18
σ_{v} (h^{-1}Mpc)	-	5.2±0.5
ξ_0^{TP}	-	0.68±0.18
$10^4 A_{\mathrm{inst}}$	$2.6^{+0.4}_{-0.8}$	$2.4^{+0.3}_{-0.5}$
$\mathrm{A_{BAO}}$	1.04±0.45	1.21±0.32
$\mathrm{N_{bin}}$	1590	4770
$\mathrm{N_{params}}$	9	13
χ^2_{min}	1711	4945
probability	0.01	0.03

Table 5.1: Table of early DESI fits on the autocorrelation and the combined auto+cross correlations. We do not perform a cross-correlation-only fit due to degeneracies between parameters. In the first part we show the mean and 68% credible region for each free parameter in our model computed from the posterior distributions from the nested sampler. We evaluate each parameter at the effective redshift of our data z_{eff} =2.376, where redshift is relevant. Where a parameter does not feature in the autocorrelation fit, we use the "-" symbol. In the second part, we give the characteristics of the fit, including the number of bins, free parameters, and the χ^2_{min} probability as estimated with a minimiser.

In figure 5.8, the residuals are spread evenly across ($r_\parallel$,$r_\perp$). There are some groups of bins with higher residuals closer to the line-of-sight ($r_\perp = 0\,h^{-1}$Mpc) where continuum distortions are strongest, and at larger separations $r \sim 150\,h^{-1}$Mpc, but these are likely due to the high degree of correlation between data points which we ignore in this plot (we use the diagonal of the covariance matrix here).

If we change the lower limit on our fit of both correlations to $r = 20\,h^{-1}$Mpc rather than $10\,h^{-1}$Mpc, our fit probability increases to 0.08 from 0.03. It's possible that our constraints on the baseline model are slightly biased when fitting down to $10\,h^{-1}$Mpc, but we will leave this mock catalogues studies in DESI year 1. Recent results using Lyα mocks on `AbacusSummit` (Hadzhiyska et al., 2023) also suggest that the Kaiser model that we employ here does not adequately model non-linear effects in the Lyα quasar cross-correlation for $r < 30h^{-1}$Mpc. If we set this as a lower limit on the cross-correlation while keeping $10\,h^{-1}$Mpc for the autocorrelation (Hadzhiyska et al., 2023 find the Kaiser model is a good approximation in this case), we increase our probability to 0.05, without biasing our measurements.

Changing the upper limit to $r = 150\,h^{-1}$Mpc also increases the fit probability slightly to 0.05, and has a negligible effect on measured parameter values. For all of these tests, the most important thing for future analyses will be the effect on the BAO scale parameters, which we do not measure here. Therefore, we keep the baseline analysis between r$\in [10, 180]\,h^{-1}$Mpc as in `dMdB20`.

In section 5.2 we outlined how we compute the Lyα flux transmission catalogue using the full resolution of the DESI spectrograph (0.8Å). Using the optimal weighting of equation 5.3, by setting σ^2_{mod}=7.5, we reach the same precision on the correlation function as using 2.4Å pixels with σ^2_{mod}=3. However, when computing the fits from the latter weighting, we find a χ^2 probability of p=0.13 for the combined fit. Despite this higher fit probability, we prefer to use the full resolution of the DESI spectrograph in this analysis and leave this effect for study in the DESI year 1 analysis. Reducing the maximum rest-frame wavelength of the Lyα forest to 1205 to 1200Å (chosen to minimise correlation function uncertainties; Ramírez-Pérez et al., 2024) increases the fit probability to p=0.18, but we refrain from making this change until there is a further study on mock data sets with greater statistical power. Several sources of contamination like DLAs, metals, and other effects such as quasar radiation are modelled in the same way as in `dMdB20` and require an in-depth study for DESI.

5.5.2 Comparison with eBOSS DR16

To compare with `dMdB20`, we fit their measured correlation functions with the settings described in this paper. Except for the instrumental systematics effect, every other part of the model is the same. In figure 5.9 we present the marginalized 2D posterior of β_α and b_α. Note that the effective redshift of the eBOSS combined correlation is z=2.334, and thus we use the Lyα bias-redshift relation in equation 5.15 to shift our bias measurement to that effective redshift.

We summarise the differences in the mean value, 68% confidence intervals, and the consistency between b_α, β_α and a selection of parameters between DESI EDR+M2 and the eBOSS DR16 fits in table 5.2. In the case of b_{q}, we also shift the DESI EDR+M2 measurement to the effective redshift of the DR16 data set, this

---

[9] The Lyα working group key paper on DESI year 1 data will be centred around measuring ($\alpha_\parallel$,$\alpha_\perp$), but studies of the Alcock-Paczyński (AP) effect using both the smooth and peak part of the Lyα forest correlations will happen.

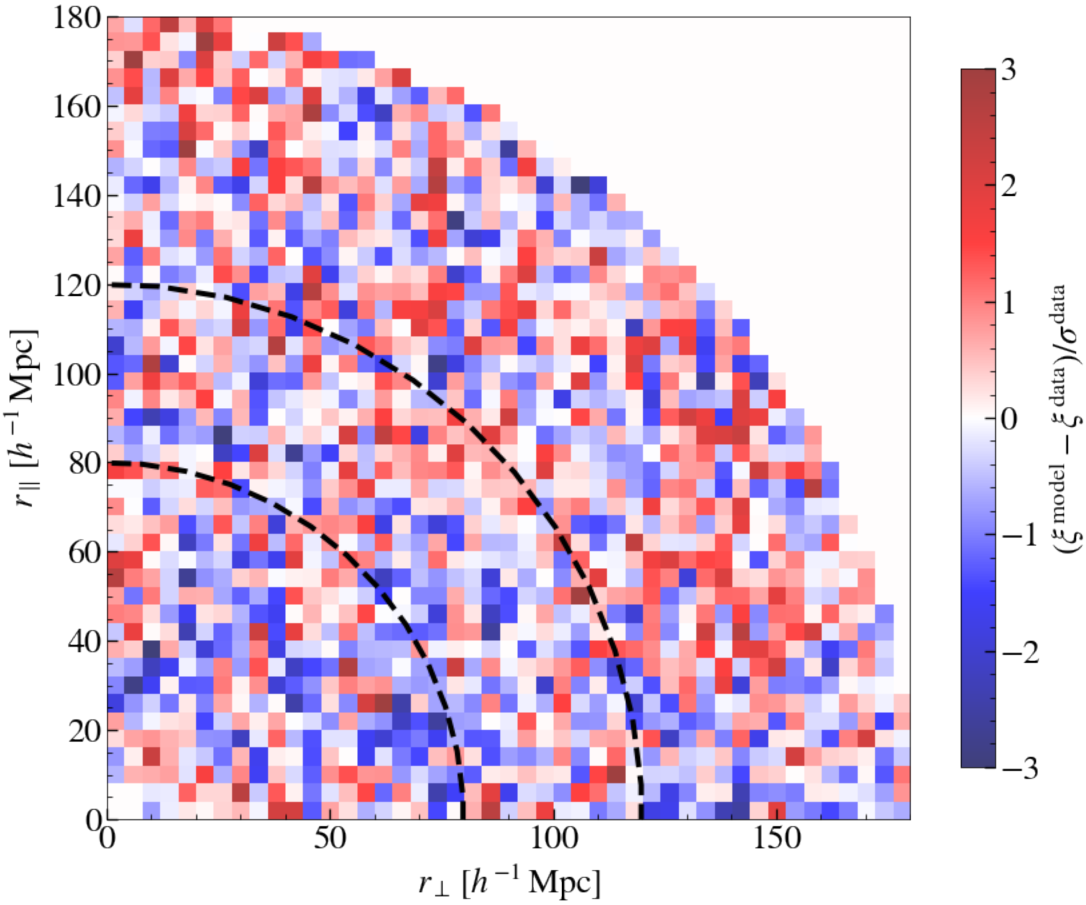

Figure 5.8: Residuals of our fit to the DESI EDR+M2 autocorrelation, defined as the difference between data and model correlation divided by the uncertainties of each data point (square root of the diagonal of the covariance matrix). The area between the black dashed lines indicates the BAO region between 80-120 h^{-1}Mpc. We see that overall there are no strong areas of discrepancy.

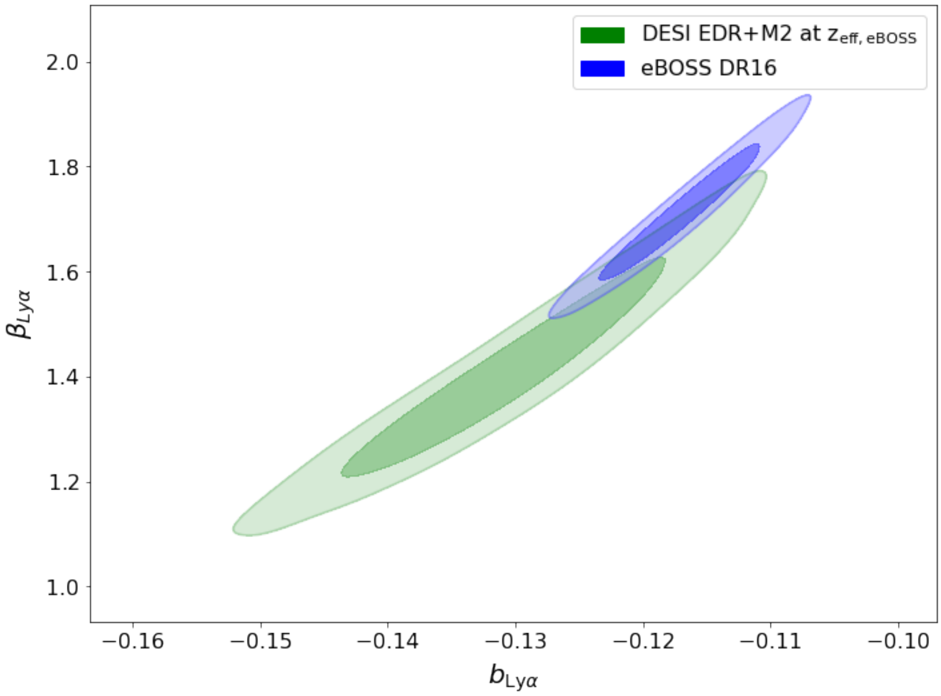

Figure 5.9: 68% and 95% credible regions on bias and redshift-space distortions from the auto- and cross-correlation of the Lyα forest for DESI EDR+M2 data at the effective redshift of eBOSS DR16 (green) and eBOSS DR16 data. We performed fits on both datasets with the same model (presented in section 5.4). To evolve the value of the DESI EDR+M2 bias with redshift we use the relation $b_{\mathrm{Ly}\alpha} \propto (1+z)^{\gamma-1}$ (McDonald et al., 2006) where γ =2.9 (Bourboux et al., 2020).

time using the relation $b_{\mathrm{q}}(\mathrm{z}) \propto (1 + \mathrm{z})^{\gamma}$ where $\gamma \sim 1.44$ (du Mas des Bourboux et al., 2019). For the other two parameters in table 5.2, we are only interested in the mean value and the relative values of the uncertainties between surveys. The σ_{v} constraint tells us that the DESI pipeline more precisely estimates quasar redshifts than was done by eBOSS DR16.

The final tracers that we constrain in our model are HCDs and metals. We measure the bias of HCDs to be b_{HCD} =-0.039±0.009 and b_{HCD} =-0.053±0.0043 in early DESI and eBOSS (data) respectively. A smaller value of b_{HCD} indicates weaker contamination by DLAs in our data, from which we conclude that the DESI pipeline (J Zou. et al., 2025) masks a higher proportion of DLAs, than the `dMdB20` pipeline. The metals are all detected with high confidence, each with a non-zero bias to a minimum of 2.25σ except SiII(1193), and are in good agreement with `dMdB20`. The effect of metal correlations is visible in figure 5.7.

The systematic error in the quasar redshift estimations is parameterised by $\Delta \mathrm{r}_{\|,\mathrm{q}}$, which is measured through an asymmetry in the cross-correlation as described in the previous section. We report a value of $\Delta \mathrm{r}_{\|,\mathrm{q}} = -2.21 \pm 0.19\, h^{-1}$Mpc, a

Parameter	DESI EDR+M2	eBOSS DR16	$\sigma_{\mathrm{DESI}}/\sigma_{\mathrm{eBOSS}}$	$\Delta\sigma$
$b_{\mathrm{Ly}\alpha}$ (z=$\mathrm{z}_{\mathrm{eff,DR16}}$)	-0.131±0.009	-0.117±0.004	2.25	1.39
$\beta_{\mathrm{Ly}\alpha}$	$1.41^{+0.12}_{-0.15}$	1.71±0.09	1.5	1.88
b_{q} (z=$\mathrm{z}_{\mathrm{eff,DR16}}$)	3.36±0.16	3.70±0.10	1.6	1.8
σ_{v} [h^{-1}Mpc]	5.2±0.5	6.77±0.29	1.72	-
$\mathrm{A}_{\mathrm{BAO}}$	1.21±0.32	1.16±0.16	2	-

Table 5.2: Table highlighting the main sampled posteriors of the auto+cross correlations for DESI EDR+M2 and eBOSS DR16 data, their relative uncertainties (68% credible region), and their consistency with each other, defined as $\Delta\sigma = |\mu_{\mathrm{eBOSS}} - \mu_{\mathrm{DESI}}| / \sqrt{\sigma^2_{\mathrm{eBOSS}} + \sigma^2_{\mathrm{DESI}}}$, where μ is the mean parameter value. We use the same model (outlined in section 5.4) for the fits in both cases, except for instrumental systematics contribution that is slightly different because of variations in the eBOSS and DESI instruments (Guy et al., 2025).

clear detection of systematic redshift errors. This offset could be related to an inconsistency in the treatment of the mean flux transmission $\overline{F}(z)$ in the quasar templates used by `Redrock` at $z \gtrsim 2$. In order to avoid this type of bias, `dMdB20` used a redshift estimator which did not use wavelengths around the Lyα emission line or lower. They measured a value of $\Delta \mathrm{r}_{\parallel,\mathrm{q}} = 0.10 \pm 0.11\, h^{-1}$Mpc, consistent with zero. The impact of redshift errors in DESI quasars is studied in more detail in Bault et al., 2025, and mitigation strategies are currently being considered in DESI.

At $r_\perp = r_\parallel = 0\, h^{-1}$Mpc, the spurious correlation produced by instrumental effects is $\mathrm{A}_{\mathrm{inst}} = 2.4^{+0.3}_{-0.5} \times 10^4$. In `dMdB20`, they used a Gaussian instrumental effects model, in which the spurious correlation at $r_\perp = r_\parallel = 0\, h^{-1}$Mpc is $\mathrm{A}_{\mathrm{sky}}/\sigma_{\mathrm{sky}}\sqrt{2\pi} = 1 \times 10^4$. We expect to see discrepancies in amplitudes here, given how distinct the two instruments are (section 5.4). To summarise the differences between early DESI and eBOSS DR16 data discussed above, we include a 13-parameter corner plot for both data sets in Appendix A, figure 5.10.

5.6 Conclusion

We present the first study of the Lyα autocorrelation and its cross-correlation with quasars from Dark Energy Spectroscopic Instrument (DESI) data. Our data sample, EDR+M2, consists of 318 691 quasar target spectra from the DESI survey validation (DESI Collaboration et al., 2023b) phase and the first two months of the DESI main survey, resulting in 88 509 Lyα forests with redshift $z > 2$.

We use the catalogue of Lyα fluctuations from Ramírez-Pérez et al., 2024, which made several adaptations to previous analysis pipelines (Bourboux et al., 2020), leading to a 20% and 10% improvement in the auto- and cross-correlations. Our 2-point correlation functions are estimated following the eBOSS DR16 analysis (Bourboux et al., 2020), with which we report high consistency and an average of 1.7× larger uncertainties across all bins in co-moving separation ($r_{\parallel}$,$r_{\perp}$). We do a combined auto+cross correlation fit with our 13 free parameter model, based on linear perturbation theory, and make constraints on key parameters like the Lyα forest bias, quasar bias, and the amplitude of the BAO peak. We detect Baryon Acoustic Oscillations to 3.8σ confidence, showing the constraining power of DESI data at an early survey stage. The model used is again similar to that in Bourboux et al., 2020, except the instrumental systematics model (Guy et al., 2025) that is modified to account for the differences between both instruments.

At this stage of the DESI survey, we have highlighted the quality of our data and validated existing analysis methods of 3D correlations using the Lyα forest. With relatively low statistical power, we do not report constraints on the scale parameters of BAO and instead leave that to the DESI year 1 analysis. This future work will require an in-depth study of systematics and contaminants like DLAs and metals before performing a proper cosmological analysis of BAO.

5.7 Appendix: complete fit results

Here we show the triangle plot of the full posterior distribution - made using the Polychord nested sampler (Handley et al., 2015) and the plotting tool `Getdist` (Lewis, 2019) - of all 13 parameters in our model of the combined auto+cross correlations, for both the DESI EDR+M2 and eBOSS DR16 data set. We measure both b_{α} and b_{q} at the effective redshift of eBOSS DR16 by using the redshift evolution of each parameter on the EDR+M2 result. For the Lyα and quasars this is $b \propto (1+z)^{\gamma}$, where $\gamma_{\alpha} = 2.9$ (McDonald et al., 2006; Bourboux et al., 2020) and $\gamma_{\mathrm{quasar}} = 1.44$ (du Mas des Bourboux et al., 2019).

We can see several parameters in the model which are correlated. The bias and RSD parameters of the Lyα forest are strongly correlated with each other and with the bias factor of HCDs. The correlations that we show between the quasar bias and the bias of HCDs, and the quasar radiation strength are why we do not fit the cross-correlation independently of the autocorrelation, as mentioned in section 5.5.

Finally, there is an apparent correlation between the instrumental systematic

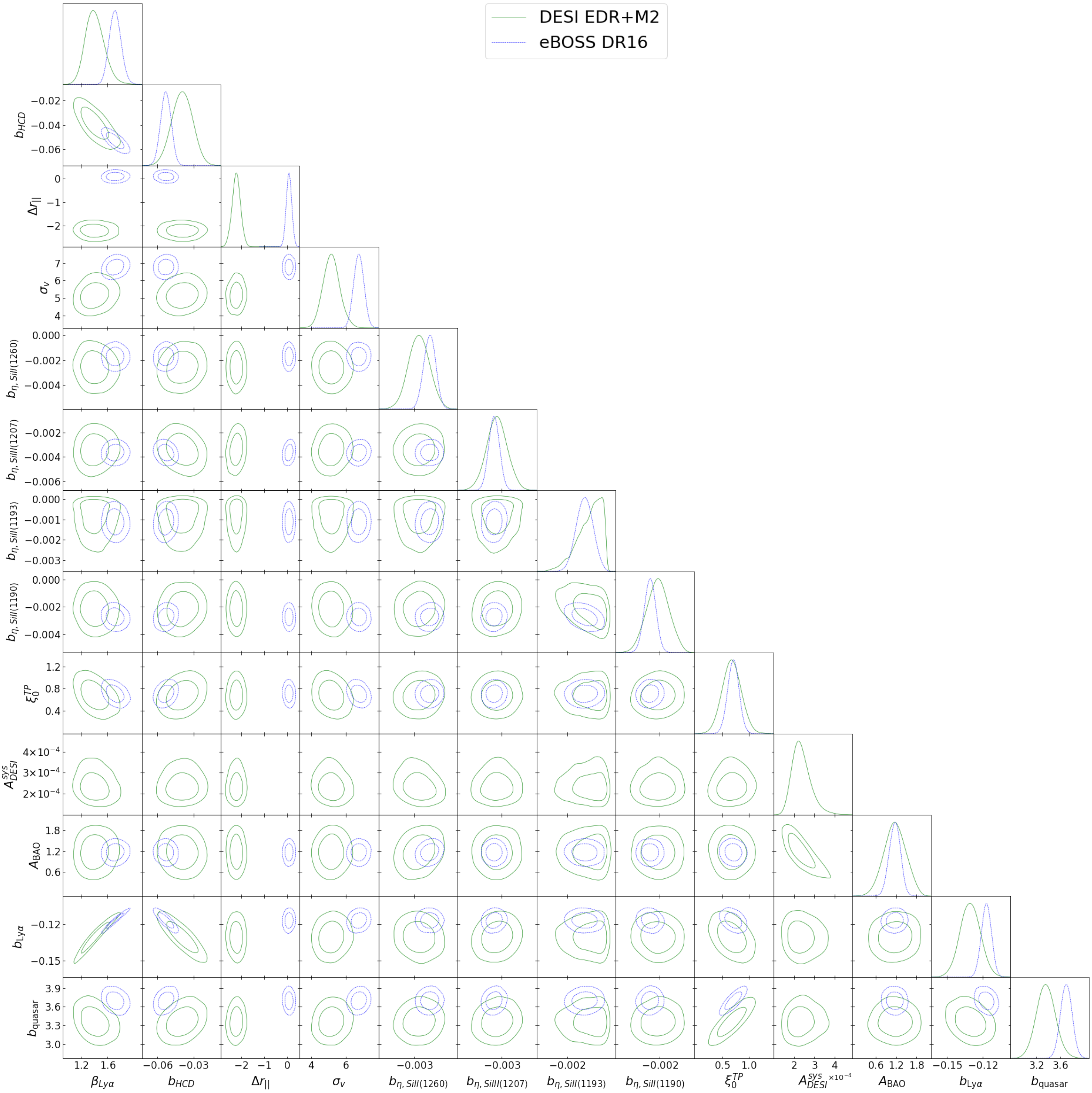

Figure 5.10: Baseline 13 free parameter fit on DESI EDR and eBOSS DR16 auto+cross correlations. We model the instrumental systematics effects of eBOSS with a slightly different model with two parameters (as opposed to just $A_{\rm DESI}^{\rm sky}$ for DESI), which we have not included here. For $b_{\rm Ly\alpha}$ and $b_{\rm q}$ we present posteriors for both surveys at the effective redshift of the eBOSS DR16 data set.

effects parameter A_{inst} and the BAO peak amplitude A_{BAO}. This is somewhat unexpected because the instrumental systematics model should only contribute at small scales ($r_{\parallel} = 0, r_{\perp} < 40\,h^{-1}\mathrm{Mpc}$) and needs to be studied further for future analyses of DESI data.

For each parameter, we see the EDR+M2 result is generally within 1-2σ of the DR16 results, except $\Delta r_{\parallel,\mathrm{q}}$, which has a much larger absolute value for our data set. As mentioned in section 5.5, this is likely because of systematic redshift errors that arise from poorly performing quasar templates. Improvements in the quasar templates used for redshift estimation will be made (Brodzeller et al., 2023) for DESI year 1 to reduce this effect, but for now, it is a well-constrained nuisance parameter in our model. Finally, we see that for EDR+M2, the posterior of SiII(1193) is not Gaussian (it is hitting the upper prior bound of 0) because of difficulties constraining the parameter from our data. From the DR16 fits, we can also see that this is one of the weaker detected metals (although it is still non-zero with 3σ confidence). Thus we expect that, with lower statistical power, it is difficult to detect. DESI year 1 data will present an exciting opportunity to make the most precise constraints on the presence of metals in the Lyα forest, including lines that are not included in our model here but were in the past (e.g. CIV(eff)).

6

Modelling the impact of quasar redshift errors on full-shape analyses of correlations in the Lyα forest.

6.1 Introduction

One of the major objectives in modern cosmology is to understand the nature of dark energy. The discovery of the accelerating expansion of the universe from measurements of distant supernovae (Riess et al., 1998; Perlmutter et al., 1999) introduced the need for some form of dark energy, in the simplest case a cosmological constant. Studies of anisotropies in the Cosmic Microwave Background (CMB) with experiments like Planck (Planck Collaboration et al., 2020) have at the same time provided percent-level constraints on other key cosmological parameters. Another more recent probe of cosmic expansion history is the measurement of Baryon Acoustic Oscillations (BAO), first in galaxy clustering (Cole et al., 2005; Eisenstein et al., 2005) and later with the Lyman-α (Lyα) forest (Busca, N. G. et al., 2013; Bautista et al., 2017; Bourboux et al., 2020). In each case the observable is a feature imprinted in the clustering measurements at comoving separation ~100 h^{-1}Mpc. This feature was created at the epoch of recombination, where sound waves that were propagating through the dense photon-baryon plasma ceased to travel because of proton-electron recombination. The overdense regions at wave peaks gave rise to the preferential clustering scale we observe as BAO today. This scale can be measured in the CMB (Planck Collaboration et al., 2020), and as such can be used as a standard ruler to measure the expansion of the universe (Adame et al., 2025c).

The Dark Energy Spectroscopic Instrument (DESI) (Levi et al., 2013; DESI Collaboration et al., 2016a,b, 2022; Guy et al., 2023; Miller et al., 2024; Poppett et al., 2024) survey will measure the spectra and redshifts of more than 40 million galaxies (Hahn et al., 2023; Raichoor et al., 2023; Schlafly et al., 2023; Zhou et al., 2023) and 3 million quasars (Chaussidon et al., 2023) over the course of a 5-year survey, covering 14 000 deg^2 of the sky. Precise measurements of large-scale structure have already been made in galaxy and quasar clustering for the first year of observations of DESI (DESI Collaboration et al., 2024a; Adame et al., 2025a), comprising spectra of targets between redshifts $0.1 < z < 2.1$, and the Lyα forest at $z > 2.1$ (Adame et al., 2025b). The Lyα forest is a series of absorption lines present in the spectra measured from very distant quasars. They are created when light from these quasars intercepts neutral hydrogen in the IGM, and as such, are an effective tracer of large-scale structure. A BAO measurement from the Lyα forest, made using a combination of the autocorrelation of the Lyα flux-transmission field and its cross-correlation with quasars (Font-Ribera et al., 2013, 2014b), has a higher effective redshift than DESI galaxies (2.33 (Adame et al., 2025b)). This allows us to probe an earlier stage of the Universe than the DESI galaxy sample, which additionally helps to break degeneracies in cosmological constraints (Adame et al., 2025c).

It is possible to drastically improve the cosmological constraining power of the Lyα forest (and galaxy) clustering by including information from the Alcock-Paczynski (AP) effect (Cuceu et al., 2021, 2023a). It was projected in Cuceu et al., 2021 that including AP information from both the BAO peak and the smooth component of the Lyα clustering measurement will give 2.5% and 1% constraints on Ω_m and $H_0 r_\mathrm{d}$, from the full DESI survey. This represents a factor of ~2 improvement over BAO-only analyses. Additionally, a sub-10% level measurement of $f\sigma_8$ (at $z \sim 2.33$) should be possible when combining AP + BAO measurements from the full set of Lyα correlation functions with redshift-space distortion (RSD) constraints derived from the quasar autocorrelation function.

It was demonstrated in Adame et al., 2025b that the analysis method used to measure BAO is quite robust to various systematic effects, in large part due to the decomposition of the peak and smooth components of the correlation function (Kirkby et al., 2013). However, since we are interested in constraining both the peak and smooth components of the AP effect, we expect to be more sensitive to contaminants than the BAO-only measurement. The main contaminants usually considered are continuum fitting distortion (Slosar et al., 2011), high column density (HCD) systems (Font-Ribera and Miralda-Escudé, 2012; Rogers et al., 2018; Pérez-Ràfols et al., 2023), broad absorption line (BAL) features (Ennesser et al., 2022), metal absorption (Font-Ribera and Miralda-Escudé, 2012; Pieri et al., 2014) and redshift errors in quasars arising from both astrophysical peculiar velocities and measurements Font-Ribera et al., 2013. Both HCDs and BALs are identified using neural-network–based finder algorithms (Guo and Martini, 2019; Wang et al., 2022) and masked to reduce their impact on the correlation functions. The former is also modelled to account for bias introduced by lower-density, undetected systems. Metals are modelled successfully by treating each absorber as a biased tracer of structure, where the biases are marginalised during the BAO fits.

Random velocities of galaxies or quasars distort clustering measurements along the line of sight, producing anisotropies relative to the transverse direction — an effect known as Fingers of God (FoGs). This effect will be present in the cross-correlation of Lyα forests with quasars (Font-Ribera et al., 2013), which we will marginalise over using a version of the model from Percival and White, 2009 (see section 6.2). In the Lyα autocorrelation, since the redshifts of Lyα absorbers are set by intergalactic hydrogen and not by quasar motion, we do not need to model this effect.

Errors in redshift measurement produce both the FoG effect and a second, distinct effect on our correlation functions. When computing the Lyα transmission field δ, we use a weighted mean continuum of all forests (equations **??**, 6.1). As first described in Youles et al., 2022, this mean continuum is smoothed by quasar redshift errors, causing spurious features in our correlation functions. They also presented a model for this contamination in the cross-correlation function, and showed that it had a small impact on the accuracy of their BAO constraints. This was also confirmed as a source of systematic bias in DESI DR1 (Cuceu et al., 2025b) and in DESI DR2 (Casas et al., 2025). In this paper we extend this work by modelling the same contamination in the Lyα autocorrelation, and introducing free parameters that we marginalise over in our full-shape analysis.

To measure redshifts of DESI quasar spectra, a package called `Redrock` is used (Bailey et al., 2025). It is a template-fitting algorithm that uses Principal Component Analysis (PCA)-based quasar templates to determine the best-fit redshift in a chi-squared (χ^2) minimisation. We expect a degree of systematic error in these estimations due to spectral variation and limited performance of templates, which was found to be $\sim 340\,\mathrm{km/s}$ (Alexander et al., 2023) for DESI survey validation. These variations are caused by Doppler shift of quasar emission lines due to the physical processes happening in the line-emitting region (Shen et al., 2016). The typical level of shift away from the systemic quasar redshift depends on the line, but highly ionised, broad emission lines tend to exhibit the largest shifts. Recent work on new templates have significantly reduced this error for DESI Data Release 1 (DR1; DESI Collaboration et al., 2025a), showing it to be ~50% smaller than survey validation (Brodzeller et al., 2023; Bault et al., 2025). This was partly done by incorporating precise measurements of the Mg II line, which has relatively low bias with respect to the systemic redshift of the quasar. Broad emission lines within the Lyα forest, listed in table 6.1 (Harris et al., 2016), can also be Doppler-shifted, potentially contributing to errors in the mean continuum estimation (see section 6.2.2).

This paper is organised as follows. In section 6.2, we describe the mock datasets used to develop and test our model, the method for generating the flux transmission field from quasar spectra, and the procedures for computing and modelling our correlation functions. In section 6.3, we show measurements of the Lyα auto- and Lyα –quasar cross-correlations from different mock datasets, both with and without redshift errors. Section 6.4 presents the redshift error model for both the auto- and cross-correlation functions, and section 6.5 shows results from AP and isotropic BAO fits on our mock datasets. In section 6.6, we discuss the application of our model to real data and how the contamination can be mitigated using data cuts. Finally, section 6.7 summarises our findings.

6.2 Method

In this section we present the mock datasets we use (section 6.2.1), and the analysis techniques employed to derive measurements of BAO and AP parameters from synthetic quasar spectra. The latter part begins with the flux transmission field measurement in section 6.2.2, followed by a description of how our correlation functions are constructed in section 6.2.3. Finally, we describe the model of the correlation functions and how we constrain BAO and AP parameters in sections 6.2.4 and 6.2.5.

6.2.1 Data sets

To test our model, we use synthetic data sets developed to validate the Lyα BAO analysis in Adame et al., 2025b, described in detail in Cuceu et al., 2025b; Herrera-Alcantar et al., 2025b. These are based on the `CoLoRe`[1] (Ramírez-Pérez et al., 2022) suite, which generates log-normal density fields with quasars distributed according to a specific biasing model. Lines-of-sight (also called skewers) are drawn from quasar positions in this field, with modified small-scale power, and

---

[1] https://github.com/damonge/CoLoRe

Line	Wavelength [Å]	Equivalent Width [Å]	σ [Å]
SiIV	1064/1074	2.9/0.7	7.7/3.5
NII	1083	1.32	5.3
PV	1118/1128	0.76/0.46	5.3/4.1
CIII*	1175	2.49	7.7

Table 6.1: The emission lines within the Lyα forest included in our synthetic spectra (Harris et al., 2016). We show from left to right, their rest-frame wavelength (in Angstroms), the mean equivalent widths and mean FWHM.

converted to transmitted flux fraction using `LyaCoLore`[2] (Farr et al., 2020b). The clustering of these noiseless skewers is designed to be realistic at scales relevant for BAO systematics studies (~100 h^{-1}Mpc). `LyaCoLore` also stores skewers of metal absorption and can include high column-density (HCD) systems, both of which are major contaminants in standard Lyα BAO analyses.

To generate realistic quasar spectra from the flux transmission skewers, we use a final package called `desisim`[3] (Herrera-Alcantar et al., 2025b). This takes the transmission skewers of `LyaCoLore` as an input, multiplies them by a quasar continuum template, and adds noise to mimic the observing conditions and the instrumental model of DESI. Continuum templates are generated using `SIMQSO`[4] (McGreer et al., 2021), which combines a broken power law with a series of Gaussian emission lines. The slopes of the power law are sampled from a Gaussian, with mean and dispersion tuned to better reflect the continuum shape and variability of quasars in the eBOSS DR16 dataset (Bourboux et al., 2020). The Lyα forest emission lines are simulated from the composite model of BOSS spectra (Harris et al., 2016), with line diversity is drawn from the distribution of equivalent widths (EWs). In table 6.1, we show the properties of these emission lines.

To emulate the effect of redshift errors in our mock datasets, we add Gaussian-distributed velocities dv with zero mean and a dispersion of 400 km/s to each quasar redshift ($1 + z_q = (1 + z_q^0)(1 + dv/c)$). The reason for doing this rather than using the error from a redshift-fitting algorithm is that the latter has been shown to perform better on simulated spectra than real data (Farr et al., 2020a). In Youles et al., 2022, they differentiate between the Fingers of God (FoG) effect and redshift errors that affect the mean continuum. In this paper we focus on the latter.

We use two synthetic datasets: one has redshift errors added only to the tracer quasar catalogue, emulating the effect of FoGs. In the other we add redshift errors immediately after quasar spectra are generated, which will additionally

---

[2] https://github.com/igmhub/LyaCoLoRe

[3] `https://github.com/desihub/desisim`

[4] https://github.com/imcgreer/simqso

give rise to mean continuum errors (see section 6.4). Note that in our measured correlation functions, the latter set will exhibit the combined effect of FoGs and mean continuum errors. This means that later in section 6.3 when we compare the differences between the two sets, we will observe only the contamination caused by continuum errors.

Each set consists of a stack of 100 mocks designed to emulate the survey conditions of DESI DR1 (DESI Collaboration et al., 2025a), including target density, footprint and signal-to-noise ratio. We will refer to these as DR1 mocks going forward. They also include all of the major contaminants affecting the analysis on data: HCDs, metal absorption and broad absorption lines (BALs). Using this synthetic dataset, we can validate our model and test correlations between various parameters in a controlled environment.

6.2.2 Continuum fitting

The flux transmission field ($\delta_F(\lambda)$) used to perform our clustering analysis is defined in equation 5.1, and is discretised into "pixels" at the dispersion of the DESI spectrograph, $\Delta\lambda$ =0.8 Å. Since we can't measure the unabsorbed continuum $C_q(\lambda)$ directly, we re-characterise its product with the mean transmission $\overline{F}(z)$:

$$\overline{F}(\lambda)C_q(\lambda) = \overline{C}(\lambda^{\mathrm{rf}})\left(a_q + b_q\frac{\log\lambda - \log\lambda_{\mathrm{min}}}{\log\lambda_{\mathrm{max}} - \log\lambda_{\mathrm{min}}}\right). \tag{6.1}$$

The term $\overline{C}(\lambda^{\mathrm{rf}})$, referred to as the "mean continuum", is the weighted mean of all forests in our analysis. This is multiplied by a first degree polynomial in $\log\lambda$ for each quasar q in our sample, to account for continuum variability. The free parameters a_q and b_q are estimated by "continuum fitting", a process that involves maximising the log-likelihood:

$$2\ln L = -\sum_q \frac{(f(\lambda) - \overline{F}(z)C_q(\lambda)(\lambda, a_q, b_q)}{\sigma_q^2(\lambda)} - \sum_q \sigma_q^2(\lambda), \tag{6.2}$$

where $\sigma_{\mathrm{q}}^2(\lambda)$ is the flux variance, a combination of pipeline noise and large-scale structure. The mean continuum $\overline{C}(\lambda^{\mathrm{rf}})$ is also computed during this maximisation. For further details on exactly how this is done, see Ramírez-Pérez et al., 2024.

Fitting each δ_q using the data from the entire forest q, means the measured transmission at λ_i will be linear combination of all other pixels in the forest. Explicitly, $\delta_{F,i} = \eta_{ij}\delta_{F,j}$. With the same methodology as section 5.3.2, we apply a projection matrix (equation 5.5) to δ_F, such that the resulting field $\tilde{\delta}_F$ will have the same distortion as our projected correlation model (section 5.4.4).

6.2.3 Correlation functions

We measure our correlations in configuration space as a function of transverse and line-of-sight separation $(r_\perp, r_\parallel)$. To convert from redshift and angular separation to $r_\parallel$ and $r_\perp$ we use equations 5.7 and 5.8 respectively. A fiducial cosmology is required to compute both distances in terms of redshift, so we use the same Planck 2018 (hereby `P18`; Planck Collaboration et al., 2020) cosmology as Gordon et al., 2023; Adame et al., 2025b. The correlation within each bin $(r_\perp, r_\parallel) \in A$ is then determined by a the weighted average given in equation 5.9, and each Lyα pixel or quasar has weight given by equation 5.3.

Note that our weights evolve with redshift proportional to $(1+z)^{\kappa-1}$, where $\kappa = 2.9$ for Lyα pixels, and $\kappa = 1.44$ for quasars. For the cross-correlation, index j in equation 5.9 represents a quasar, which we treat as points that effectively have $\delta_j = 1$. This is equivalent to computing the weighted mean of δ_F at set distances from a quasar. In both cases we choose a bin size of 4 h^{-1}Mpc following Gordon et al., 2023; Adame et al., 2025b. In the autocorrelation we measure between 0 and 200 h^{-1}Mpc in both directions, giving a total of 2500 bins. The cross-correlation however is not symmetric along the line-of-sight under permutation of pixels and quasars. Therefore, we define $r_\parallel$ between -200 and 200 h^{-1}Mpc, where negative separations correspond to a quasar being behind the Lyα pixel with respect to the observer and vice-versa for positive separations. This gives us 5000 bins in total for the cross-correlation.

The covariance of our measurement, explained in section 5.3.5, is estimated by splitting our dataset into sub-samples s and computing equation 5.12. In our case, each Lyα forest or quasar is uniquely set in one `HEALpix` sample s. The covariance is then smoothed by averaging all off-diagonal elements of the correlation matrix $(C_{AB}/\sqrt{(C_{AA}C_{BB})})$ with the same $\Delta r_\perp = r_\perp^A - r_\perp^B$ and $\Delta r_\parallel = r_\parallel^A - r_\parallel^B$. Furthermore, as introduced in Adame et al., 2025b, we now include the cross-covariance between the Lyα autocorrelation and Lyα-quasar cross-correlation functions in our analysis.

In Adame et al., 2025b and Cuceu et al., 2025a they use two additional correlation functions that measure the autocorrelation of Lyα absorption in a bluer region of the quasar ("region B", between 920-1020 Å DESI Collaboration et al., 2025b) and its cross-correlation with quasars. For simplicity, and because the relative contribution to cosmological constraining power is small ($\sim 10\%$), we choose not to include them in our analysis.

6.2.4 BAO and full-shape information

In this section we outline the model used to extract cosmological information and capture behaviour of contaminants in our analysis. The full-shape analysis that we outline here was first described for the Lyα forest in Cuceu et al., 2021, and performed on eBOSS data in Cuceu et al., 2023a. We follow the method of Kirkby et al., 2013 and employ a template power spectrum, based on `P18` cosmology, that we allow to vary from the measured cosmology by re-scaling our co-ordinate grid:

$$r'_{\parallel} = q_{\parallel} r_{\parallel}, \quad r'_{\perp} = q_{\perp} r_{\perp}. \tag{6.3}$$

We wish to distinguish between isotropic re-scaling α of the correlation function and anisotropic re-scaling ϕ (AP effect), therefore we define:

$$\alpha(z) = \sqrt{q_{\parallel} q_{\perp}} \text{ and } \phi(z) = \frac{\mathrm{q}_{\perp}}{\mathrm{q}_{\parallel}} \tag{6.4}$$

The first step in standard Lyα BAO analyses is to take the linear isotropic input power spectrum from `P18`, $P_{\mathrm{L}}(k)$, and decompose it into peak ($P_{\mathrm{L}}^{\mathrm{peak}}(k)$) and smooth ($P_{\mathrm{L}}^{\mathrm{sm}}(k)$) components following Kirkby et al., 2013. This is possible because the BAO is a distinct feature in the full correlation function, which also makes the standard peak-only analysis more robust to systematics affecting the smooth component. In the full-shape analysis we keep this peak-smooth decomposition, but introduce scaling parameters α and ϕ for each component:

$$\xi(r_{\perp}, r_{\parallel}) = \xi_{\mathrm{p}}(r_{\perp}, r_{\parallel}, \alpha_{\mathrm{p}}, \phi_{\mathrm{p}}) + \xi_{\mathrm{s}}(r_{\perp}, r_{\parallel}, \alpha_{\mathrm{s}}, \phi_{\mathrm{s}}), \tag{6.5}$$

where "s" and "p" refer to the smooth and peak components respectively. The peak component is defined in terms of distances and the sound horizon scale r_{d} as:

$$\alpha_{\mathrm{p}} = \sqrt{\frac{d_{\mathrm{m}}(z) d_h(z)/r_{\mathrm{d}}^2}{[d_{\mathrm{m}}(z) d_h(z)/r_{\mathrm{d}}^2]_{\mathrm{fid}}}} \tag{6.6}$$

where d_h is the Hubble distance and "fid" is the fiducial `P18` cosmology. From this one can constrain the combination of $H_0 r_{\mathrm{d}}$ and Ω_{m}, where H_0 is the Hubble constant. Both AP components are equivalent to $\phi = d_{\mathrm{m}}(z)H(z)/[d_{\mathrm{m}}(z)H(z)]_{\mathrm{fid}}$. From this we can directly measure Ω_{m}, and therefore break the degeneracy between Ω_{m} and $H_0 r_{\mathrm{d}}$. A more detailed description of this is given in Cuceu et al., 2021.

In Cuceu et al., 2023b, it was shown that including the smooth component AP parameter ϕ_{s} doubled the precision of constraints on Ω_{m}. Note that in Cuceu et al., 2023b and in this work, we treat α_{s} as a nuisance parameter since it is not clear how to extract cosmological information from it, and it's more strongly correlated with the Lyα bias. Finally, while the information gain of including the smooth-component is substantial, it is also more susceptible to contaminants, thus motivating us to create the model presented in this paper.

6.2.5 Correlation model

To compute our Lyα and Lyα-quasar correlation function models we begin with the corresponding power spectra, which are derived from linear perturbation theory, with a Kaiser (Kaiser, 1987) model of RSD. Our power spectra are the same as those given by equation 5.14, with an additional term Gaussian smoothing term F_{sm} that accounts for our log-normal simulations having limited grid size (~2.4 h^{-1}Mpc):

$$P_\alpha(k, \mu_{\mathrm{k}}, z) = b_\alpha^2(1 + \beta_\alpha \mu_{\mathrm{k}}^2)^2 F_{\mathrm{sm}} P_{\mathrm{L}}(k, z), \tag{6.7}$$

$$P_{\mathrm{X}}(k, \mu_{\mathrm{k}}, z) = b_\alpha b_{\mathrm{q}}(1 + \beta_\alpha \mu_{\mathrm{k}}^2)(1 + \beta_q \mu_{\mathrm{k}}^2) F_{\mathrm{sm}} F_{\mathrm{NL,q}} P_{\mathrm{L}}(k, z)\,. \tag{6.8}$$

b_i are the cosmological biases and β_i are the RSD parameters for tracer i (quasar or Lyα pixel), and μ_{k} is the cosine of the k vector with respect to the line-of-sight direction. Our correlation models are then computed by taking the Fast Hankel Transform of the above equations. All of this is done with the package `Vega`[5], which we also use to fit cosmological parameters, given in section 6.5.

Note that the cross-correlation term $F_{\mathrm{NL,q}}$ accounts for quasar peculiar velocities, and is given in equation 5.19. These produce a smoothing effect at large $k_{||}$, and are a distinct effect from the redshift error continuum smoothing we model in this paper. We also account for any systematic quasar redshift errors by introducing a free parameter to our cross-correlation model: $\Delta r_{||} = r_{||,\mathrm{true}} - r_{||,\mathrm{measured}}$.

In the contaminated mock dataset we introduce metal absorption and HCD contamination. For HCDs, we mask systems with column density $> 2 \times 10^{20}\mathrm{cm}^{-2}$, following the analysis of Adame et al., 2025b. The remaining HCDs, which in real data are too small to detect for masking, are biased tracers of large-scale structure. In this case we treat the Lyα bias and RSD parameters as effective combinations of signal from Lyα forest and HCDs, given in equations 5.25 and 5.26. L_{HCD} in these equations is the typical length-scale of the unmasked HCDs, which along with the HCD bias and RSD parameters, we marginalise over.

---

[5] https://github.com/andreicuceu/vega

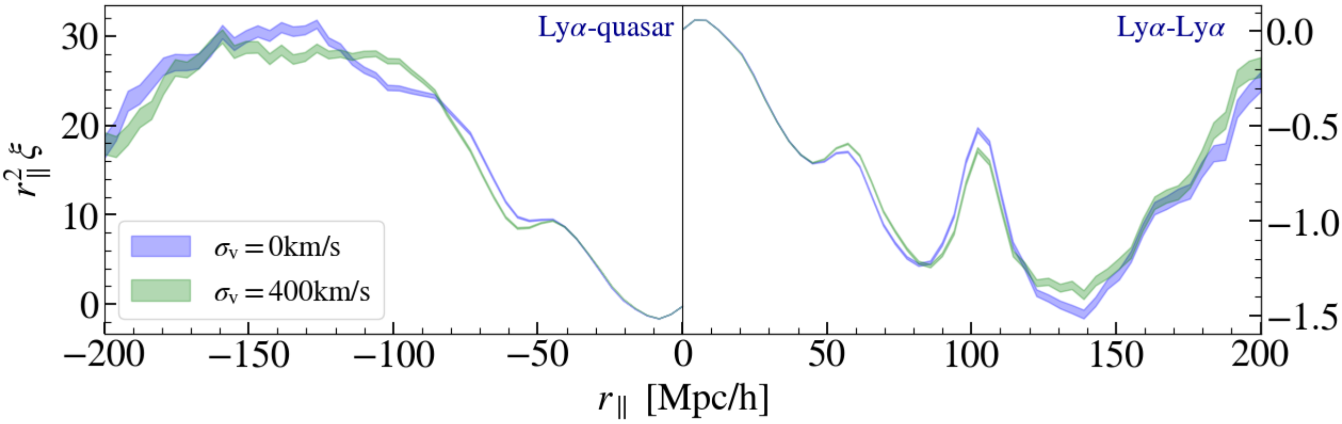

Figure 6.1: The effect of continuum redshift errors on the Lyα-quasar cross-correlation (left) and the Lyα-Lyα autocorrelation (right) computed from our mock datasets. Each dataset is the stack of 100 DESI DR1 mocks. We take the weighted average over the first 4 bins in transverse separation $r_\perp$ ([0,16] h^{-1}Mpc) where the effect of redshift errors is strongest, and plot as function of line-of-sight separation $r_{\parallel}$. We only include negative $r_{\parallel}$ for the cross-correlation, since as shown in figure 6.2 there is no impact at $r_{\parallel} > 0$ h^{-1}Mpc. Note that at $\sim \pm 60$ h^{-1}Mpc we see bumps in the correlations due to SiII(1190/1193) lines.

In our analysis, we assume all of the absorption within the Lyα forest is due to the Lyα transition. However, in some cases we are mistaking Lyα absorption with absorption from different metals. This leads to a contamination which we correct for using the same process outlined in Adame et al., 2025b. In our fits, we marginalise over a set of bias parameters for 4 metal lines: SiIII(1207), SiII(1190), SiII(1193), and SiII(1260) (section 6.5).

As mentioned in section 6.2.2, distortions introduced to δ during continuum fitting lead us to use a "projected" field $\tilde{\delta}$ (equation 5.4). This is designed such that the projected distorted field has the same distortion as the projected true field. We write the distorted model (A) in terms of the undistorted model (B) using a "distortion matrix": $\xi_A = \sum_{AB} D_{AB}\xi_B$. D is given for the autocorrelation and cross-correlation in equations 5.32 and 5.33 respectively.

6.3 Impact of redshift errors

In this section we will show the impact of continuum redshift errors on the measured correlation functions of our mock datasets. Then, in section 6.4, we will discuss the origin of this contamination and propose a model for it.

We use the two synthetic datasets described in section 6.2.1, both of which are stacks of 100 DESI DR1 mocks. The "uncontaminated" set has redshift errors added to the tracer quasar catalogue, emulating FoGs. The "contaminated" set has errors added to the spectra, emulating pipeline redshift errors. Using this method we isolate the effect of redshift errors which enter during the continuum fitting process (section 6.2.2). We refer to this contamination as continuum redshift errors, to differentiate it from the effect of FoGs.

We then compute the Lyα autocorrelation and its cross-correlation with quasars using the method outlined in the previous section. For each of the two sets, we compute the weighted mean and covariance of the 100 correlation functions. We use these two stacks to model and fit the effect of continuum redshift errors in our correlation functions in the following sections. The auto- and cross-correlation functions from the uncontaminated and contaminated sets are shown in figure 6.1. These are plotted as a function of the line-of-sight separation, averaged over the first 4 bins in transverse separation ([0,16] h^{-1}Mpc) where the effect is strongest. We can also clearly see a peak at $\sim 100\ h^{-1}$Mpc, a combination of BAO and SiII(1260) contamination, and other peaks caused by metal contamination (section 6.2.5).

The contamination of continuum redshift errors (the difference between the two datasets) in the auto- and cross-correlations is shown in figure 6.2. As discussed in Youles et al., 2022, there is a strong dependence of redshift error distortion in the cross-correlation on the small-scale quasar autocorrelation, which is significantly weaker for positive $r_{\|}$ where host quasars of Lyα pixels are much further from the correlating quasars. Therefore, as we see in figure 6.2, the distortion from redshift errors is only visible for negative $r_{\|}$. This dependence on the quasar autocorrelation is also the reason the contamination becomes negligible at transverse separations $r_{\perp} \gtrsim 20\ h^{-1}$Mpc. The behaviour is similar in the Lyα autocorrelation, except the dependence is now on the small-scale Lyα-quasar cross-correlation. In section 6.4, this dependence can be seen explicitly in our model. The distortions in figure 6.2 are also localised around specific values of $r_{\|}$. In section 6.4.1 we show how this pattern arises due to smoothing of the mean continuum.

We can see that the continuum redshift errors will contaminate our full-shape measurement, but also that part of the contamination overlaps with the fiducial BAO position ($\sim 100\ h^{-1}$Mpc). In section 6.5, we show the effect of this contamination on the set of cosmological parameters measured from the BAO peak (α_{p}, ϕ_{p}) and the broadband component (ϕ_{s}, $f\sigma_8$) of our correlation functions.

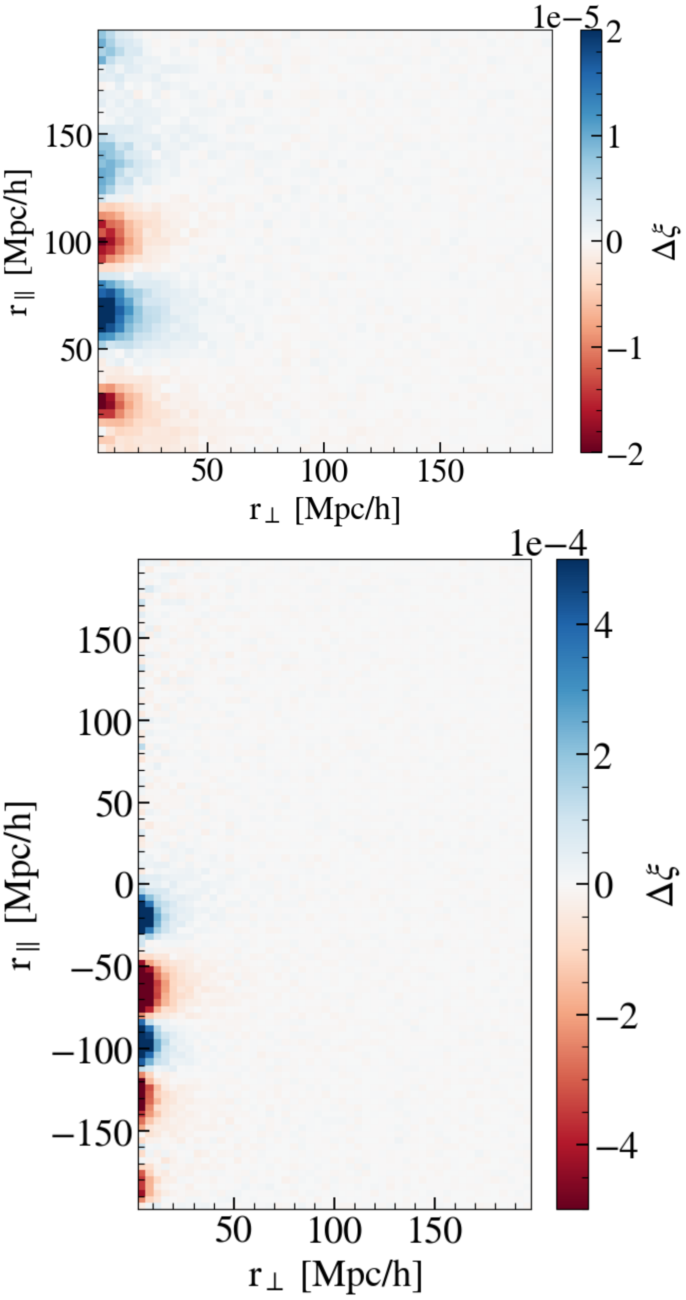

Figure 6.2: (Top) difference in the Lyα autocorrelation measured from contaminated (with continuum redshift errors) and uncontaminated datasets, as a function of $(r_\perp, r_\parallel)$. (Bottom) difference in the Lyα-quasar cross-correlation from contaminated and uncontaminated datasets.

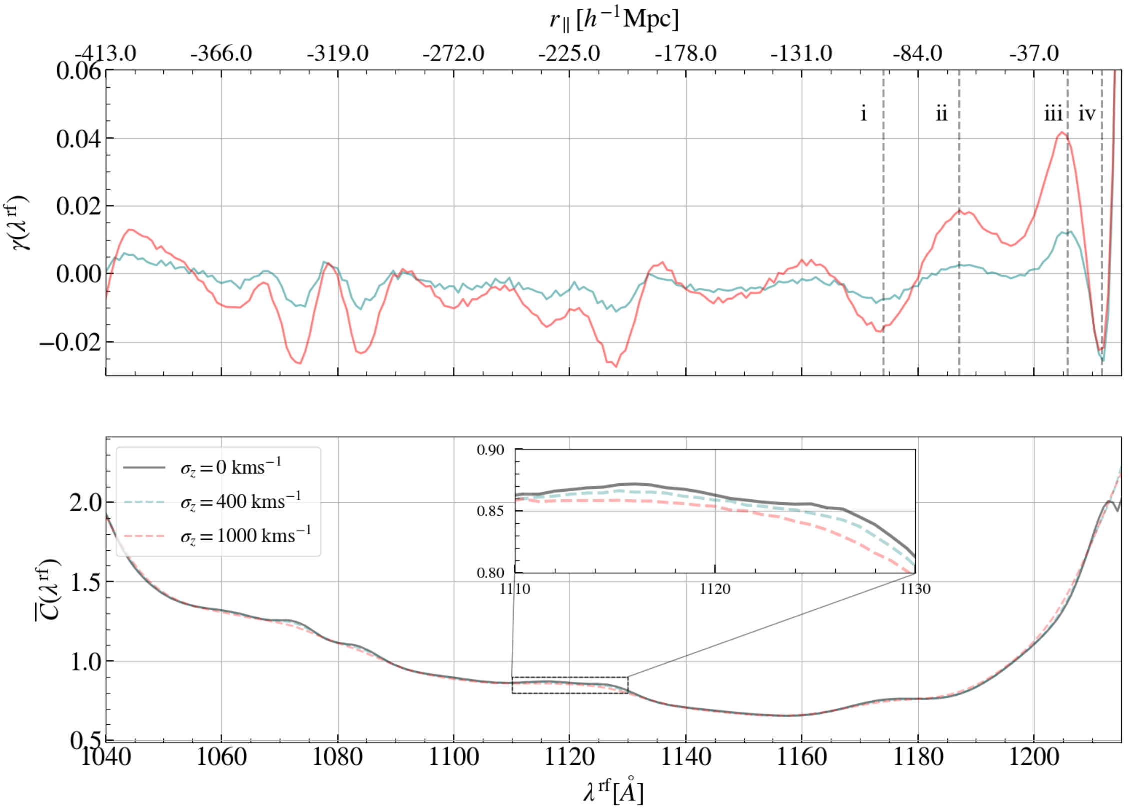

Figure 6.3: (Top) the mean continuum distortion function $\gamma = \hat{\overline{C}}/\overline{C} - 1$ as a function of rest-frame wavelength. $\hat{\overline{C}}(\lambda_{\mathrm{rf}}, \sigma_{\mathrm{v}})$ is the mean continuum with redshift errors σ_{v}, as shown in the plot above. Roman numerals mark the location of prevalent features in γ, which contaminate our correlation functions. We include the approximate comoving distance between a quasar at $z = 2.3$ and pixels in its forest, along the top axis. (Bottom) the Lyα forest mean continuum (equation 6.1) of our mock datasets with $0\,\mathrm{kms}^{-1}$, $400\,\mathrm{kms}^{-1}$ and $1000\,\mathrm{kms}^{-1}$ of redshift errors added. The later is for visualisation while $400\,\mathrm{kms}^{-1}$ is used in the actual analysis.

6.4 Modelling the contamination

In this section we will outline our model for continuum redshift error contamination. We start describing the effect of redshift errors on the mean continuum and flux-transmission field (following Youles et al., 2022), and carry this through to a full model of the contamination in the autocorrelation. We also highlight the cross-correlation redshift error model of Youles et al., 2022.

In Youles et al., 2022, redshift errors were identified to smooth the mean continuum (equation 6.1) that is fit during the transmission field estimation. They subsequently defined a term $\gamma(\lambda^{\mathrm{rf}})$:

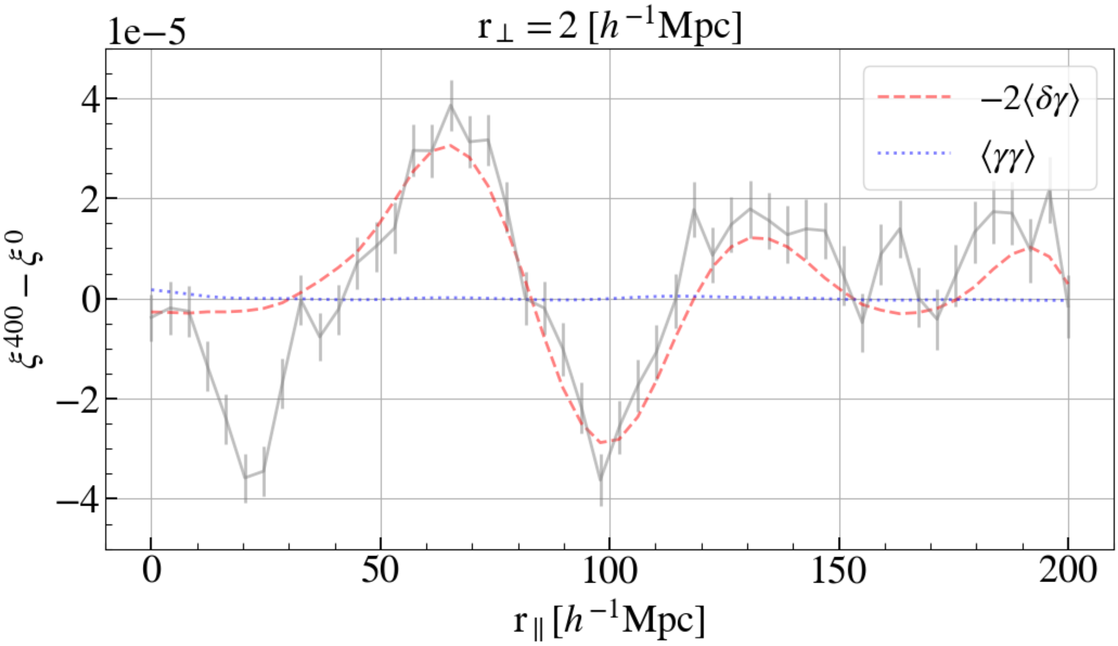

Figure 6.4: The difference in the Lyα autocorrelation function between contaminated (with continuum redshift errors) and uncontaminated datasets (grey), overlaid with direct measurements of $\langle\delta\gamma\rangle$ (red dashed) and $\langle\gamma\gamma\rangle$ (blue dotted). We plot only the first $r_\perp$ bin where the contamination is strongest.

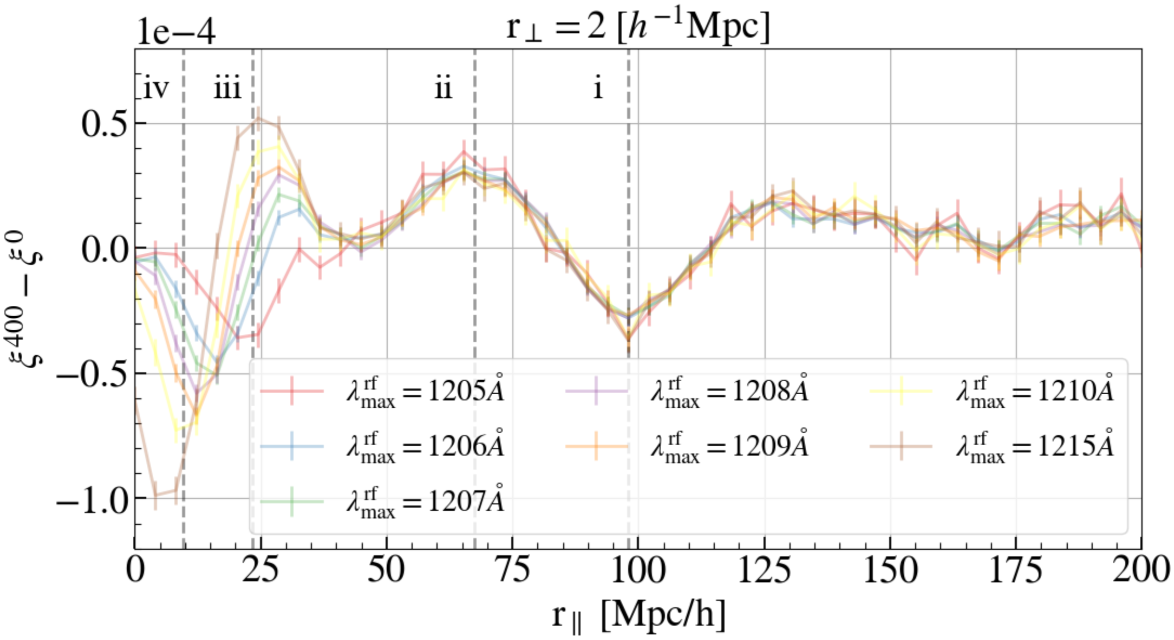

Figure 6.5: The evolution of continuum redshift error contamination with the maximum rest-frame wavelength of the Lyα forest. The 1205Ålimit (red) is the limit used in the analysis of this paper. We include Roman numerals to indicate the features in γ (figure 6.3) which correspond to spurious correlations in this figure.

$$\gamma(\lambda^{\mathrm{rf}}) = \frac{\hat{\overline{C}}(\lambda^{\mathrm{rf}})}{\overline{C}(\lambda^{\mathrm{rf}})} - 1, \tag{6.9}$$

where $\hat{\overline{C}}$ is the mean continuum estimate with redshift errors and $\overline{C}$ is the true mean continuum. The Lyα forest region of our mock continua consists of a power law smooth component, superimposed with many Gaussian emission lines. If quasar redshift errors are randomly distributed, the mean continuum will be smoothed around emission lines as shown in figure 6.3. In appendix 6.9, we write an expanded version of $\gamma(\lambda^{\mathrm{rf}})$, accounting for the fact that the wavelength grid λ^{rf} for each forest is shifted from the true grid, due to errors in z_{q}.

In the top panel of figure 6.3, we show a measurement of $\gamma(\lambda^{\mathrm{rf}})$ formed using mean continua from our mock datasets with redshift errors of 400 kms^{-1} and 0 kms^{-1}. Combining equation 6.9 with our definition of the flux transmission field (equation 5.1) we get:

$$\hat{\delta}(\lambda) = \frac{1 + \delta(\lambda)}{1 + \gamma(\lambda^{\mathrm{rf}})} - 1. \tag{6.10}$$

Assuming we can ignore second-order terms we have:

$$\hat{\delta}(\lambda) \approx \delta(\lambda) - \gamma(\lambda^{\mathrm{rf}}). \tag{6.11}$$

Thus the measured autocorrelation function can be written as:

$$\langle \hat{\delta}(\lambda)\hat{\delta}(\lambda) \rangle = \langle \delta(\lambda)\delta(\lambda) \rangle - 2\langle \delta(\lambda)\gamma(\lambda^{\mathrm{rf}}) \rangle + \langle \gamma(\lambda^{\mathrm{rf}})\gamma(\lambda^{\mathrm{rf}}) \rangle. \tag{6.12}$$

To determine the most dominant term in this equation, we directly compute $\langle \delta\gamma \rangle$ and $\langle \gamma\gamma \rangle$ using δ from mock datasets with no added redshift errors, and the measured gamma in figure 6.3. The result is shown in figure 6.4, where both terms are plotted over the measured contamination in the autocorrelation. Clearly, $\langle \gamma\gamma \rangle$ is negligible. Therefore, we only model the contribution of the $\langle \delta\gamma \rangle$ term. Noticeably, the trough in the correlation difference at $\sim 25\,h^{-1}$Mpc is not captured by either term. In the following section we discuss the behaviour of this feature.

6.4.1 Variation with wavelength

Spurious features in $\langle \delta\gamma \rangle$ caused by redshift errors in the Lyα continuum appear at specific separations. These separations (in comoving distance) are determined by the difference in rest-frame wavelength between the Lyα line and strong features in $\gamma(\lambda^{\mathrm{rf}})$:

$$r^{\text{feature}} \approx \frac{c(1+\overline{z})}{H(\overline{z})}\frac{(\lambda^{\text{feature}} - \lambda^{\alpha})}{\lambda^{\alpha}}, \tag{6.13}$$

where $\overline{z}$ is the effective redshift of the dataset and H is the Hubble parameter. In the top panel of figure 6.3, we highlight key features in Roman numerals: i at 1174 Å, ii at 1187 Å, iii at 1205.7 Å and iv at 1211.6 Å, the latter two of which fall outside of the fiducial rest-frame limits (1205 Å) of our clustering analysis. Using these values in equation 6.13 with an effective redshift of 2.3 predicts spurious correlation features at separations of 99 h^{-1}Mpc, 70 h^{-1}Mpc, 26 h^{-1}Mpc and 9 h^{-1}Mpc respectively. In figure 6.5, we show the contamination in the Lyα autocorrelation as a function of the upper rest-frame wavelength limit of the forest. We also include Roman numerals here, indicating the aforementioned comoving separations, and see clearly that they align with the strongest spurious correlation features.

Looking specifically at the rest-frame limit of our analysis, 1205 Å(red line in figure 6.5), we see a trough at $\sim 25\,h^{-1}$Mpc. This roughly overlaps with the feature iii (at 1205.7 Å) in figure 6.3, but is a trough where we expect a peak (peaks in γ produce peaks in $\hat{\xi}_{\alpha\alpha}$). Therefore we posit that this trough is instead caused by feature iv in figure 6.3, at a rest-frame wavelength of 1211.6 Å. This is possible because quasar redshift errors not only smooth the mean continuum, but also shift the rest-frame wavelength grid with respect to the truth. For any forest in our dataset, we apply the upper rest-frame wavelength limit by converting observed wavelength λ into rest-frame, where $\lambda^{\text{rf}} = \lambda/(1+\hat{z}_{\text{q}})$ for a measured quasar redshift $\hat{z}_{\text{q}}$. If $\hat{z}_{\text{q}} > z_{\text{q}}$, pixels which have $\lambda^{\text{rf}} > 1205$Å will fall into the accepted Lyα region. Consequently, values of $\gamma(\lambda^{\text{rf}} > 1205\,\text{Å})$ will contaminate our analysis.

In figure 6.6, we show (blue dashed line) the result of removing pixels (for each forest in our correlation measurement) which are in reality above the Lyα rest-frame wavelength limit, but are accepted into the sample because of redshift errors. In this case, the trough feature is no longer visible, and the contamination at small-scales is now roughly consistent with 0. We have also verified that including pixels that are in reality below 1205 Å, but are excluded due to redshift errors, has no effect on the shape of the contamination.

In appendix 6.9, we make the shifting of the rest-frame wavelength grid where we evaluate each δ explicit, by expanding $\hat{\overline{C}}$ about small redshift deviations. This produces extra terms that contaminate the correlation functions, but to 1st order they do not capture the trough at 25 h^{-1}Mpc in our fiducial analysis. We also try to empirically model the trough feature (equation 6.33), which works well in a direct fit (see figure 6.8), but biases our full-shape analysis.

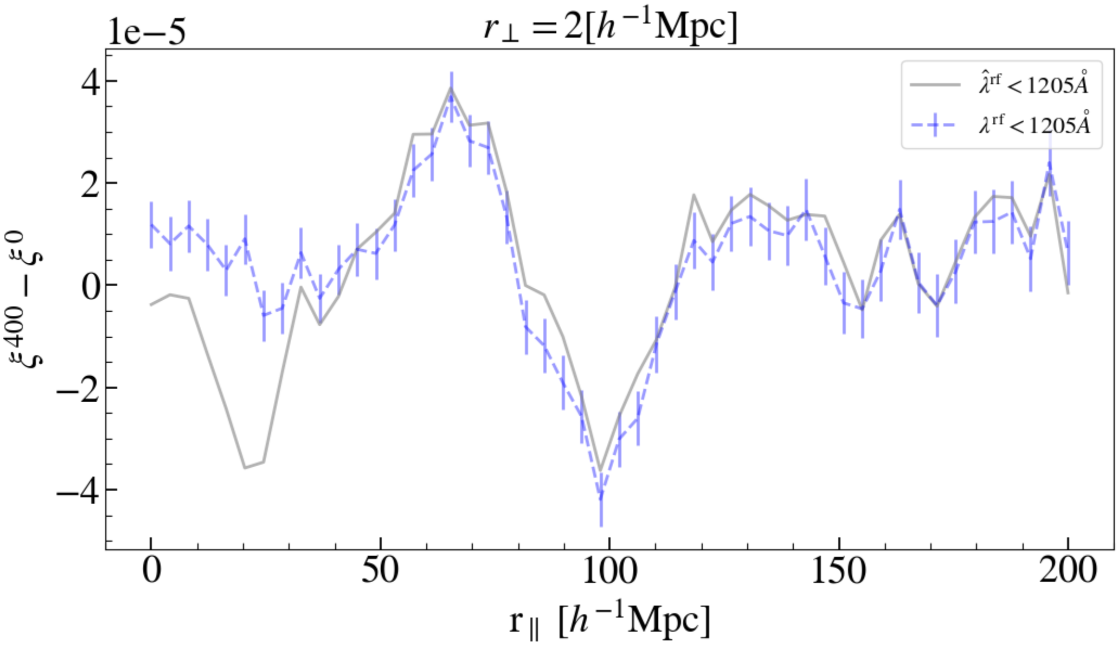

Figure 6.6: Contamination caused by redshift errors (grey), and the same contamination after removing pixels with true rest-frame wavelength greater than 1205 Å (blue dashed). After removing these pixels the trough feature at $\sim 25\,h^{-1}$Mpc disappears, while the rest of the contamination is unaffected.

6.4.2 Autocorrelation model

To model the $\langle\delta(\lambda)\gamma(\lambda^{\mathrm{rf}})\rangle$ term in equation 6.12, we follow a similar approach to that used in Youles et al., 2022 to model the $\langle\gamma(\lambda^{\mathrm{rf}})\rangle$ term in the Lyα-quasar cross-correlation model.

We begin by considering a general expression for the expectation value of $\langle\delta\gamma\rangle$:

$$\langle\delta\gamma\rangle_A = \int_{-1}^{\infty} d\delta \int_{-1}^{\infty} d\gamma P(\delta,\gamma|\mathbf{r}^A)\delta\gamma \tag{6.14}$$

where $P(\delta,\gamma)$ is the probability of having a particular δ and γ, for separation $\mathbf{r}^A = \{\mathbf{r}_{\parallel}{}^A, \mathbf{r}_{\perp}{}^A\}$ between two Lyα pixels. This can be equivalently expressed as:

$$\langle\delta\gamma\rangle_A = \int_{-1}^{\infty} d\delta \int_{\lambda^{\mathrm{rf,min}}}^{\lambda^{\mathrm{rf,max}}} d\lambda^{\mathrm{rf}} P(\delta,\lambda^{\mathrm{rf}})\delta\gamma, \tag{6.15}$$

since γ is uniquely defined by λ^{rf}, and the integration limits are given by the fact that we only consider pixels within the defined Lyα forest rest-frame region. We can write the quasar rest-frame wavelength of any pixel as:

$$\lambda^{\mathrm{rf}} = \lambda^{\alpha}(1+z^{\alpha})/(1+z^{\mathrm{Q}}), \tag{6.16}$$

where λ^{α} is the rest-frame wavelength of the Lyα transition, and z^{α} is the redshift of a Lyα pixel in the Lyα forest region of a quasar with redshift z^{Q}. Therefore, for a given λ_i^{rf} for forest i, we have a set of possible $\{z_i^{\mathrm{Q}}, z_i^{\alpha}\}$. Furthermore, for separations $r_{\parallel} \in A$ and pixel redshift z_i^{α}, we have a unique z_j^{α} in forest j. Thus, we re-write equation 6.15 as:

$$\langle \delta\gamma \rangle_A = \int_{-1}^{\infty} d\delta_j \int_{z^{\mathrm{Q,min}}}^{z^{\mathrm{Q,max}}} \int_{z^{\alpha,\mathrm{min}}}^{z^{\alpha,\mathrm{max}}} dz_i^{\mathrm{Q}} dz_j^{\alpha} P(\delta_j, z_j^{\alpha}, z_i^{\mathrm{Q}} | r_{\parallel}^{A}) \delta_j \gamma_i. \tag{6.17}$$

We can expand this further as:

$$\langle \delta\gamma \rangle_A = \int_{-1}^{\infty} d\delta_j \int_{z^{\mathrm{Q,min}}}^{z^{\mathrm{Q,max}}} \int_{z^{\alpha,\mathrm{min}}}^{z^{\alpha,\mathrm{max}}} dz_i^{\mathrm{Q}} dz_j^{\alpha} P(\delta_j | z_j^{\alpha}, z_i^{\mathrm{Q}}) \times P(z_j^{\alpha} | z_i^{\mathrm{Q}}) P(z_i^{\mathrm{Q}}) \delta_j \gamma_i, \tag{6.18}$$

where we have made the dependence on $r_{\parallel}^{A}$ explicit. Expanding the first term in this equation, we get:

$$P(\delta_j | z_j^{\alpha}, z_i^{\mathrm{Q}}) = \frac{P(z_i^{\mathrm{Q}} | \delta_j, z_j^{\alpha}) P(\delta_j | z_j^{\alpha})}{P(z_i^{\mathrm{Q}} | z_j^{\alpha})} \tag{6.19}$$

$$= \frac{P(z_i^{\mathrm{Q}} | \delta_j, z_j^{\alpha}) P(\delta_j | z_j^{\alpha}) P(z_j^{\alpha})}{P(z_j^{\alpha} | z_i^{\mathrm{Q}}) P(z_i^{\mathrm{Q}})}. \tag{6.20}$$

Now we will make some approximations to evaluate the first two terms in the numerator of this equation - note that the first term in the denominator cancels with the second term in equation 6.18. We begin by writing $P(\delta_j | z_j^{\alpha}) \approx P(\delta_j)$. This is equivalent to ignoring the redshift evolution of δ, normally proportional to $(1 + z_j^{\alpha})^{\kappa}$, which we find does not change the shape of $\langle \delta\gamma \rangle_A$.

Next, we express the first term in equation 6.19 as $P(z_i^{\mathrm{Q}} | \delta_j, z_j^{\alpha}) = P(z_i^{\mathrm{Q}})(1 + \delta_j \xi^{\mathrm{X}}(\mathbf{r}^{\mathrm{X}}))$, where $\mathbf{r}^{\mathrm{X}}$ is the separation between z_j^{α} and z_i^{Q}, and ξ^{X} is the Lyα-quasar cross-correlation. Note that the dependence of $P(z_i^{\mathrm{Q}} | \delta_j, z_j^{\alpha})$ on z_j^{α} is accounted for by the redshift evolution of the cross-correlation. This is not an analytically derived expression, but rather an ansatz that produces the expected behaviour of $P(z_i^{\mathrm{Q}} | \delta_j, z_j^{\alpha})$. The definition of ξ^{X} here is somewhat analogous to the definition of the quasar autocorrelation as a measure of excess probability, $\xi^{\mathrm{Q}} = P(z_j^{\mathrm{Q}} | z_i^{\mathrm{Q}}) / P(z_j^{\mathrm{Q}}) - 1$, but with a modulating δ term. Making the subsequent substitutions into equation 6.18 gives:

$$\langle \delta\gamma \rangle_A = \int_{-1}^{\infty} d\delta_j \int_{z^{\mathrm{Q,min}}}^{z^{\mathrm{Q,max}}} \int_{z^{\alpha,\mathrm{min}}}^{z^{\alpha,\mathrm{max}}} dz_i^{\mathrm{Q}} dz_j^{\alpha} P(\delta_j) P(z_j^{\alpha}) P(z_i^{\mathrm{X}}) \\ \times \left(1 + \delta_j \xi^{\mathrm{X}}(\mathbf{r}^{\mathrm{X}})\right) \delta_j \gamma_i. \quad (6.21)$$

Multiplying the terms in brackets results in two parts (with and without the cross-correlation). The integration over δ is independent of the other two variables, and since it has mean 0, the part with one δ_j term vanishes. In the second term, $\int_{-1}^{\infty} P(\delta_j)\delta_j^2 d\delta_j$ evaluates to ~ 0.6. We introduce a free amplitude for this model $A_{\mathrm{cont}}^{\mathrm{auto}}$, which will absorb this value, but here we call this factor a_δ, and write our final expression as:

$$\langle \delta\gamma \rangle_A = a_\delta A_{\mathrm{cont}}^{\mathrm{auto}} \int_{z^{\mathrm{Q,min}}}^{z^{\mathrm{Q,max}}} \int_{z^{\alpha,\mathrm{min}}}^{z^{\alpha,\mathrm{max}}} dz_i^{\mathrm{Q}} dz_j^{\alpha}\, P(z_j^{\alpha}) P(z_i^{\mathrm{Q}}) \xi^{\mathrm{X}}(\mathbf{r}^{\mathrm{X}}) \gamma_i. \quad (6.22)$$

Qualitatively, we evaluate this model by iterating over correlation function bins $(r_\perp, r_\parallel)$. For a bin A (width $4\,h^{-1}$Mpc in our analysis), we have N bins of z_i^{Q} covering the redshift range of the quasars in the dataset. This results in an array of N×M pixel redshifts z_i^{α} (equation 6.16), using M bins of rest-frame wavelength in the Lyα forest region ($\lambda^{\mathrm{rf}} \in [1040, 1205]$Å). With these values we can compute $P(z_i^{\mathrm{Q}})$ from our normalised quasar redshift distribution, and the γ_i term for each rest-frame wavelength in M. Then, given z_i^{α} and $r_\parallel \in A$, we can derive z_j^{α} and the vector $\mathbf{r}^{\mathrm{X}}$ between z_j^{α} and z_i^{Q}. Finally, we compute the probability of z_j^{α} using a normalised pixel redshift distribution, and evaluate the cross-correlation $\xi(\mathbf{r}^{\mathrm{X}})$.

In figure 6.7, we present the model compared to the difference of correlation functions with and without redshift errors. To compute this, we directly measure $\gamma(\lambda)$ from the mock datasets (via equation 6.9), and input it to equation 6.22. In theory, one could also use a measurement of γ from mock datasets to fit our model on data. However, this relies on knowing the amount of redshift error present in the data, so we opt instead for building a simple model described in the next section.

6.4.3 Modelling γ

In real data we cannot directly measure $\gamma(\lambda^{\mathrm{rf}})$, since we don't know the true continuum (without redshift errors). Therefore, we need to construct a model of $\gamma(\lambda^{\mathrm{rf}})$.

Recalling equation 6.9, we see γ contains the ratio of the mean continuum in the presence of redshift errors ($\hat{\overline{C}}$), to the true mean continuum ($\overline{C}$). Therefore, to

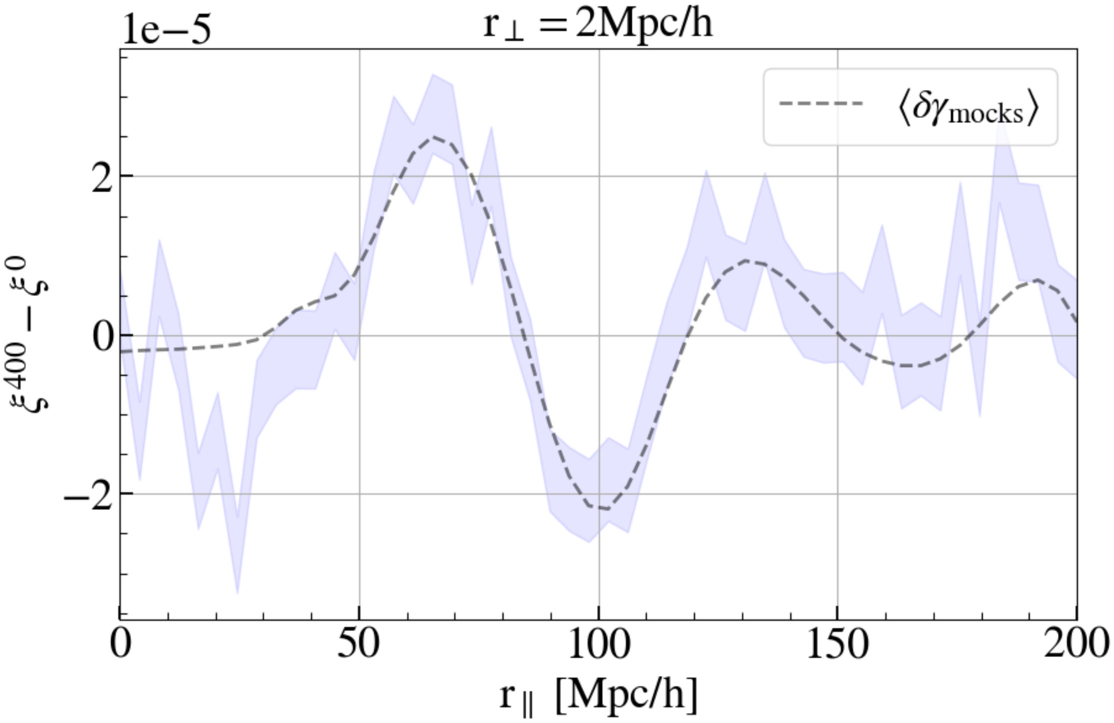

Figure 6.7: The contamination introduced by continuum redshift errors (blue shaded) as a function of $r_\parallel$, for the first (most contaminated) $r_\perp$ bin. The black dashed line is the contamination model (equation 6.22), estimated using a measurement of $\gamma(\lambda^{\rm rf})$.

construct a model of γ, we need a template quasar continuum and a parameter that controls the amount of redshift error to input. The continuum template is a convolution of a set of emission lines and a smooth broadband component. For the former we use the emission line properties of the composite model of BOSS spectra DR9 (Harris et al., 2016). We then approximate the smooth component as flat, motivated by the fact that the majority of the contamination we observe comes from the smoothing of emission lines. To produce $\hat{\overline{C}}$ from our template $\overline{C}$, we need to emulate the same smoothing. We do this by broadening the emission lines in our template $\overline{C}$ (listed in table 6.1), via a parameter $\sigma_{\rm cont}$ representing the amount of the velocity dispersion in kms^{-1} [6]. The width of each emission line is then broadened according to:

$$\hat{\sigma}_{\rm line} = \sqrt{\sigma_{\rm line}^2 + \left(\frac{\lambda_{\rm line}\sigma_{\rm cont}}{c}\right)^2}, \tag{6.23}$$

where $\sigma_{\rm line}$ is the intrinsic line width, $\lambda_{\rm line}$ is the rest-frame wavelength of the line, and c is the speed of light in kms^{-1}. Note that since the emission lines are

---

[6] The velocity dispersion here is related to redshift errors.

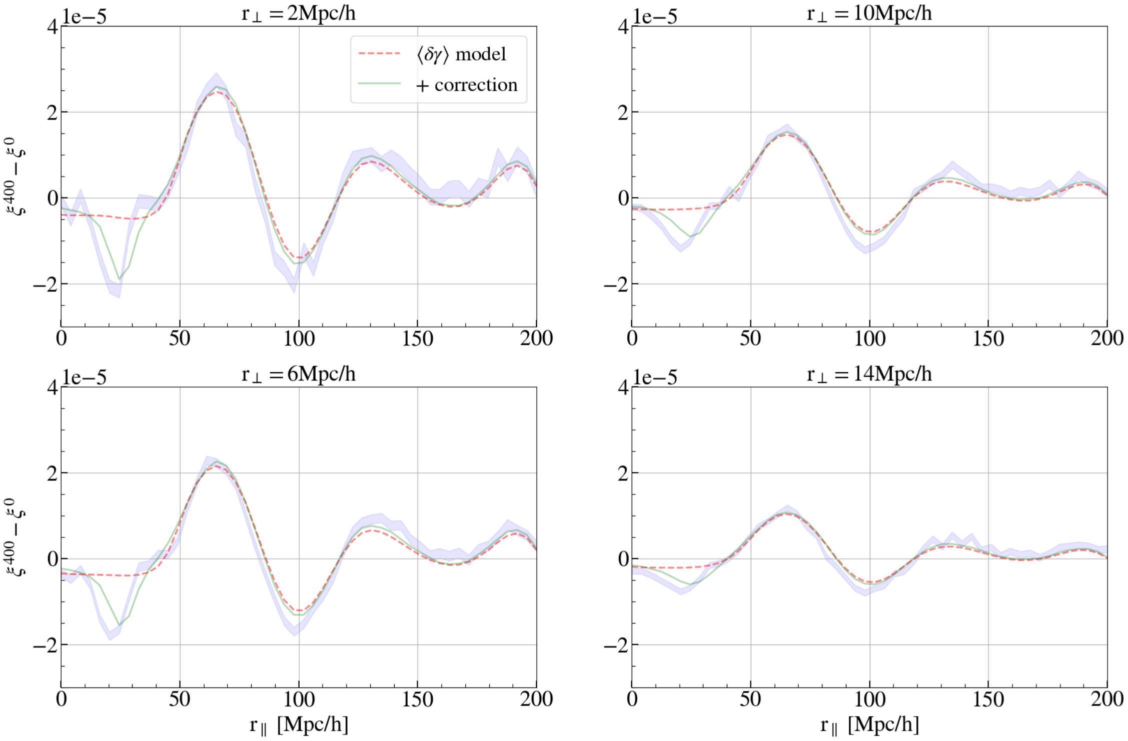

Figure 6.8: Direct fits to the contamination introduced by redshift errors, or the difference in the contaminated and uncontaminated Lyα autocorrelation, for the first 4 bins in $r_\perp$. We plot our model with (+correction) and without ($\langle \delta\gamma \rangle$ model) the addition of small-scale correction described in appendix 6.9. Note we use the $\gamma(\lambda^{\mathrm{rf}})$ model described in section 6.4.3 to compute $\langle \delta\gamma \rangle$, rather than a measurement from mocks. In this plot we show the stack of 100 DESI DR1 mocks, used throughout this analysis.

normalised Gaussians, the line amplitudes are $\propto 1/\hat{\sigma}_{\mathrm{line}}$. With the broadened emission lines we can construct $\hat{\overline{C}}(\lambda, \sigma_{\mathrm{cont}})$, and obtain $\gamma(\lambda)$. As mentioned earlier, this is robust to the choice of smooth component but highly dependent on the emission line model.

The result of fitting our model directly to the redshift error contamination is shown in figure 6.8 (using a model for γ). We show fits with the small-scale correction of appendix 6.9 (green solid line) and without (red dashed line), which we see are identical with the exception of the small-scale trough. In the full analysis (section 6.5) we choose not to use the small-scale correction model, to avoid biasing our results. We are still able to fit our model and recover the true input σ_{cont} without it, but may fail to capture the full impact of the trough at $\sim 25\,h^{-1}$Mpc. We see that the contamination is well fit at all scales, and that it drops off quickly with transverse separation.

6.4.4 Cross-correlation model

In Youles et al., 2022 they introduced a model of continuum redshift errors for the Lyα-quasar cross-correlation. The contamination in this case is characterised as: $\langle\hat{\delta}\rangle = \langle\delta\rangle - \langle\gamma(\lambda^{\mathrm{rf}})\rangle$, where $\langle\hat{\delta}\rangle = \hat{\xi}^{\mathrm{X}}$. The model of $\gamma(\lambda^{\mathrm{rf}})\rangle$ is then constructed using a similar method to the one in section 6.4.2, resulting in:

$$\langle\gamma\rangle_A = A^{\mathrm{cross}}_{\mathrm{cont}} \int\int dz^{\mathrm{Q}}_i dz^{\mathrm{Q}}_j P(z^{\mathrm{Q}}_i) P(z^{\mathrm{Q}}_j) \xi^{\mathrm{Q}}(\mathbf{r}^{\mathrm{Q}}) \gamma_i, \tag{6.24}$$

where there is now dependence on the quasar autocorrelation ξ^{Q}, rather than the Lyα-quasar cross-correlation. To include the Youles et al., 2022 model in our fits, we use the same γ model as in the autocorrelation (section 6.4.3). For this we have two free parameters, σ_{cont} and $A^{\mathrm{cross}}_{\mathrm{cont}}$. Note that in the cross-correlation the model of Youles et al., 2022 also fails to capture the behaviour of redshift errors at $\sim 25\,h^{-1}$Mpc, where instead of a trough we now see a peak (visible in figure 6.10). However, because we limit our cross-correlation to scales above $40\,h^{-1}$Mpc (see section 6.5), the impact of this peak is much smaller.

We ultimately combine both correlation functions in a joint fit, shown in section 6.5. This allows us to break degeneracies between certain nuisance parameters and key parameters like ϕ_{s}, α_{s}. As explained in Cuceu et al., 2021, we can also measure the combination $f\sigma_8$ from the joint fit. We expect γ to be the same for the auto- and cross-correlations, so we use only one parameter σ_{cont} for both. We have verified that allowing each correlation function to have a different set of parameters give perfectly consistent results.

6.5 Results

In this section we will explain our model fitting procedure, and show the impact of redshift errors on the cosmological parameters of our full-shape analysis. In particular, we discuss how large this contamination is with respect to the precision of DESI DR1 and DR2. We will then present the full-shape fits including the redshift errors model for the auto- (section 6.4.2) and cross-correlations (section 6.4.4). We show results of fits to these correlations individually, and together in a joint fit which gives constraints on the combination $f\sigma_8$.

6.5.1 Parameter degeneracies

The RSD parameter of the Lyα forest is given by:

$$\beta_\alpha = \frac{b_{\eta,\alpha(z)} f(z)}{b_\alpha(z)}, \tag{6.25}$$

where $b_{\eta,\alpha}$ is the velocity divergence bias of the Lyα forest and f is the logarithmic growth-rate. In practice we must assume a linear matter power spectrum template (Planck Collaboration et al., 2020), which has a fixed normalisation proportional to $\sigma_8(z)$ - the amplitude of perturbations in $8h^{-1}$Mpc spheres. In linear theory f and σ_8 are fully degenerate (Percival and White, 2009), so we are sensitive to the combinations $b_{\eta,\alpha} f \sigma_8$ and $b_\alpha \sigma_8$. Since $b_{\eta,\alpha}$ is unknown, we treat β_α as a nuisance parameter to marginalise over.

For the cross-correlation, we are sensitive to both the quasar and Lyα RSD and bias parameters. Since for quasars $\beta_\mathrm{q} = f/b_\mathrm{q}$, we are sensitive to: $b_\alpha \sigma_8$, $b_{\eta,\alpha} f \sigma_8$, $b_\mathrm{q} \sigma_8$ and $f\sigma_8$. $f\sigma_8$ is difficult to constrain from the cross-correlation alone, since it's dependent on several biases ($b_{\eta,\alpha}, b_\alpha, b_\mathrm{q}$). But we break these degeneracies by combining both correlation functions in a joint fit, as the autocorrelation provides precise measurements of $b_{\eta,\alpha} f \sigma_8$ and $b_\alpha \sigma_8$.

6.5.2 Fits

Following the full-shape analysis of DESI year-1 Lyα data (Cuceu et al., 2025a), we restrict the autocorrelation to $r \in [25, 180]\,h^{-1}$Mpc, and cross-correlation to $r \in [40, 180]\,h^{-1}$Mpc. This is a conservative minimum scale cut designed to limit the impact of increasingly non-linear scales and systematics (including redshift errors). We highlight at this point that there are some key differences between the analysis we perform here, and that of Cuceu et al., 2025a. First of all, as mentioned in section 6.2, Cuceu et al., 2025a use two additional correlation functions measured from Lyα absorption in Lyman-β region (between 920-1020 Å). These are the autocorrelation of Lyα absorption in the Lyman-β region, and its cross-correlation with quasars. We omit these correlations from our analysis for simplicity, since we would need to extend our continuum redshift errors model to incorporate this new region. We also do not expect redshift errors in the Lyman-β region to have a large impact overall, since these correlation functions have low statistical power relative to the two we analyse in this paper. Furthermore, in this paper we measure the AP effect separately on the broadband (ϕ_s) and peak component (ϕ_p) of the correlation function, as opposed to Cuceu et al., 2025a who measure the AP effect across the full correlation function with one parameter (ϕ_f). We opt for the former to characterise the impact of redshift errors more precisely, but it may result in a larger overall shift. Finally, we use a Gaussian prior on L_HCD, compared to Cuceu et al., 2025a who use a wide uniform prior. We do this to make fitting our new redshift error model parameters easier, but it may also lead to slight differences in our results.

Since we are only interested in our set of cosmological parameters $\{\alpha_\mathrm{p}, \phi_\mathrm{p}, \phi_\mathrm{s}, f\sigma_8\}$ and the parameters of the model introduced in this paper $\{\sigma_\mathrm{cont}, A^\mathrm{auto}_\mathrm{cont}, A^\mathrm{cross}_\mathrm{cont}\}$, we treat all other parameters as nuisance. Note that we have not included α_s in our parameters of interest, since it is not simple to extract cosmological from it.

In the autocorrelation fits, we have 15 free parameters when using the redshift errors model, and 13 without it. We fix β_HCD and L_HCD to values constrained in the joint fit, and set $f\sigma_8$ to value of the fiducial cosmology. For our model (equation 6.22), we also need the redshift distributions of Lyα pixels and quasars in our dataset, our gamma model $\gamma(\lambda^\mathrm{rf})$ and an input cross-correlation function ξ^X. Ideally, one would use a model (i.e. one that we fit in this analysis), but for simplicity and to reduce computing time we will use the measured cross-correlation from our mock dataset (figure 6.1). This is a reasonable approximation since the stack of mocks has very high signal-to-noise. For the DESI year-1 dataset, as we discuss in section 6.6, the cross-correlation is noisy and therefore is a not a good substitute for a model.

For the cross-correlation we have 16 free parameters with the redshift errors model and 14 without. We additionally fit for systematic redshift errors ($\Delta r_\parallel$), and smoothing at large $k_\parallel$ due to redshift errors and peculiar velocities (σ_v). However, we now fix $\sigma_\parallel$[7] to the joint fit value since it is degenerate with σ_v. For this we also provide a quasar redshift distribution, the continuum smoothing function $\gamma(\lambda^\mathrm{rf})$ (see section 6.4.3) and the quasar autocorrelation ξ^Q (section 6.4.4). Again we make a simplification by using a measurement of the quasar autocorrelation in-place of a model.

The joint fits have have 22 free parameters, combining the auto- and cross-correlations. Our full list of priors for all free parameters in the joint fit is given in table 6.4 in appendix 6.8. In most cases we use a wide uniform prior, except for the HCD length-scale (L_HCD) and RSD parameter (β_HCD), which where assign Gaussian priors based on previous studies (Pérez-Ràfols et al., 2018; Adame et al., 2025b). Our joint fits in both cases have an effective redshift $z \sim 2.3$. We compute our full-shape model using the `Vega`[8] library, which also contains a set of analysis tools. For the auto- and cross-correlation fits, we use the `Vega` interface for `Iminuit`, which performs a fast χ^2 minimisation. For the joint correlations we use the nested

---

[7] The line-of-sight smoothing due to limited simulation grid size.

[8] https://github.com/andreicuceu/vega

sampler `Polychord`[9] (Handley et al., 2015) to sample the posterior space of our parameters. The latter is slower, but is able to capture any non-Gaussianity in our posterior space, and gives better error estimates than a χ^2 minimisation.

In figure 6.9 we present the posteriors of ϕ, α, ϕ_s, $f\sigma_8$, from nested sampler runs on the joint correlation function for both the uncontaminated and contaminated datasets. For the latter we include two sets of contours, for runs including (red dashed) and not including (blue solid) the redshift errors model presented in this paper. We first note that the errors on these differences are very small, because we are analysing the stack of 100 DESI year-1 mocks. Later, when looking at the full set of results in table 6.2, we will compare our results to the precision of DESI DR1 (Cuceu et al., 2025a). We also note that the uncontaminated contours are slightly shifted from our fiducial cosmology (Planck Collaboration et al., 2020), but we will not discuss this here since we are specifically concerned about the impact of redshift errors (shift from green to blue). In appendix 6.8, we show the full triangle plot of our joint correlation run, and discuss what effect our model has on other nuisance parameters. We see that, in general, the shifts introduced by continuum redshift errors are small compared to the precision of DESI DR1. In this paper, we consider any shift greater than $\sigma_\text{DR1}/3$ to be significant, in-line with the requirements of Adame et al., 2025b. It should be noted however, that the shift we refer to here is only due to the impact of continuum redshift errors. In other words, this shift is not equivalent to a bias with respect to the fiducial cosmology (dashed crosshairs in figure 6.9).

From figure 6.9, we see ϕ_p is more affected than α_p, but the shift is still minor. The smooth AP parameter (ϕ_s) is most biased by continuum redshift errors because it depends on the broadband component of our correlation functions, which is more affected by contaminants in general. The observed shift may also be smaller because of the conservative minimum separations we choose for the auto- ($25\,h^{-1}$Mpc) and cross-correlations ($40\,h^{-1}$Mpc). A small but insignificant shift occurs in the $f\sigma_8$ posterior, which is also sensitive to the broadband of the correlation functions.

After introducing our model (red dashed contours), we see an impact on each of our posteriors. Firstly, we successfully recover the BAO peak parameters (α_p and ϕ_p) from the uncontaminated case. This result shows that our model can be used to correct for continuum redshift errors on BAO measurements, without considering a small-scale correction (e.g. equation 6.33).

The shift introduced by continuum redshift errors on ϕ_s is reduced significantly

---

[9] https://github.com/PolyChord/PolyChordLite

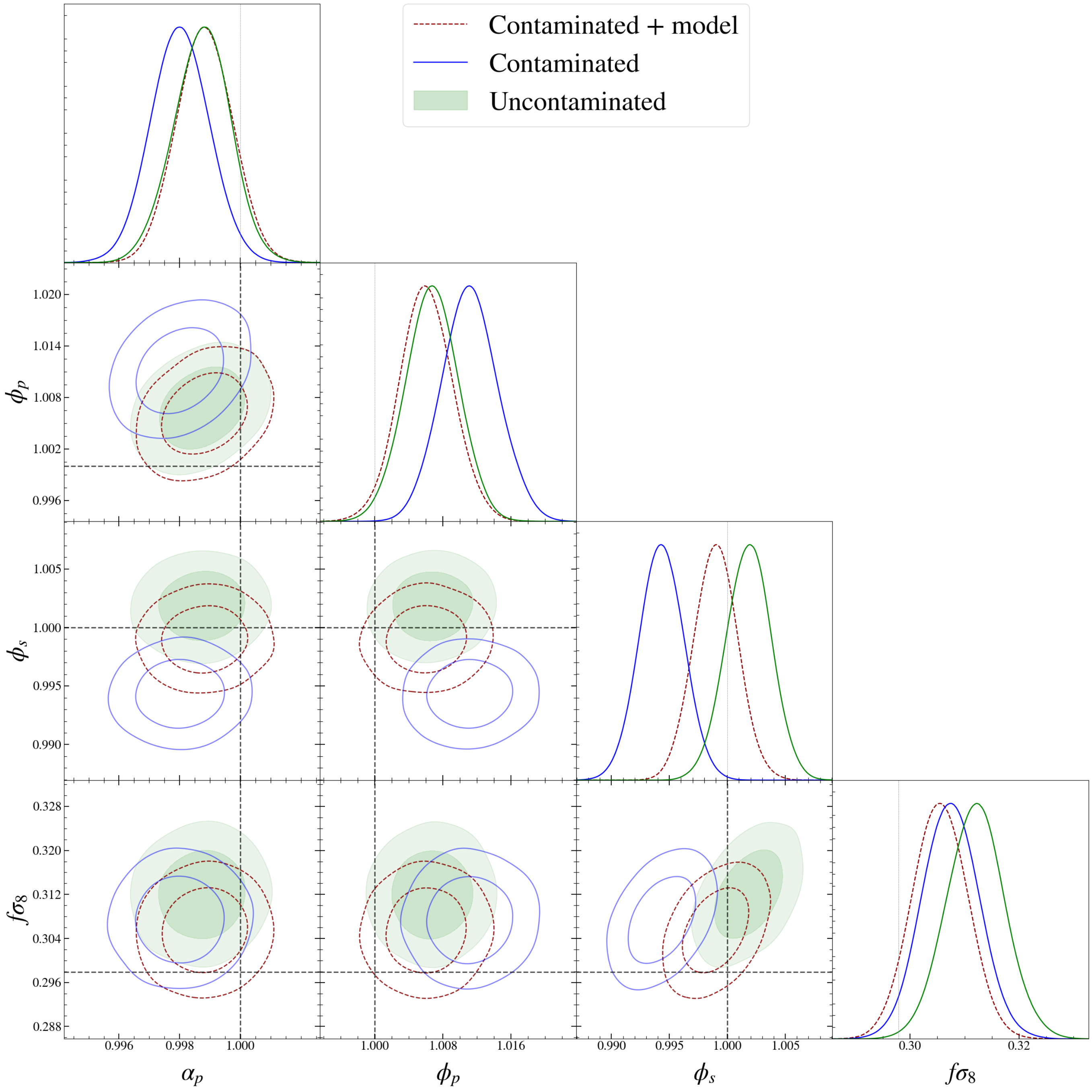

Figure 6.9: Full-shape posteriors from fits to a stack of 100 DESI DR1 mocks. We compare fit results from an uncontaminated (without redshift errors) dataset (green shaded), to fits on a contaminated dataset with (red dashed) and without our model (blue solid). Our fiducial cosmology is also indicated with black dashed crosshairs.

upon introduction of our model, but still differs slightly from the uncontaminated case. As discussed in section 6.4, our model does not capture the redshift error contamination present at $\sim 25\,h^{-1}$Mpc, which is responsible for the remainder of the shift. The $f\sigma_8$ constraint is shifted very marginally further from the uncontaminated case when using our model, towards the fiducial value (black dashed crosshairs). Indeed, it seems that all of our posteriors are shifted marginally towards the fiducial value, which suggests our new model is capturing some imperfections in the existing model. However, as we will show in appendix 6.8, our model parameters are not correlated with the cosmological parameters of interest, thus we are fitting for some other "nuisance" effect. The shift in fσ_8 is still small and consistent with the uncontaminated case to within 1-sigma.

In table 6.2 we show the differences in the cosmological parameters of our full-shape model, between analyses on uncontaminated and contaminated mocks. The columns under "+Model" give the same differences after adding our models of the contamination (section 6.4) to the full-shape fits. We ignore nuisance parameters constraints, except those of the model introduced in this paper, which we show in table 6.3. As noted earlier, we are analysing the stack of 100 DESI DR1 mocks with high signal-to-noise. For that reason, we include the projected precision of the DESI year-1 full-shape analyses on data in the final column of the table (see Cuceu et al., 2025a).

The shift on α_{p} (isotropic BAO parameter, equation 6.6) is never more than $\sim$ 10% of the DR1 68% confidence level ($0.1\sigma_{\mathrm{DR1}}$). For ϕ_{p} (AP effect on the BAO peak, section 6.2.4) and the growth-rate, we also see shifts of $\sim 0.1\sigma_{\mathrm{DR1}}$ in the joint fit. For each of the latter 3 parameters, the shifts introduced by redshift errors are notably consistent with zero to $\sim 1\sigma$. However, the shift on ϕ_{s} is $0.44(\pm 0.13)\sigma_{\mathrm{DR1}}$, which is larger than the $\sigma_{\mathrm{DR1}}/3$ criterion we set earlier. Due to large computing time, errors on these shifts are estimated from the stack of mocks, rather than the distribution of shifts across the 100 individual mocks. This approximation is justified because the shift is systematic in nature, and we are using the same mock realisations on both sides of the comparison, so the impact of noise is minimised.

While we have only included DESI DR1 projections, it is important to note that the relative sizes of the shifts shown here will increase significantly for the DESI DR2 and future datasets. The DR2 constraints for example should be approximately ~40% tighter than DR1, making the ϕ_{s} shift $\sim 0.7\sigma_{\mathrm{DR2}}$. It is also important to stress that this shift is not a bias with respect to the fiducial cosmology, rather it characterises the impact of a single contaminant. It may be the case that parametrising the AP effect over the full correlation function instead of the peak and broadband components separately, as is done in Cuceu et al., 2025a, reduces the impact of redshift errors.

The next 3 columns (+ model) show the same differences as the first 3, but including the continuum redshift errors model. For the peak component parameters (α_P, ϕ_P), our model is very successful in removing the bias for all 3 fit cases. For ϕ_s, our model successfully reduces the bias in the joint case to $0.19(\pm 0.13)\sigma_\mathrm{DR1}$, now consistent with zero to 1.5σ. Our model has almost no impact on $f\sigma_8$, but the shift is small $\sim 0.1\sigma_\mathrm{DR1}$ and consistent with zero to $\sim 1\sigma$. The parameter values of our model constrained in the analysis are shown in the section below.

As previously mentioned, we have a single set of parameters controlling the shape of γ: σ_cont, $A^\mathrm{auto}_\mathrm{cont}$ (autocorrelation amplitude) and $A^\mathrm{cross}_\mathrm{cont}$ (cross-correlation amplitude), which have values given in table 6.3. In each case we are able to recover the input redshift error (400 kms^{-1}) to within a 1σ confidence interval, and have strong detections of the other 3 parameters. The cross-correlation amplitude parameter is consistent between the cross and joint cases; however, this is not true for the autocorrelation amplitude. The tension between the auto and joint cases may reflect the incompleteness of the model at small-scales. For each of these parameters we apply wide uniform priors, shown in table 6.4 with the rest of the parameters in the joint fit.

The χ^2 probabilities of the analyses both with and without our model are very low. This is expected, since we are fitting an extremely high signal-to-noise dataset with a linear theory model (section 6.2.5) that is suboptimal. Furthermore, it has been shown in Cuceu et al., 2025a that our model is suitable for at least the level of precision of DESI DR1. We see a large improvement in χ^2 value when using our model in the autocorrelation ($\Delta\chi^2 = 655$, 1564 bins, 15 parameters), the cross-correlation ($\Delta\chi^2 = 726$, 3030 bins, 16 parameters) and the joint ($\Delta\chi^2 = 970$, 4594 bins, 22 parameters) correlation runs.

6.6 Discussion

In the previous section, we showed and discussed the results of fits to synthetic data sets with the model introduced in this paper. In section 6.6.1 we will discuss the application of this model to real data, and highlight potential differences with the synthetic datasets we use. We will also show in section 6.6.2, that the effect of continuum redshift errors can be mitigated by removing pairs where the quasar autocorrelation (for contamination in the Lyα-quasar cross-correlation) or the Lyα-quasar cross-correlation (for contamination in the Lyα autocorrelation) is large.

Parameter	Contaminated fit			+ Model			$\sigma_{\rm Y1}$
	auto	cross	joint	auto	cross	joint	
$10^3\Delta\alpha_{\rm p}$	0.1 ± 1.4	1 ± 1.2	0.8 ± 1.0	-0.2 ± 1.4	0.03 ± 1.2	-0.1 ± 0.9	11
$10^3\Delta\phi_{\rm p}$	-8 ± 5	-4 ± 4	-4 ± 3	-5 ± 5	1 ± 4	-0.7 ± 3	38
$10^3\Delta\phi_{\rm s}$	4 ± 2.4	10 ± 4	-7 ± 2	3 ± 2.4	10 ± 3	-3 ± 2	16
$10^2\Delta f\sigma_8$	-	-	0.5 ± 0.5	-	-	0.6 ± 0.5	6.2

Table 6.2: Differences in parameter constraints on fits to mocks with and without redshift errors. We fit a stack of 100 DESI DR1 contaminated mocks, without (Contaminated fit) and with (+ Model) our continuum redshift errors model, and measure the shift with respect to the uncontaminated set. The last column shows the projected errors (68% confidence interval) on the same parameters for the joint constraints from DESI DR1 dataset. Note that the values presented here are shifts introduced solely by continuum redshift errors, not the bias with respect to the fiducial cosmology.

Parameter	auto	cross	joint
$\sigma_{\rm cont}[{\rm kms}^{-1}]$	401 ± 19	406 ± 11	411 ± 18
$A_\gamma^{\rm auto}$	-6.6 ± 0.4	-	-8.9 ± 0.4
$A_\gamma^{\rm cross}$	-	-6.2 ± 0.3	-5.8 ± 0.2

Table 6.3: Constraints on free parameters of the model we introduce to capture continuum redshift errors (section 6.4), from a fit to a stack of 100 DESI DR1 contaminated mocks. We show the constraints (section 6.4) for the auto-, cross- and joint correlations separately. The input amount of redshift error was $\sigma = 400\,{\rm kms}^{-1}$, which we recover to within 1σ in the each case.

6.6.1 Contamination on real data

In Youles et al., 2022, they found a strong dependence of $\langle\gamma\rangle$ on small-scale quasar clustering. Likewise, we verify that $\langle\delta\gamma\rangle$ has strong dependence on the small-scale cross-correlation function. This is discussed further in section 6.6.2. We also find that the quasar clustering in our lognormal mocks deviates more from linear theory at small-scales than more realistic n-body simulations (see figure 6 of Youles et al., 2022), which is likely to increase the amount of contamination relative to real data.

As mentioned in section 6.5.2, we use a measurement of the cross-correlation in our $\langle\delta\gamma\rangle$ model, and a measurement of the quasar autocorrelation in $\langle\gamma\rangle$. For our mock datasets, where we have high signal-to-noise these measurements are smooth, and are a better alternative to using a model of the respective correlation functions. However, in e.g. DESI DR1, our correlations are relatively noisy and, in the case of $\hat{\xi}^X$, and are themselves affected by redshift errors. It is also important to note that because of our strong dependence on small-scale clustering, using a linear model of ξ^X in equation 6.22 (or ξ^Q in equation 6.24) would likely be a bad approximation. For these reasons, we choose to leave a full analysis on data for future work where we have a higher signal-to-noise data set, or a cross-correlation (and quasar autocorrelation) model that is realistic at small-scales. At the level of precision of DESI DR1, our model is also not yet necessary to perform an un-biased analysis (i.e. Cuceu et al., 2025a).

Another contributing factor in our analysis realism is our synthetic spectra, or specifically the quasar continua that we use. As we discussed in section 6.2, we generate quasar spectral templates using convolutions of power law smooth components, and emission lines (table 6.1) from the composite spectrum of BOSS quasars (Harris et al., 2016; McGreer et al., 2021). We sample from a Gaussian distribution of power law slopes, tuned to have mean and scatter that better reflected the eBOSS DR16 dataset (Bourboux et al., 2020). Likewise, for the Lyα forest emission lines we sample from a Gaussian distribution of EWs, tuned on BOSS DR9 data. The distribution of emission line EWs is particularly important for our study, since the dominant contribution to the contamination is around the positions of these. If, for example, the mean EW in our spectra was higher than in e.g. DESI DR1, we would observe a higher level of contamination. Also, since the scales of the spurious correlation features are directly related to the rest-frame wavelength of the emission lines in our Lyα forests (see section 6.4.1), we are sensitive to the relative strength of individual lines.

The way we introduce redshift errors into our spectra should also reflect the magnitude and nature of redshift errors in real data. However, this is difficult to emulate because there are several ways in which redshift errors are produced. We add errors drawn from a Gaussian with dispersion $\sigma = 400\,\mathrm{kms}^{-1}$ to each quasar in our sample, representing errors which might occur in our redshift estimation pipeline `Redrock`. However, as mentioned in section 6.1, it is also typical for emission lines to shift away from the systemic redshift of the quasar. The level of shift depends on the emission line, but is larger in general for broad, high-ionisation lines like CIV, SiIV and CIII. Therefore, the level of error depends on the set of lines being used to estimate the quasar redshift. There has been progress made in this area recently, with work on improving spectral templates in DESI (Brodzeller et al., 2023; Bault et al., 2025).

There is further complication due to the fact that the contaminating lines within the Lyα forest are high-ionisation. If these are strongly correlated with the high-ionisation lines used to estimate quasar redshift, then there would be little impact on the mean continuum. The latter point requires further studies like the one of Shen et al., 2016; Brodzeller et al., 2023; Bault et al., 2025, but is difficult in practice because Lyα forest lines are less prominent, and visible at a higher redshift ($z \gtrsim 2.1$). One could improve the way we add redshift errors to mocks using these studies to input typical relative line shifts into our spectra, and use a redshift estimator (i.e. `Redrock`) to estimate our redshifts.

6.6.2 Mitigating contamination in real data

Youles et al., 2022 identified that the level of redshift error contamination in the cross-correlation between Lyα and quasars, depends on the correlation between the host quasar of the forest and the correlating quasar. This correlation is larger for neighbouring pairs of quasars, and avoiding these configurations significantly reduces the contamination. Here we extend this further by explicitly removing Lyα-quasar pairs where the forest host-quasar and correlating quasar are separated by less than a given separation. The situation is similar for the Lyα autocorrelation, except that, as we showed in section 6.4.2, the contamination is now dependent on the Lyα-quasar cross-correlation.

The result of removing contributions to our model from pairs where the host quasar of a Lyα pixel and the correlating quasar are separated by less than r_{q}, is shown in figure 6.10 (bottom). We choose two cuts at $5\,h^{-1}$Mpc and $10\,h^{-1}$Mpc, which removes only $\sim 0.1 - 0.3\%$ of our total number of pairs, but $\sim 10\%$ in the first 5 transverse bins (up to $r_{\perp} \sim 20\ h^{-1}$Mpc). From figure 6.10, we see that by $10\,h^{-1}$Mpc, the contamination is effectively negligible.

In the top panel of figure 6.10 we show an analogous test with the contamination in the Lyα autocorrelation. Here we remove pairs where the host quasar of one forest is separated from the correlating pixel in the other forest by less than $r_{\mathrm{X}} = 5, 10\,h^{-1}$Mpc. In this case, we see a similar same level of success in removing the contamination. The fraction of correlating pairs that is removed in this test is also sub-percent level overall, but $\sim 10\%$ in the most line-of-sight bins.

These tests were undertaken in Casas et al., 2025; DESI Collaboration et al., 2025b and shown to have very little impact on the precision of BAO constraints, where the most line-of-sight bins are at much smaller separation than the BAO feature. However, there was a significant impact on the Lyα bias, which may suggest this approach is not suitable for full-shape analyses. We leave this to be studied in future work, as a possible alternative to the model we present in this paper.

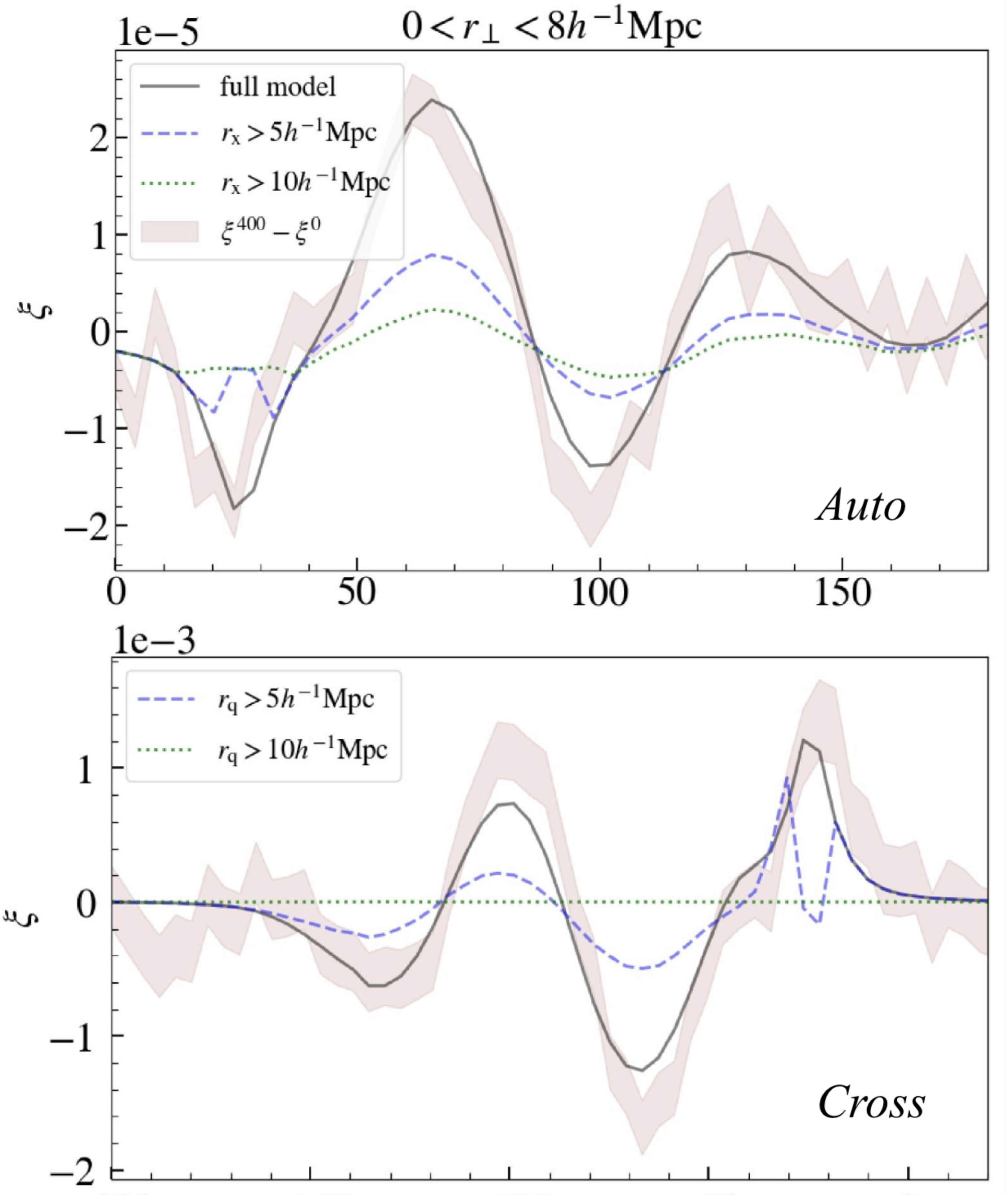

Figure 6.10: (Top) the contamination in the Lyα autocorrelation introduced by redshift continuum errors and our model of it (black solid). Also plotted is the same model evaluated only for separations r_X greater than 5 h^{-1}Mpc (blue dashed) and 10 h^{-1}Mpc (green dotted), where r_X is the separation between one pixel and the host quasar of the other pixel. (Bottom) continuum redshift errors contamination in the Lyα-quasar cross-correlation, over-plotted with the model of the same contamination. The model is evaluated with the same cuts as the autocorrelation, but with r_q now referring to the separation between a correlating quasar and the host quasar of the correlating Lyα forest pixel. In both cases we have used the small-scale correction of appendix 6.9.

6.7 Summary

Random errors in our quasar redshift measurements smooth the mean continuum used to estimate the Lyα flux transmission field. This smoothing gives rise to spurious features across the length of the Lyα forest, which in-turn distort the full-shape of the Lyα autocorrelation and its cross-correlation with quasars.

In this paper we presented a model of these spurious correlations for the Lyα autocorrelation function, building upon work in Youles et al., 2022, where they presented an equivalent model for the Lyα-quasar cross-correlation. We then created a model for the distortion as a function of rest-frame wavelength ($\gamma(\lambda^{\rm rf}, \sigma_{\rm cont})$), and introduced three new parameters to our standard full-shape model. These control the amount of smoothing introduced by redshift errors ($\sigma_{\rm cont}$), and the amplitude of the effect in the auto- ($A_\gamma^{\rm auto}$) and cross-correlation functions ($A_\gamma^{\rm cross}$).

We show using a high signal-to-noise synthetic dataset that continuum redshift errors shift the measurements of the growth rate ($f\sigma_8$), isotropic BAO parameter ($\alpha_{\rm p}$), anisotropic BAO parameter ($\phi_{\rm p}$) and anisotropic broadband parameter ($\phi_{\rm s}$) to varying degrees, with respect to a dataset which does not contain redshift errors. At the level of precision of DESI DR1, this shift is relatively minor in the first 3 parameters (~10%). However, we find that $\phi_{\rm s}$ is shifted from the uncontaminated dataset (without redshift errors) value by $0.44(\pm 0.13)\sigma_{\rm DR1}$, where $\sigma_{\rm DR1}$ is projected the 68% confidence region of the DESI DR1 constraint. Note that this relative shift will increase as analyses on future datasets (e.g. DESI DR2) improve the constraining power on these parameters. We demonstrated that our model reduces the shift introduced to $\phi_{\rm s}$ by $\sim 60\%$, and completely removes any shift on $\alpha_{\rm p}$ and $\phi_{\rm p}$. We also recover to within 1σ the input dispersion of the Gaussian distribution of redshift errors we add ($400\,{\rm kms}^{-1}$), and strongly detect each of the amplitude parameters of the model.

On real data, one should use models of the Lyα-quasar cross-correlation and the quasar autocorrelation as inputs for the redshift errors model in the autocorrelation ($\langle\delta\gamma\rangle$) and the cross-correlation ($\langle\gamma\rangle$) respectively. This is because measurements of these functions are noisy compared to the high signal-to-noise stack of mocks we used throughout this work. We also show that continuum redshift error contamination could be mitigated in real data by removing pairs in the cross-correlation where the quasar autocorrelation is strongest, and removing pairs in the autocorrelation where the cross-correlation is strongest. In section 6.6.2, we showed that removing pairs where $r_{\rm q}, r_{\rm X}$[10]$< 10\,h^{-1}$Mpc is enough to make $\langle\gamma\rangle$

---

[10] The separation between quasars in the quasar autocorrelation, and Lyα pixels and quasars in

and $\langle \delta\gamma \rangle$ negligible. This corresponds to $\sim 0.3\%$ of total correlating pairs, but $\sim 10\%$ of the most line-of-sight (up to $r_\perp \sim 20\,h^{-1}$Mpc) pairs, meaning this cut may be suitable for removing the effect of redshift errors on BAO constraints, but not for full-shape analyses.

6.8 Appendix: full model results

In section 6.5 we showed the posteriors of our set of cosmological parameters from nested sampler runs, for three different cases: uncontaminated (no redshift continuum errors, baseline model), contaminated (baseline model) and contaminated (baseline + redshift errors model). We showed that the model introduced in this paper (see section 6.4) successfully removed the effect of continuum redshift errors on the BAO peak parameters (ϕ_p,α_p). It also removes the majority of the shift on ϕ_s, but has little impact on $f\sigma_8$.

In table 6.4, we show the priors on parameters in the joint fit. In the first section, we show the cosmological parameters of interest, and in the second we show the nuisance parameters of the baseline model. In the final section, we show the parameters introduced in this paper to capture the effect of continuum redshift errors. For most of the parameters, we choose conservative uniform priors, except in the case of β_HCD and L_HCD, where we follow Adame et al., 2025b and place more informative priors based on HCD studies to improve the performance of the sampler. The priors placed on parameters in our model are physically motivated, i.e. we can't have negative velocity dispersions, or positive values for the amplitude of the effect (see section 6.4). The value of σ_cont directly reflects the size of the redshift errors in our dataset, meaning the upper bound of 2000 kms^{-1} would be very extreme.

In figure 6.11, we show the triangle plot of posteriors for each parameter in our sample runs. Looking first at the parameters introduced in this paper ($A^\mathrm{cross}_\mathrm{cont}$, $A^\mathrm{auto}_\mathrm{cont}$, σ_cont), we see that they are mostly uncorrelated with all of the other parameters. There is a correlation between σ_cont and the amplitude parameters ($A^\mathrm{cross}_\mathrm{cont}$, $A^\mathrm{auto}_\mathrm{cont}$), due to the fact that broadening the (Gaussian) emission line profiles in the Lyα forest also reduces their amplitude. We can also see that while our model successfully reduces the bias on the key cosmological parameters, it often fails to recover the value of other nuisance parameters in the uncontaminated case, and in one case ($b_\mathrm{SiII(1206)}$), is significantly more discrepant than the fit to the contaminated dataset without our model. One of the main reasons for both

---

the cross-correlation respectively.

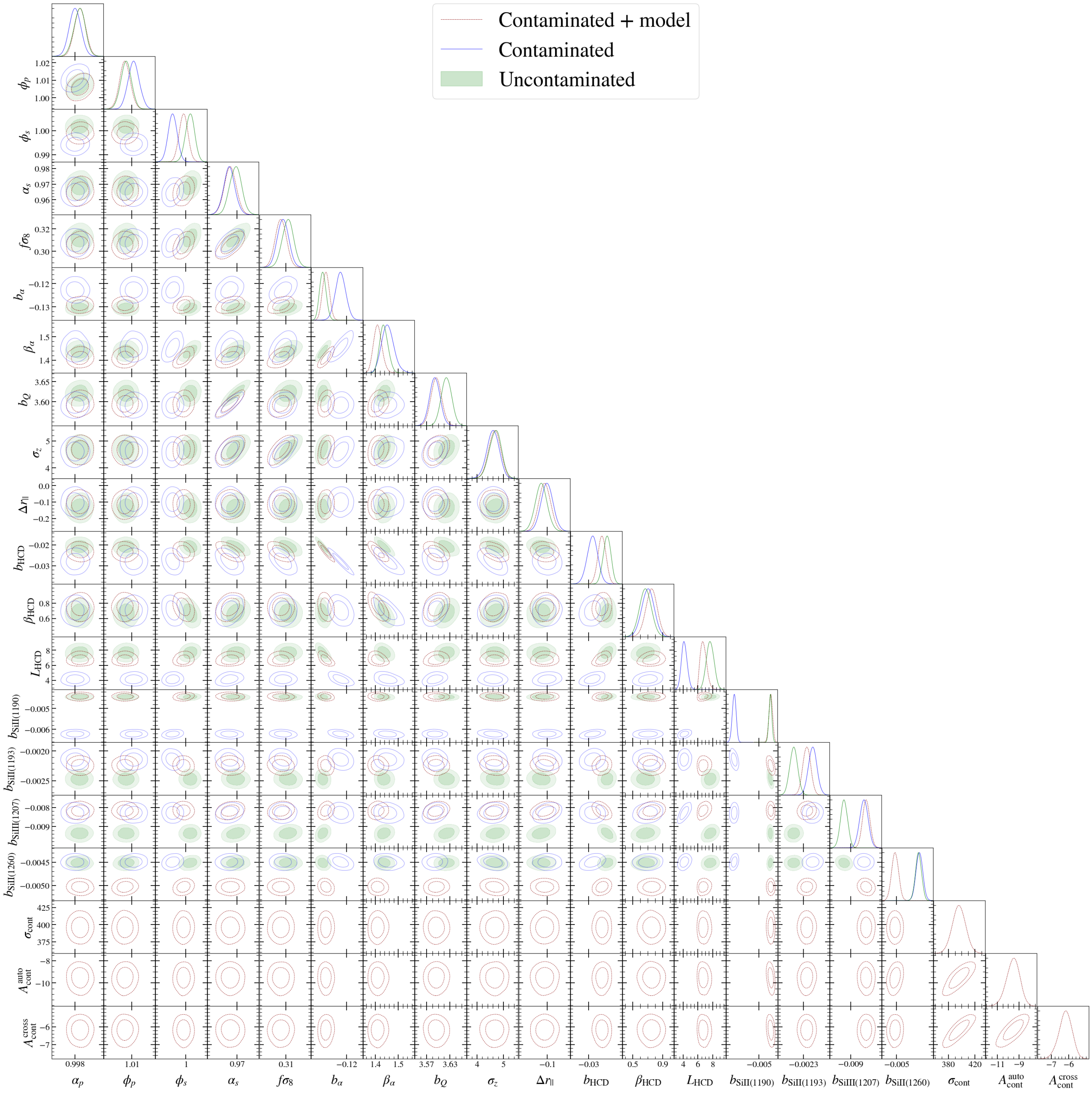

Figure 6.11: Posteriors of all free parameters in our sampler runs on the joint auto+cross correlation function, measured from a stack of 100 DESI DR1 mocks. The uncontaminated (without continuum redshift errors) run is shown in green, and the contaminated runs with and without the redshift errors model are shown in dark red and blue respectively. We exclude the smoothing parameters $\sigma_\parallel$ and $\sigma_\perp$, since they marginalise the effect of grid size in our simulated data sets, and are not relevant to real data.

Parameter	Prior
α, ϕ^*	$\mathcal{U}$[0.01,2]
$f\sigma_8$	$\mathcal{U}$[0,2]
b_α	$\mathcal{U}$[-2, 0]
β_α	$\mathcal{U}$[0, 5]
b_q	$\mathcal{U}$[0, 10]
σ_v	$\mathcal{U}$[0, 15]
$\Delta r_{\Vert}$	$\mathcal{U}$[-3, 3]
b_HCD	$\mathcal{U}$[-0.2, 0]
β_HCD	$\mathcal{N}$[0.5, 0.09]
L_HCD	$\mathcal{N}$[5, 1]
$b_\mathrm{SiII(1190)}$	$\mathcal{U}$[-0.02, 0.02]
$b_\mathrm{SiII(1193)}$	$\mathcal{U}$[-0.02, 0.02]
$b_\mathrm{SiIII(1207)}$	$\mathcal{U}$[-0.02, 0.02]
$b_\mathrm{SiII(1260)}$	$\mathcal{U}$[-0.02, 0.02]
$\sigma_{\Vert}$	$\mathcal{U}$[0, 10]
$\sigma_{\perp}$	$\mathcal{U}$[0, 10]
σ_cont	$\mathcal{U}$[0, 2e3]
$A_\mathrm{cont}^\mathrm{auto}$	$\mathcal{U}$[-50,0]
$A_\mathrm{cont}^\mathrm{cross}$	$\mathcal{U}$[-50,0]

Table 6.4: Priors on free parameters used in the joint fits to our stack of 100 DESI DR1 contaminated mocks. Recall that there are two α and two ϕ parameters, corresponding to the BAO peak and broadband of the correlation function.

of these issues is the large parameter space with, in some cases, high levels of degeneracy. These differences in nuisance parameters result in the very marginal over-fitting we see in figure 6.9, but really only serve to indicate that there are some systematic signals that we do not account for with our previous baseline.

6.9 Appendix: mean continuum expansion

For a forest of a given quasar q, the delta field is evaluated at λ in the observed frame. The mean continuum $\overline{C}$ is a function of rest-frame, which we can write as a function of λ and z_q as:

$$1 + \delta(\lambda) = \frac{f(\lambda)}{\hat{\overline{C}}(\frac{\lambda}{1+\hat{z}_\mathrm{q}})(a_\mathrm{q} + b_\mathrm{q}\Lambda)}, \tag{6.26}$$

where Λ is a function of $\log\lambda$ given in equation 6.1, and $\hat{\overline{C}}$ is the measured mean continuum in the presence of redshift errors. We can see from this equation that δ will be contaminated by the smoothed $\hat{\overline{C}}$, but also that the conversion from λ^rf

to λ will be incorrect, given we measure a quasar redshift that has some error, $\hat{z}_{\rm q} = z_{\rm q} + \epsilon$. We can therefore expand the mean continuum about the true redshift as:

$$\hat{\overline{C}}(\hat{\lambda}^{\rm rf}) = \hat{\overline{C}}(\lambda^{\rm rf}) + \epsilon \frac{d\hat{\overline{C}}}{dz_{\rm q}} + \frac{\epsilon^2}{2} \frac{d^2\hat{\overline{C}}}{dz_{\rm q}^2} + \ \ldots \tag{6.27}$$

Now, we re-write the differential with redshift as:

$$\frac{d\hat{\overline{C}}}{dz_{\rm q}} = \frac{d\lambda^{\rm rf}}{dz_{\rm q}} \frac{d\hat{\overline{C}}}{d\lambda^{\rm rf}} = -\frac{\lambda^{\rm rf}}{1 + z_{\rm q}} \frac{d\hat{\overline{C}}}{d\lambda^{\rm rf}}. \tag{6.28}$$

Substituting equations 6.27 and 6.28 into equation 6.9, we can derive a new expression for our measured transmission field $\hat{\delta}_{\rm q}$:

$$\hat{\delta}_{\rm q} = \delta_{\rm q} - \gamma + \gamma_{\rm z}, \tag{6.29}$$

where as usual we ignore 2nd order terms, and

$$\gamma_{\rm z} = \frac{\lambda^{\rm rf} \epsilon}{\overline{C}(1 + z_{\rm q})} \frac{d\hat{\overline{C}}}{d\lambda^{\rm rf}}. \tag{6.30}$$

Now, as we did in section 6.4, we propagate our expression for our measured delta through to the cross- and autocorrelation functions:

$$\langle \hat{\delta} \rangle = \langle \delta \rangle - \langle \gamma \rangle + \langle \gamma_{\rm z} \rangle \tag{6.31}$$

$$\langle \hat{\delta}\hat{\delta} \rangle = \langle \delta\delta \rangle - 2\langle \delta\gamma \rangle + 2\langle \delta\gamma_{\rm z} \rangle - 2\langle \gamma\gamma_{\rm z} \rangle + \langle \gamma\gamma \rangle + \langle \gamma_{\rm z}\gamma_{\rm z} \rangle. \tag{6.32}$$

Modelling the additional terms analytically may prove to be difficult, but as we did for figure 6.4, we can use mock datasets to measure each term in equation 6.31. In figure 6.12, From this measurement, we see that only $\langle \delta\gamma_{\rm z} \rangle$ is non-negligible, and that this is roughly constant across $r_{\parallel}$ for the usual range of $r_{\perp} (\lesssim 20\,h^{-1}{\rm Mpc})$. An explanation for the trough feature $\sim 25\,h^{-1}$Mpc is therefore still required, where perhaps higher-order terms cannot be ignored.

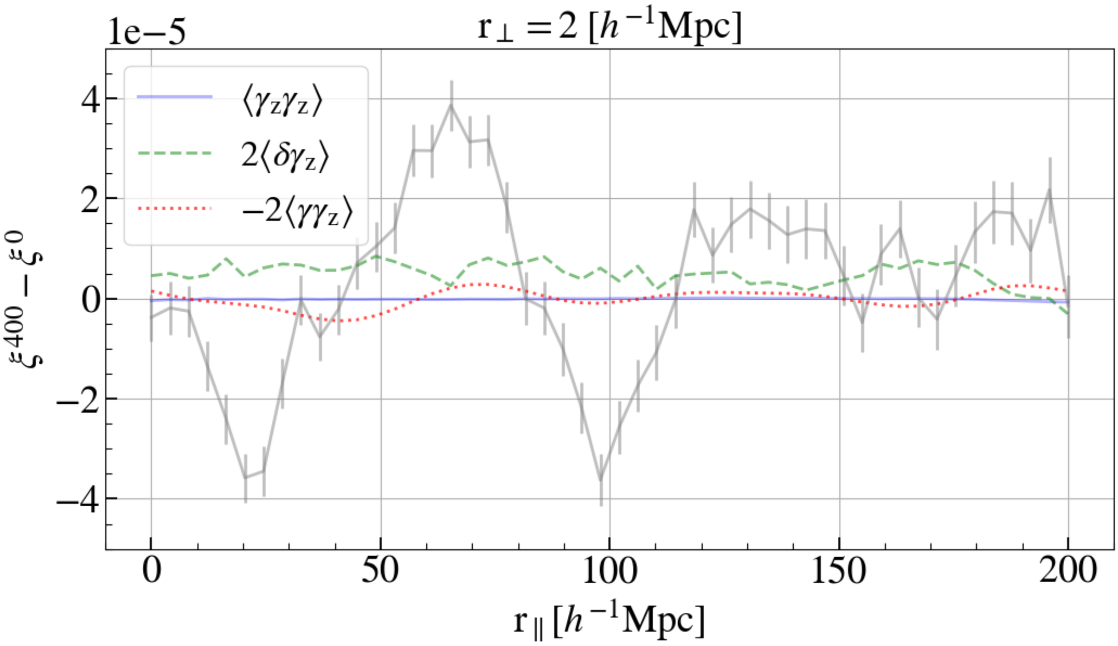

Figure 6.12: Contaminating terms in the Lyα autocorrelation, due to rest-frame wavelength grids shifts in the presence of quasar redshift errors. γ_z is defined in equation 6.30, and is the result of expanding the measured mean continuum $\hat{\overline{C}}$ about small redshift errors ϵ. The contamination in the first transverse bin is shown in grey, as a function of line-of-sight separation ($r_{||}$).

6.9.1 Small-scale contamination model

Here, we try to model the 25 h^{-1}Mpc contaminating trough feature of our fiducial analysis (see e.g. figure 6.4) empirically. We do this by adding an extra term to the original γ function (equation 6.9):

$$\gamma(\lambda^{\mathrm{rf}}) = \overline{\gamma_0}(\lambda^{\mathrm{rf}}, \sigma_{\mathrm{cont}}) + \frac{a_1}{(\lambda^{\mathrm{rf}} - \lambda_1^{\mathrm{rf}})^2}, \tag{6.33}$$

where the function is offset by $\lambda_1^{\mathrm{rf}} = 1207.6$ Å, and $\overline{\gamma_0}$ is the function we model in the previous section. This value was constrained in a direct fit to the difference in the Lyα autocorrelation with and without continuum redshift errors, together with a_1, the redshift error parameter (σ_{cont}, section 6.4.2) and a free amplitude. Note this function diverges for $\lambda^{\mathrm{rf}} = \lambda_1^{\mathrm{rf}} = 1207.6$ Å, but this is not an issue since in our analysis we always have $\lambda^{\mathrm{rf}} < \lambda_1^{\mathrm{rf}}$.

We show in figure 6.8 that in a direct fit to the redshift error contamination, our additional correction captures the trough feature very well. However, when performing a full-shape fit to our auto- and cross-correlations we find that a_1 takes much larger than expected values [11]. This consequently biased our measurements

---

[11] Our expected value of a_1 is computed in a direct fit to the contamination, where the only free

of ϕ_s, so it was decided to proceed for now without the additional correction.

---

parameters are σ_{cont}, A_{cont} and a_1.

7

LYMAN-α BAO FORECASTS IN DESI-II

The DESI-II survey is a proposed successor to the DESI survey, due to start its 6-year observing programme in 2029. DESI operations will extend for three years beyond the nominal DESI observing period (2021-2026), in a programme called DESI-extension (DESI-ext; 2026-2029). In this work we will mostly make comparisons between the nominal DESI survey and DESI-II.

DESI-II will provide a complementary dataset to DESI, by observing 5000 $\deg^2$ of sky for three new surveys: the high-redshift survey, high-density survey and the dark matter survey. The latter two surveys focus on the Universe at $z \lesssim 1$ and the local group respectively, allowing new tests of gravity and cosmic structure at fine detail, as well as studies of the nature of dark matter halos and constraints on dark matter physics. The high-redshift survey, which we will focus on here, is a study of the Universe at $z \gtrsim 2$ which will target more than three million Lyman-α emitters (LAEs) and Lyman break galaxies (LBGs). These new targets will provide better BAO measurements than DESI quasars at even higher redshifts, giving us sub-percent constraints on various dark energy models. They will also allow us to study in even greater detail primordial non-Gaussianities and neutrino masses. In addition to the new targets, DESI-II will also observe quasars to greater depth than DESI, increasing the target density from 60 to $\sim 100\,\deg^{-2}$.

LBGs can be used both as tracers in clustering measurements, and as sources of Lyα absorption from which we can perform clustering measurements. LAEs on the other hand are only useable as tracers, because they have very faint and often unobservable continua. In this work we create a Fisher forecast to calculate the expected level precision on the transverse and parallel BAO scaling parameters, $\alpha_\perp$ and $\alpha_\parallel$[1], attainable from these tracers and quasars.

---

[1] Equivalent to forecasting the precision on D_M/r_d and $H(z)r_d$ respectively.

We forecast BAO from several measurements, including the auto-power spectrum of Lyα fluctuations from LBGs (hereby `Ly`α`(LBG)`$^{\text{auto}}$), the Lyα(LBG)-LBG cross-power spectrum (`Ly`α`(LBG)`$^{\text{cross}}$), the Lyα(LBG)-LAE cross-power spectrum (`Ly`α`(LBG)xLAE`), and the LBG-LBG auto-power (`LBG`$^{\text{auto}}$) and LAE-LAE auto-power (`LAE`$^{\text{auto}}$) spectra. We also include the cross-power spectrum of Lyα fluctuations in quasar spectra with LBGs (`Ly`α`(QSO)xLBG`) and LAEs (`Ly`α`(QSO)xLAE`), to take advantage of the higher forest SNR in quasars, and the significantly larger target density of LBGs and LAEs. Note that the Lyα forests from quasars are from the combined DESI + DESI-II projections.

We first begin by outlining the method by which we perform our forecast in section 7.1, introducing the Fisher formalism, and the covariance matrices used to estimate it. Then in section 7.2 we discuss the set of survey conditions and results of our DESI-II forecast.

7.1 Fisher forecast

The Fisher information matrix F_{ij} is defined in terms of the log-likelihood ($\mathcal{L}$) of parameters θ as:

$$F_{ij} = \left\langle \frac{\partial^2 \mathcal{L}}{\partial\theta_i \partial\theta_j} \right\rangle . \tag{7.1}$$

If our likelihood is Gaussian, we can also derive an expression for this in terms of the covariance matrix (Tegmark, 1997):

$$F_{ij} = \frac{1}{2}\mathrm{Tr}\left(C^{-1}\frac{\partial C}{\partial\theta_i} C^{-1}\frac{\partial C}{\partial\theta_j} \right) , \tag{7.2}$$

where $C_{ij}(\theta) = \langle \delta_i \delta_j \rangle$, for a Gaussian distributed data vector δ.

The Cramer-Rao limit (Rao, 1945) gives the lower bound on the variance of any unbiased estimator $\hat{\theta}$ of a parameter θ. This limit can be equivalently expressed in terms of the inverse of the Fisher matrix:

$$\sigma_i \geq (F^{-1})_{ii} . \tag{7.3}$$

Thus, using this relation we can determine the best possible constraints on a set of cosmological parameters given any covariance matrix and its dependence on said parameters. The question we will address in the following sections is how we generate $C(\theta)$ for a simulated Lyman-α (Lyα) survey.

7.1.1 Lyα power

We begin this section by outlining our overall aims. To compute errors on BAO ($\alpha_{||}$ and $\alpha_{\perp}$) we need to construct a Fisher matrix for a specific survey, which in-turn requires a covariance matrix and its derivates with $\alpha_{||}$ and $\alpha_{\perp}$. We could in principle choose the data vectors of our covariance matrix to be Lyα fluctuations (δ_F) in pixels of some width $\Delta\lambda$ (e.g. 0.8 Å, the DESI spectrograph pixelisation), but this would have enormous dimensions due to the large number of pixels in our survey. Thus, we follow previous analyses (Seo and Eisenstein, 2003; McDonald and Eisenstein, 2007; McQuinn and White, 2011). and use summary statistics, in particular the power spectrum of fluctuations, as our data vector. This is computationally much less expensive and contains effectively the same amount of information.

Thus we begin here, by deriving an expression for our observed power spectrum $P_T(\mathbf{k})$, which will contain signal from Lyα fluctuations, as well as shot noise and spectral noise. It is convenient to begin in configuration space, defining our observed flux fluctuations, then computing the correlation of these and finally the power spectrum $P_T(\mathbf{k})$ through a Fourier transform. We start with a weighted flux transmission field:

$$\delta(\mathbf{x}) = \frac{w(\mathbf{x}_\perp)}{\overline{w}}\left(\delta_F(\mathbf{x}) + \delta_N(\mathbf{x})\right), \tag{7.4}$$

where $\delta_F(\mathbf{x}) = \delta_F(x_{||}, \mathbf{x}_\perp)$ is the flux-transmission field (equation 5.1), δ_N is spectral noise, and $\overline{w}$ is the mean weight. Our weights $w(\mathbf{x}_\perp)$ will take into account quasar brightness, which in-turn determines spectral noise, and we assume they are only dependent on transverse separation. This is akin to assuming that our weights are are completely correlated within forests - along the line-of-sight (LOS) - and completely uncorrelated between forests, and that spectral noise is the same at every point in the same forest. $w(\mathbf{x}_\perp)$ is fully defined over our survey volume, where any location not probed by a quasar LOS has zero weight.

Now, the covariance of the field in equation 7.4 is defined as:

$$\begin{aligned}\langle\delta(\mathbf{x})\delta(\mathbf{x}')\rangle &= \left\langle\frac{w(\mathbf{x}_\perp)w(\mathbf{x}'_\perp)}{\overline{w}^2}\left(\delta_F(\mathbf{x})\delta_F(\mathbf{x}') + \delta_F(\mathbf{x})\delta_N(\mathbf{x}') + \delta_F(\mathbf{x}')\delta_N(\mathbf{x}) + \delta_N(\mathbf{x})\delta_N(\mathbf{x}')\right)\right\rangle. \\ &= \left\langle\frac{w(\mathbf{x}_\perp)w(\mathbf{x}'_\perp)}{\overline{w}^2}\left(\delta_F(\mathbf{x})\delta_F(\mathbf{x}') + \delta_N(\mathbf{x})\delta_N(\mathbf{x}')\right)\right\rangle\end{aligned} \tag{7.5}$$

where we have removed the cross-terms, which vanish since the flux transmission field (δ_F) and noise field (δ_N) are independent. Taking first the part of equation

7.5 which includes δ_F, we have:

$$\begin{aligned}\left\langle \frac{w(\mathbf{x}_\perp)w(\mathbf{x}'_\perp)}{\overline{w}^2}\delta_F(\mathbf{x})\delta_F(\mathbf{x}')\right\rangle &= \langle(1+\delta_w(\mathbf{x}_\perp))(1+\delta_w(\mathbf{x}'_\perp))\delta_F(\mathbf{x})\delta_F(\mathbf{x}')\rangle \\ &= \xi_F(\mathbf{r}) + \langle\delta_w(\mathbf{x}_\perp)\delta_w(\mathbf{x}'_\perp)\rangle\,\xi_F(\mathbf{r})\,,\end{aligned} \tag{7.6}$$

where $\xi_F(\mathbf{r}) = \langle\delta_F(\mathbf{x})\delta_F(\mathbf{x}')\rangle$, and $\xi_F(\mathbf{r})$ is the Lyα autocorrelation function for points separated $\mathbf{r}$. We have also introduced the weight overdensity, defined as $\delta_w(x) = w(\mathbf{x}_\perp)/\overline{w} - 1$, such that $\langle\delta_w\rangle = 0$. Note that expectation values involving δ_w and δ_F separate, since they are uncorrelated.

Our weights are a white noise field (uncorrelated) in the transverse direction ($\mathbf{x}_\perp$) and constant (degenerate) along the LOS, which means $\left\langle\delta_w(\mathbf{x}_\perp)\delta_w(\mathbf{x}'_\perp)\right\rangle = P_w^\perp\delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp)$, where δ^D is the Dirac delta function and $P_w^\perp$ is a constant power spectrum in the 2D plane of the sky ($\mathbf{x}_\perp$). In general, a discrete Gaussian white noise field in 2-dimensions has power per mode $P = \sigma^2 s^2$, where s is the angle of one side of a pixel in the limit where $s \to 0$, and σ^2 is the pixel variance [2]. Then, we can write:

$$P_w^\perp = \sigma_w^2 s^2 = \left(\frac{\langle w^2\rangle}{\overline{w}^2} - 1\right) s^2. \tag{7.7}$$

.

and equation 7.6 becomes:

$$\left\langle \frac{w(\mathbf{x}_\perp)w(\mathbf{x}'_\perp)}{\overline{w}^2}\delta_F(\mathbf{x})\delta_F(\mathbf{x}')\right\rangle = \xi_F(\mathbf{r}) + P_w^\perp\delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp)\xi_F(\mathbf{r}). \tag{7.8}$$

The last term in this equation is known as the "aliasing" term, and it arises due to the fact that Lyα fluctuations only sample our 3D volume along the LOS. The second part of equation 7.5 is the weighted noise correlation, which we write as:

$$\xi_N^{\text{eff}}(\mathbf{x}_\perp, \mathbf{x}'_\perp) = \left\langle \frac{w(\mathbf{x}_\perp)w(\mathbf{x}'_\perp)}{\overline{w}^2}\delta_N(\mathbf{x}_\perp)\delta_N(\mathbf{x}'_\perp)\right\rangle\,, \tag{7.9}$$

where now weights are dependent on noise, so the term in brackets does not separate. We have also written the spectral noise purely in terms of transverse location, since for simplicity we are assuming it is the same at each point along the LOS [3]. Both our weights and noise should be uncorrelated in the transverse direction, leading to two scenarios:

---

[2] Since all k-modes in a white noise field are independent, the covariance is diagonal

[3] Note that as a noise field arising from instrumental effects, we are no longer assuming statistical homogeneity

$$\xi_N^{\text{eff}}(\mathbf{x}_\perp, \mathbf{x}'_\perp) = \begin{cases} 0 & , \mathbf{x}_\perp \neq \mathbf{x}_\perp \\ \left\langle \frac{w^2(\mathbf{x}_\perp)}{\overline{w}^2} \delta^2(\mathbf{x}_\perp) \right\rangle & , \ \mathbf{x}_\perp = \mathbf{x}_\perp \,. \end{cases} \tag{7.10}$$

Substituting the expressions from equations 7.8 and 7.9 into leaves us with:

$$\langle \delta(\mathbf{x})\delta(\mathbf{x}') \rangle = \xi_F(\mathbf{r}) + P_w^\perp \delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp)\xi_F(\mathbf{r}) + \xi_N^{\text{eff}}(\mathbf{x}_\perp, \mathbf{x}'_\perp)\,\delta^D(\mathbf{x} - \mathbf{x}')\,, \tag{7.11}$$

where have multiplied the weighted noise by the Dirac delta, to ensure the conditions of equation 7.10 are met. At this point, we can move into Fourier space by taking the transforming the equation:

$$\begin{aligned} \left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^*(\mathbf{k}') \right\rangle = \int d\mathbf{x}\, d\mathbf{x}'\, e^{-i\mathbf{k}\cdot\mathbf{x}+i\mathbf{k}'\cdot\mathbf{x}'} \big[\, & \xi_F(\mathbf{x} - \mathbf{x}') + P_w^\perp\, \delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp)\, \xi_F(\mathbf{x} - \mathbf{x}') \\ & + \xi_N^{\text{eff}}(\mathbf{x}_\perp, \mathbf{x}'_\perp)\, \delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp) \,\big]. \end{aligned} \tag{7.12}$$

Using the fact that $\delta(\mathbf{k}) = \int \delta(\mathbf{x}) e^{-i\mathbf{k}\mathbf{x}}\, d\mathbf{x}$ and $\left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^*(\mathbf{k}') \right\rangle = (2\pi)^3 \delta^D(\mathbf{k}-\mathbf{k}')P(\mathbf{k})$, we have:

$$\begin{aligned} \left\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^*(\mathbf{k}') \right\rangle = {} & (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}') P_F(\mathbf{k}) + \\ & P_w^\perp \int d\mathbf{x}_\perp d\mathbf{x}'_\perp e^{-i\mathbf{k}_\perp \mathbf{x}_\perp} e^{i\mathbf{k}'_\perp \mathbf{x}'_\perp} \int dx_\parallel dx'_\parallel e^{-ik_\parallel x_\parallel}\, e^{ik'_\parallel x'_\parallel}\, \xi_F(\mathbf{x} - \mathbf{x}')\delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp) \\ & + \int d\mathbf{x}_\perp d\mathbf{x}'_\perp e^{-i\mathbf{k}_\perp \mathbf{x}_\perp} e^{i\mathbf{k}'_\perp \mathbf{x}'_\perp} \int dx_\parallel dx'_\parallel e^{-ik_\parallel x_\parallel}\, e^{ik'_\parallel x'_\parallel}\, \xi_N^{\text{eff}}(\mathbf{x}_\perp, \mathbf{x}'_\perp)\delta^D(\mathbf{x}_\perp - \mathbf{x}'_\perp)\,, \end{aligned} \tag{7.13}$$

where $P_F(\mathbf{k})$ is the 3-dimensional Lyα forest power spectrum (P3D). Since we are interested in BAO (linear regime), we use a simple Kaiser model (Kaiser, 1987):

$$P_F(\mathbf{k}, z) = b_F^2(z)\,(1 + \beta_F(z)\mu^2(\mathbf{k}))^2 P_L(k, z). \tag{7.14}$$

The linear matter power spectrum we use is computed using Planck 2018 cosmological results (Planck Collaboration et al., 2020) for a flat ΛCDM model. In the aliasing (second) term, applying the Dirac delta collapses the integral to:

$$(2\pi)^2 \delta^D(\mathbf{k}_\perp - \mathbf{k}'_\perp)\, P_w^\perp \int dx_\parallel dx'_\parallel e^{-ik_\parallel x_\parallel}\, e^{ik'_\parallel x'_\parallel}\, \xi_F(x_\parallel - x'_\parallel)\,. \tag{7.15}$$

Evaluating the LOS integral, we arrive at:

$$(2\pi)^3 \delta^D(\mathbf{k}_\perp - \mathbf{k}'_\perp)\, \delta^D(k_\parallel - k'_\parallel)\, P_w^\perp\, P_F^{\text{1D}}(k_\parallel) \ = (2\pi)^3 \delta^D(\mathbf{k} - \mathbf{k}')\, P_w^\perp\, P_F^{\text{1D}}(k_\parallel)\,, \tag{7.16}$$

where $P_F^{\rm 1D}$ is the 1-dimensional Lyα forest power spectrum (P1D), for which we employ the empirical model of Palanque-Delabrouille et al., 2013. This fits the observed Lyα fluctuations down to relatively small scales ($\sim 2\times10^{-2}\,{\rm km^{-1}s}$), but it is also suitable for our scales of interest ($\sim 10^{-4}\,{\rm km^{-1}s}$):

$$P_F^{\rm 1D}(k_\parallel) = \frac{\pi A_F}{k_\parallel}\left(\frac{1+z}{1+z_0}\right)^{B_F}\left(\frac{k_\parallel}{k_0}\right)^{3+n_F+\alpha_F\ln\frac{k_\parallel}{k_0}+\beta_F\ln\frac{1+z}{1+z_0}}\,. \tag{7.17}$$

A_F is the amplitude of the P1D (in ${\rm kms^{-1}}$) at redshift z_0, and pivot mode k_0, equal to 3 and $0.009\,{\rm km^{-1}s}$ respectively. n_F and α_F are the slope and curvature parameters, and B_F and β_F govern the redshift evolution of the P1D. Note that this function is also corrected such that it becomes flat at low-$k_\parallel$.

Using the same methodology as before, the final integral in equation 7.13 becomes $(2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')\,P_N^{\rm eff}$, with:

$$P_N^{\rm eff} = s^2\Delta\lambda_p\frac{\langle w^2\sigma_N^2\rangle}{\overline{w}^2}\,, \tag{7.18}$$

where noise variance $\sigma_N^2(\mathbf{x}_\perp) = \langle\delta_N^2(\mathbf{x}_\perp)\rangle$, and the prefactor $s^2\Delta\lambda_p$ converts the dimensionless noise variance per pixel into 3D power units. We reiterate that s is only used to strictly define our equations, and that it will cancel out in our final expressions.

Although we have not written it explicitly here, we will show in section 7.1.2 that both $P_w^\perp$ and $P_N^{\rm eff}$ depend on $\mathbf{k}$ through their dependence on the weights (see equation 7.27). Our final expression for the measured Lyα power, P_T, is then:

$$\left\langle\tilde\delta(\mathbf{k})\tilde\delta^*(\mathbf{k}')\right\rangle = (2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')\left(P_F(\mathbf{k}) + P_w^\perp P_F^{\rm 1D}(k_\parallel) + P_N^{\rm eff}\right) \tag{7.19}$$

$$= (2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')P_T(\mathbf{k})\,. \tag{7.20}$$

7.1.2 Evaluating $P_N^{\rm eff}$ and $P_w^\perp$

To estimate our observed Lyα power (equation 7.19), the last thing we need to do is evaluate the aliasing term (specifically $P_w^\perp$) and noise term ($P_N^{\rm eff}$) in equation 7.19. Starting with the former (equation 7.7), our mean weights are defined in terms of quasar magnitude:

$$\overline{w} = \int_{-\infty}^{m_{\rm max}} dm\, p(m)\, w(m) = L_F\, s^2\int_{-\infty}^{m_{\rm max}} dm\,\frac{dn_q}{dm}\,w(m) \tag{7.21}$$

where m is apparent magnitude and $L_F = c\ln\frac{\lambda^{\rm rf,max}}{\lambda^{\rm rf,min}}$ is forest length in ${\rm kms^{-1}}$, defined in terms of the rest-frame ($\lambda^{\rm rf}$) forest region. dn_q/dm is the quasar density per magnitude, and the probability of a quasar (of magnitude m) forest probing a small volume of space is $p(m) = L_F\, s^2 \frac{dn_q}{dm}$. From the definition of $P_w^{\perp}$ (equation 7.7), we also need an expression for $\langle w^2 \rangle$:

$$\langle w^2 \rangle = L_F\, s^2 \int_{-\infty}^{m_{\max}} dm \frac{dn_q}{dm}\, w^2(m). \tag{7.22}$$

Combining the two equations above, with our definition of $P_w^{\perp}$, gives:

$$P_w^{\perp} = \frac{\int_{-\infty}^{m_{\max}} dm \frac{dn_q}{dm}\, w^2(m)}{L_F \left[\int_{-\infty}^{m_{\max}} dm \frac{dn_q}{dm}\, w(m) \right]^2}, \tag{7.23}$$

in the limit where $s \longrightarrow 0$. Going forward, we will call the term in brackets in the denominator I_1, and the term in the numerator I_2, such that:

$$P_w^{\perp} = \frac{I_2}{L_F I_1^2}. \tag{7.24}$$

These integrals isolate all parts of the covariance which contain all information related to the quasar (and therefore Lyα forest pixel) distribution.

Our expression for the effective noise power (P_N^{eff}; equation 7.18) can also be evaluated by expressing the weights and variance as functions of magnitude:

$$P_N^{\mathrm{eff}} = \frac{s^2 \Delta\lambda}{\overline{w}^2} \int_{-\infty}^{m_{\max}} dm\ \sigma_N^2(m)\, L_F\, s^2 \frac{dn_q}{dm}\, w^2(m)\,. \tag{7.25}$$

Introducing a new term $I_3 = \int dm\ \sigma_N^2(m) \frac{dn_q}{dm}\, w^2(m)$, and substituting for I_1 ($= \overline{w}/s^2 L_F$), we arrive at:

$$P_N^{\mathrm{eff}} = \frac{\Delta\lambda_p\, I_3}{L_F I_1^2}\,. \tag{7.26}$$

To evaluate equations 7.26 and 7.24, all that remains to define are the weights $w(m)$ and the spectral variance $\sigma^2(m)$. For the former, we assume the Feldman, Kaiser and Peacock (FKP) form:

$$w^{\mathrm{FKP}}(m) = \frac{P_S}{P_S + P_N(m)}, \tag{7.27}$$

where P_S is the signal power at $\mathbf{k_w} = (k_w, \mu_w)$. Our weights are not very sensitive to the mode ($\mathbf{k}_w$) we choose to evaluate the signal power at, so we simply and choose the central mode of our covariance, $k = 0.07\, h\mathrm{Mpc}^{-1}$, $\mu = 0.5$. The aliasing term is not formally noise, since it comes from projected LSS, and is correlated with real signal. Thus, we do not include it in the noise term, as we find that down-weighting lines-of-sight with larger aliasing increases our variance.

The noise power at magnitude m, denoted $P_N(m(\mathbf{x}))$, combines contributions from Poisson and spectral noise. We write it as:

$$P_N(m) = \frac{\sigma_N^2(m)}{n_{\rm 3D}^{\rm eff}}, \tag{7.28}$$

where

$$n_{\rm 3D}^{\rm eff} = \frac{\overline{w}}{s^2 \Delta\lambda} \tag{7.29}$$

is the effective 3D Lyα pixel density. This is also equivalent to (removing dependence on s) $L_F I_1/\Delta\lambda$. Note that $P_N(m)$ is an approximation of the effective noise $P_N^{\rm eff}$. For galaxies (either LBGs or quasars), we can write the noise power in an analogous way:

$$P_N(m) = \frac{\sigma_g^2(m)}{\overline{n}_{\rm g}} = \frac{\overline{n}_g}{n_g(m)} \frac{1}{\overline{n}_g} = \frac{1}{n_g(m)}, \tag{7.30}$$

where $\overline{n}_{\rm gal}$ is the mean galaxy density at magnitude m (our magnitude dependence is equivalent to position dependence). In this case σ_g^2 is just Poisson (shot) noise.

There is mutual dependence between the weights, and the factors I_1 and I_2, therefore we must compute them iteratively. We begin by assuming an initial form for the weights based on the 1D power:

$$w_{\rm init}(m) = \frac{P^{\rm 1D}(k_\parallel)}{P^{\rm 1D}(k_\parallel) + \sigma_N^2(m)}. \tag{7.31}$$

It is important to note that the final converged weights are not very sensitive to the choice of initial weights. Once we have $w_{\rm init}$, we compute I_1, I_2 and $P_N(m)$, then $w(m)$ as defined in equation 7.27. We repeat this latter process around 3 times, until the values $w(m)$ have converged.

In figure 7.1 we show the relative contribution to the total power of each term, for DESI-II LBG (left) and quasar (right) forests (see section 7.2 for a description of the samples). We plot each term as a function of k, for $\mu = 0$ (solid lines) and $\mu = 1$ (dashed lines). First of all, we see that the effective noise (purple) is a much larger contribution to P_T for LBG forests - a consequence of the fact that the SNR/Å for LBGs is relatively low. For quasars on the other hand, the aliasing term - which depends strongly on target density - is the dominant contribution. For $\mu = 1$, we see the aliasing term start to fall off towards high k ($k_\parallel$), as the P1D signal power decreases. We can gain an intuition from this plot, without running a full forecast, that the best way to increase BAO precision from LBG forests is to increase exposure time (and SNR).

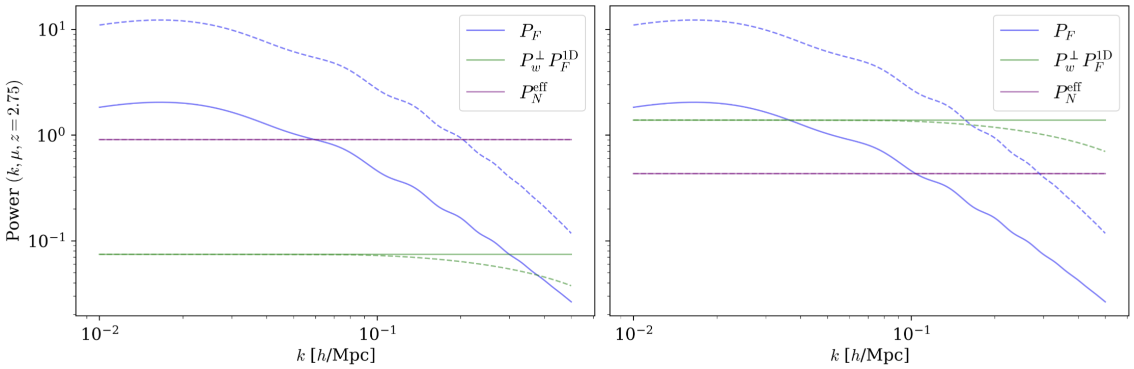

Figure 7.1: The three components contributing to the total observed Lyα power spectrum (P_T): the 3D power (P_{3D}, blue), aliasing (green), and effective noise (purple), shown for LBG forests (left) and quasar forests (right). Each term is plotted as a function of k, for $\mu = 0$ (solid lines) and $\mu = 1$ (dashed lines). The relatively lower SNR per Å in LBG forests leads to a larger effective noise contribution, while for quasars, the aliasing term dominates.

Now that we have a clear way to compute the observed Lyα power, we will move on to estimating our covariance matrix, and getting projections on BAO parameters from the resulting Fisher matrix.

7.1.3 Covariance estimation

The general covariance of our measured power spectra is equal to:

$$C_{ijnm} = \langle \Delta P_{ij} \Delta P_{nm} \rangle = \frac{2\pi^2}{V_s\, k^2\, dk\, d\mu} (P_{im} P_{jn} + P_{in} P_{jm}) \,, \tag{7.32}$$

where $\Delta P_{ij} = \hat{P}_{ij} - P_{ij}$ is the band power error, and i, j, n, m are tracers of structure in the same survey volume V_s. The prefactor $2\pi^2 / V_s k^2\, dk\, d\mu$ is equivalent to $1/N_k$, where N_k is the number of independent k modes. For the Lyα forest covariance ($i = j = n = m$), and Gaussian fluctuations δ_k, equation 7.32 reduces to:

$$C_F(\mathbf{k}) = \frac{2 P_T^2(\mathbf{k})}{N_k} \,. \tag{7.33}$$

The observed power for point tracers has no aliasing term ($P_w^\perp P_F^{1D}$), or dependence on spectral noise, and can be approximated by (Tegmark, 1997):

$$C_{\text{g}}(\mathbf{k}) \approx 2 \frac{P_{\text{g}}^2(\mathbf{k})}{V_k V_{\text{eff}}(\mathbf{k})} \,, \tag{7.34}$$

where $V_k = N_k \,/\, V_s$ is the volume element of a k-mode, P_{g} is the galaxy/quasar power spectrum and

$$V_{\mathrm{eff,g}}(\mathbf{k}) = V_s \frac{P_{\mathrm{g}}(\mathbf{k})}{P_{\mathrm{g}}(\mathbf{k}) + \overline{n}_{\mathrm{g}}^{-1}}. \tag{7.35}$$

$\overline{n}_{\mathrm{g}}$ is the mean galaxy/quasar number density. $V_{\mathrm{eff,g}}$ is the effective volume used to measure power, where modes at which $\overline{n}_{\mathrm{g}}P(\mathbf{k}) < 1$ contribute only a small amount. Note that for a high-density sample of galaxies where $\overline{n}P \gg 1$, the covariance becomes solely dependent on the survey V_s.

The Lyα forest covariance can also be written in terms of effective volume, as is done in McQuinn and White, 2011. This is defined as:

$$V_{\mathrm{eff,F}}(\mathbf{k}) = V_s \frac{P_{\mathrm{F}}(\mathbf{k})}{P_{\mathrm{F}}(\mathbf{k}) + P_F^{\mathrm{1D}}/n_{\mathrm{2D}}^{\mathrm{eff}}}\,, \tag{7.36}$$

where they introduce the effective pixel density in 2D (in the plane of the sky):

$$\frac{1}{n_{\mathrm{2D}}^{\mathrm{eff}}} = P_w^{\perp} + \frac{P_N^{\mathrm{eff}}}{P_F^{\mathrm{1D}}}\,. \tag{7.37}$$

The cross-power between point tracers and the Lyα forest, which we also wish to forecast, has variance given by (McQuinn and White, 2011):

$$\sigma^2_{\mathrm{g\,x\,Ly}\alpha} = P^2_{\mathrm{g\,x\,Ly}\alpha}(\mathbf{k}) + P_T(\mathbf{k})(P_{\mathrm{g}}(\mathbf{k}) + \overline{n}_{\mathrm{g}}^{-1})\,, \tag{7.38}$$

where $P_{\mathrm{g\,x\,Ly}\alpha}(\mathbf{k}) = \langle \tilde{\delta}_F(\mathbf{k})\tilde{\delta}_{\mathrm{g}}(\mathbf{k})\rangle$ is the galaxy/quasar-Lyα cross-power spectrum. The cross-power is a complimentary measurement to the Lyα auto-power, as it can break degeneracies between cosmological parameters (see section 6.5.1), and in some cases be more constraining. Specifically, $P_{\mathrm{g\,x\,Ly}\alpha}$ is more sensitive to $P_L(k)$ when (McQuinn and White, 2011):

$$\overline{n}_{\mathrm{g}}\, b_{\mathrm{g}}^2 \geq \frac{n_{\mathrm{2D}}^{\mathrm{eff}}\, b^2_{\mathrm{Ly}\alpha}}{P^{\mathrm{1D}}(k_{\parallel})}\,. \tag{7.39}$$

The spectral noise in the Lyα forests of LBGs is smaller compared with quasars. We have already shown the affect of this on our total power contributions (figure 7.1), and in section 7.2 where we discuss our DESI-II forecast, we will show how this affects the relative BAO errors from the auto- and cross-power spectra.

7.1.4 Estimating BAO

Now that we have all of the necessary components to evaluate the Fisher matrix (equation 7.2), we can turn towards calculating projected BAO errors. For a large enough number of independent modes ($V_n V_{\rm eff} \gg 1$), we can simplify expressions for the Fisher matrix for both the Lyα forest and point tracers:

$$F_{ij}^{{\rm Ly}\alpha} = \frac{V_s}{4\pi^2} \int_{\mathbf{k}_{\rm min}}^{\mathbf{k}_{\rm max}} \int_0^1 \frac{k^2\, dk\, d\mu}{P_T^2(\mathbf{k})} \frac{\partial P_F(\mathbf{k})}{\partial \theta_i} \frac{\partial P_F(\mathbf{k})}{\partial \theta_j}\,, \tag{7.40}$$

and

$$F_{ij}^{\rm g} = \frac{1}{4\pi^2} \int_{\mathbf{k}_{\rm min}}^{\mathbf{k}_{\rm max}} \int_0^1 \frac{V_{\rm eff}(\mathbf{k})\, k^2\, dk\, d\mu}{P_{\rm g}^2(\mathbf{k})} \frac{\partial P_{\rm LBG}(\mathbf{k})}{\partial \theta_i} \frac{\partial P_{\rm LBG}(\mathbf{k})}{\partial \theta_j}\,. \tag{7.41}$$

Note that since the density field is real-valued, $-\mathbf{k}$ and $+\mathbf{k}$ are degenerate, so we integrate across $0 < \mu < 1$ rather than $-1 < \mu < 1$.

Next, we want to estimate the Fisher matrix for parameters $i, j = \alpha_{||}, \alpha_{\perp}$. The power spectra derivates in the above equations can be computed as follows:

$$\begin{aligned}
\frac{\partial P}{\partial \alpha_{||}} &= \frac{\partial P}{\partial \log k'} \frac{\partial \log k'}{\partial \alpha_{||}} && (7.42)\\
&= \frac{1}{k'} \frac{\partial P}{\partial \log k'} \frac{\partial k'}{\partial \alpha_{||}} && (7.43)\\
&= \frac{1}{k'} \frac{\partial P}{\partial \log k'} \frac{\partial \left(\sqrt{\alpha_{||}^2 \mu^2 + \alpha_{\perp}^2 (1-\mu^2)}\right)}{\partial \alpha_{||}} && (7.44)\\
&= \frac{\partial P}{\partial \log k'} \mu^2\,, && (7.45)
\end{aligned}$$

where $k'_{||} = \alpha_{||}\, k_{||}$ is the observed LOS wave number, shifted from the fiducial cosmology by $\alpha_{||}$. Note that to arrive at the final step in this equation, we set $\alpha_{||} = \alpha_{\perp} = 1$, which is fine since we are running a forecast. Following the same steps as equation 7.42 for the transverse BAO parameter, we get:

$$\frac{\partial P}{\partial \alpha_{\perp}} = \frac{\partial P}{\partial \log k'} (1-\mu^2)\,. \tag{7.46}$$

Before estimating the variances of $\alpha_{||}$ and $\alpha_{\perp}$ we must separate the peak and broadband components of the linear matter power spectrum, to avoid including the broadband signal in the summation of equations 7.40 and 7.41. We do this by fitting a high-degree polynomial, such that:

$$P_L(\mathbf{k}) = P_L^{\rm sm}(\mathbf{k}) + P_L^{\rm peak}(\mathbf{k}) \exp\left(-\frac{k_{\perp}^2 \Sigma_{\perp}^2}{2} - \frac{k_{||}^2 \Sigma_{||}^2}{2}\right)\,, \tag{7.47}$$

where $P_L^{\rm sm}(\mathbf{k})$ is the broadband component of P_L, and $P_L^{\rm peak}(\mathbf{k})$ is the linear BAO component, which appear as a series of wiggles. The factor multiplying the peak component accounts for non-linear smoothing of the BAO feature, where $\Sigma_{\perp}^2, \Sigma_{\parallel}^2 = 3.26, 6.42\,h^{-1}$Mpc (Eisenstein et al., 2007).

In the computation of F_{ij}, we sum over discrete bins dk and $d\mu$, of width $0.001\,h\,\mathrm{Mpc}^{-1}$ and 0.1 respectively. We choose $k_{\mathrm{min}} = 0.01\,h\,\mathrm{Mpc}^{-1}$ in the summation of F_{ij}, a much smaller value than previous Lyα forecasts. In McQuinn and White, 2011 for example, they choose $k_{\mathrm{min}} = 0.1\,h\,\mathrm{Mpc}^{-1}$, motivated by the fact that below this value, P_F is distorted by the continuum fitting procedure in data (see section 6.2.2)[4]. However, we assume that the continuum distortion contribution to the 1D power can be removed, using techniques developed for BAO analyses of 3D Lyα correlations (Bautista et al., 2017; du Mas des Bourboux et al., 2017).

At higher k, we start to become affected by non-linear processes, and as such choose a limiting scale of $k_{\mathrm{max}} = 0.5\,h\,\mathrm{Mpc}^{-1}$ following DESI Collaboration et al., 2024b. Additionally, the exponential correction in equation 7.47 should provide more realistic results. In contrary, for measurements that include the broadband signal (e.g. $f\sigma_8$), scales above $\sim 0.1\,h\,\mathrm{Mpc}^{-1}$ been shown to give unrealistic forecasts (Foroozan et al., 2021).

7.2 DESI-II forecast results

In this section we will forecast the high-redshift DESI-II survey, using LBGs and quasars as tracers and sources of Lyα forest fluctuations, and LAEs as tracers only. A version of this forecast was also included in the results section of Herrera-Alcantar et al., 2025a, which gave measurements of Lyα clustering in LBGs and LAEs in DESI, and prospects for DESI-II.

Starting from 2029, DESI-II will observe 5000 deg^2 of sky, to a greater depth than its predecessor. LBGs and LAEs will be observable up to a limiting magnitude of ~ 24.5 (Payerne et al., 2025). We will also benefit from extensive imaging from the LSST survey (Ivezić et al., 2019), which overlaps with the entire DESI-II footprint. These depths, along with many other instrumental upgrades, should be able to provide $\sim$ 1000 LBG and $\sim$ 1000 LAE targets per deg^2. Furthermore, DESI-II will provide a complete quasar catalogue up to $r \sim 23.5$, leading to an improved Lyα quasar density (with respect to DESI) of 100 deg^{-2} at $z > 2.15$. This choice of redshift limit comes from DESI science requirements for Lyα quasars (DESI Collaboration et al., 2016a) - in reality the Lyα forest is observable in quasars with $z \gtrsim 2$.

---

[4] McDonald et al., 2005 found that at this limit, the power contributed by continuum distortions (in the LOS) is < 1% of P_F.

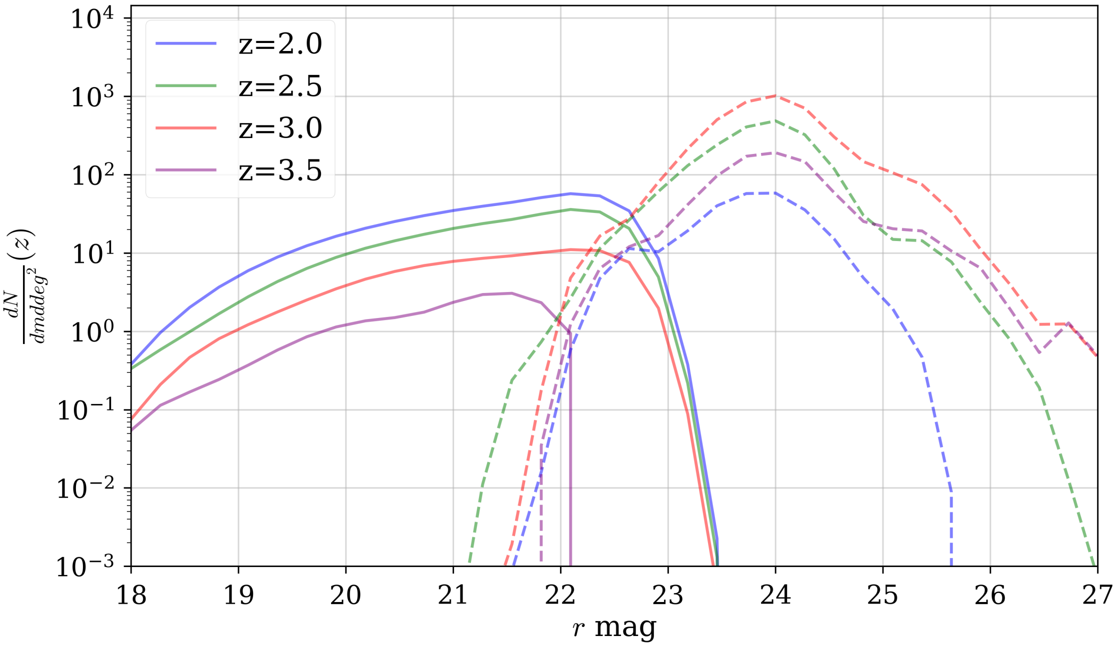

Figure 7.2: Luminosity function (number per $\deg^{-2}$) of quasar (solid) and LBG (dashed) targets in DESI-II as a function of r band magnitude, for 4 different redshifts. For LAEs we assume the same distribution function of LBGs.

To run our Fisher forecast (detailed in section 7.1), we require these expected total target densities, magnitude limits, and distributions with magnitude and redshift for LBGs, LAEs and quasars (we work in $\deg^{-2}(\mathrm{km/s})^{-1}$). Note that the distributions (dn_q/dm; see e.g. equation 7.21) are dependent on redshift, although we did not explicitly write it. In figure 7.2, we show the input distributions for quasars (solid lines) and LBGs (dashed lines). Because we have limited LAE data, we approximate their distribution as following that of our LBGs.

The quasar distributions are taken from preliminary DESI target selection (Yèche et al., 2020), which give a total integrated density of $n_q \sim 204\,\deg^{-2}$, or $n_q \sim 50\,\deg^{-2}$ at $z > 2.15$. To achieve the desired DESI-II Lyα quasar density ($100\,\deg^{-2}$ at $z > 2.15$), we re-normalise the distributions of figure 7.2, multiplying dn_q/dm by $n_q^{\mathrm{DESI-II}}(z > 2.15)/n_q^{\mathrm{DESI}}(z > 2.15) \approx 2$. Note that this simple re-scaling is not entirely realistic since most of the newly selected targets would come from increasing the survey limiting magnitude, where forests generally have lower SNR. Ideally, one would use an accurate quasar luminosity function (LF) which extends to magnitude $r \sim 23.5$, to avoid underestimating our parameter variance.

The LBG (which we use also for LAEs) distribution in figure 7.2 is taken from DESI measurements in the COSMOS field (Ruhlmann-Kleider et al., 2024; Herrera-Alcantar et al., 2025a; Payerne et al., 2025), and has a total selected target density of $n_{\mathrm{LBG}} = 453\,\deg^{-2}$ across all redshifts (practically all of these targets

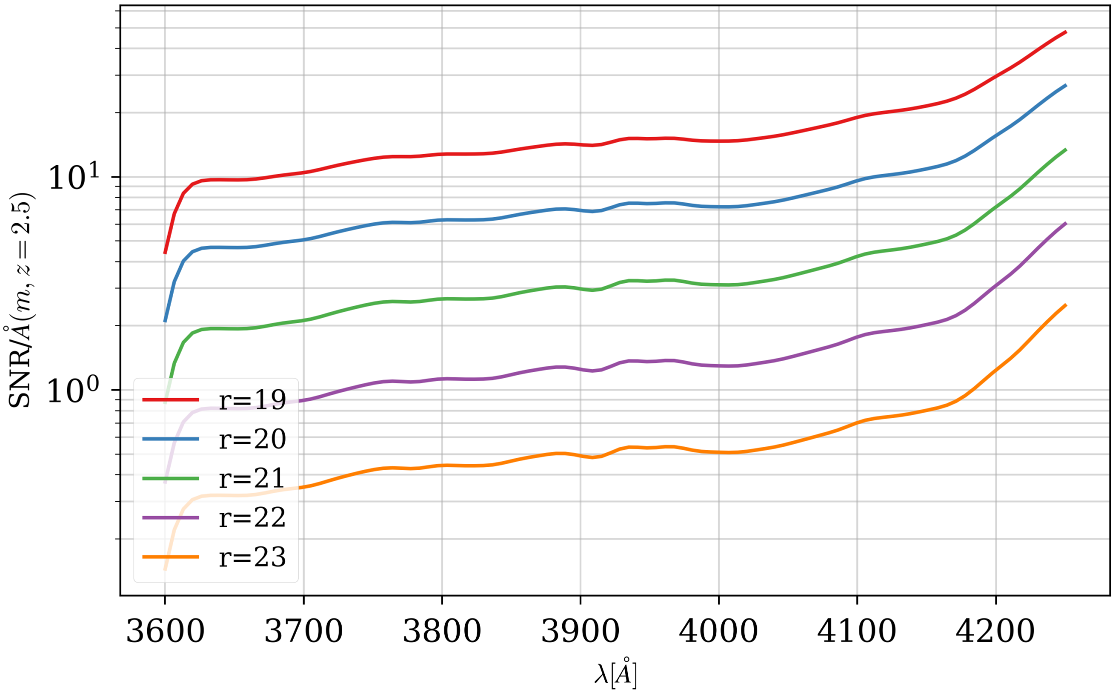

Figure 7.3: SNR per Angstrom in the Lyα forest of DESI quasars, as a function of r band magnitude, for $z_q = 2.5$. Note that the SNRs peak around 4200Å due to the Lyα emission line.

are at $z > 2.15$). To re-scale this to the expected DESI-II density (1000 deg^{-2}), we multiply $dn_{\rm LBG}/dm$ by $n_{\rm LBG}^{\rm DESI-II}/n_{\rm LBG}^{\rm DESI} \approx 2$.

To estimate the weights and covariance (see equation 7.25) of our Lyα forest survey, we will also need a function $\sigma(m, z, \lambda)$ which gives pixel variance as a function of tracer (quasar or LBG) magnitude and redshift, and observed wavelength (λ). For Lyα quasars, we obtain this by simulating observations with 4000s of effective exposure time on the DESI spectrograph, at a series of magnitudes and redshifts across the full observed wavelength grid of the instrument (3600-9800 Å). The result is shown in figure 7.3 for a quasar at $z \sim 2.5$ and selection of magnitudes. The wavelength range shown is that over which we observe almost all of our Lyα absorption. When computing our survey weights and covariance, we simply interpolate onto this function.

For LBGs, we do not have access to such simulations, so we approximate that each forest in our analysis has the same SNR per Angstrom. We use COSMOS field measurements (Herrera-Alcantar et al., 2025a) with a minimum of 2 hours exposure time, shown in figure 7.4, and get a mean SNR of 0.35 Å^{-1}. This will primarily impact the effective noise $P_N^{\rm eff}$ (equation 7.25), which depends on pixel noise. However, since the distribution of LBGs is strongly peaked around $r \sim 24$, this should not bias us significantly.

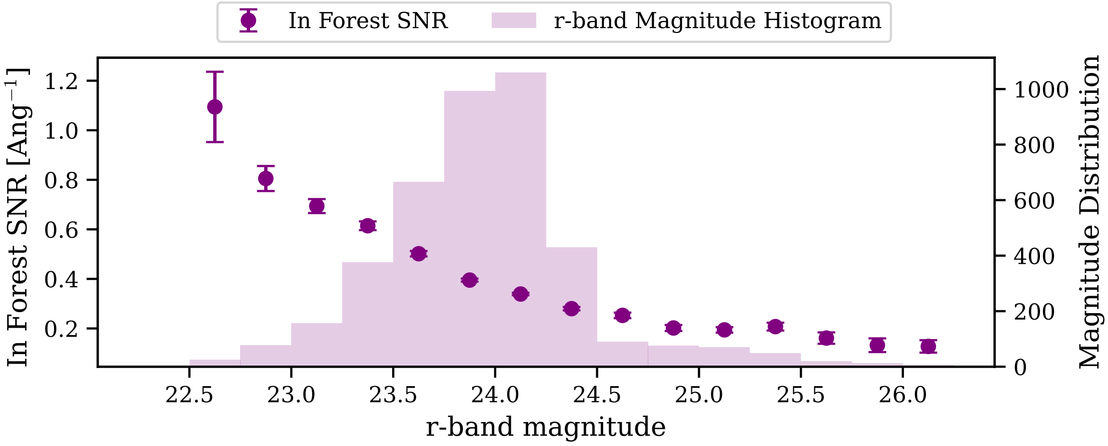

Figure 7.4: Measurements of SNR per Angstrom in the Lyα forest vs r band magnitude of Lyman break galaxies (LBGs), from DESI (Herrera-Alcantar et al., 2025a). These were performed in the COSMOS field, and contain a total of $\sim$ 5000 forests. These SNRs are obtained from observations with a minimum of 2 hours exposure time.

7.2.1 Computing the Fisher matrix

Now that we have all of the necessary inputs, we compute the weights, noise terms, covariance and Fisher matrices for each tracer in our forecast, for a given survey set-up, using the package `lyaforecast`[5] and following the methodology of section 7.1.

Firstly, we run our forecast for a redshift range between $z \in [2, 3.5]$, in 15 individual bins of width $dz = 0.1$. These are chosen to sample the quasar/galaxy distributions finely enough, such that our BAO errors converge. It also allows us to include the evolution of the relevant power spectra. Then, for each bin dz, we define the observed wavelength limits of the bin for Lyα absorption:

$$\lambda_{\text{min}} = \lambda_\alpha(1 + z_{\text{min}})\,, \tag{7.48}$$

$$\lambda_{\text{max}} = \lambda_\alpha(1 + z_{\text{max}}). \tag{7.49}$$

From these, we define the mean observed wavelength of the bin $\lambda_c = \sqrt{\lambda_{\text{min}}\lambda_{\text{max}}}$, and the redshift of the quasar/LBG that centres the Lyα forest in our bin:

$$z_o = \frac{\lambda_c}{\lambda_c^{rf}} - 1\,, \tag{7.50}$$

where $\lambda_c^{rf} = \sqrt{\lambda_{\text{min}}^{rf}\lambda_{\text{max}}^{rf}} \approx 1119\,\text{Å}$ is the mean of the Lyα forest rest-frame region. Note that we are using only the primary Lyα forest region between 1040 Åand

---

[5] https://github.com/igmhub/lyaforecast

Parameter	Lyα	Quasar	LBG	LAE
b	-0.16	3.8	3.0	1.9
β	1.45	0.26	0.33	0.52

Table 7.1: Bias parameter values for each tracer in our analysis, evaluated at $z = 2.5$.

1205 Å, and excluding the Lyβ region (see e.g. DESI Collaboration et al., 2025b) for simplicity. Note that internally, `lyaforecast` works in kms^{-1}, even though often we refer to e.g. k values in h Mpc^{-1}. The pixel width ($\Delta\lambda$) corresponds to the spacing of the DESI spectrograph, 0.8Å . Again, we use kms^{-1} for pixel widths internally, which average ~ 70 kms^{-1} (changes with redshift).

For each redshift bin dz, we compute the weights and noise terms P_N^{eff} and $P_w^{\perp}$ (at $k = 0.07\,h$Mpc^{-1}, $\mu = 0.5$), by integrating over magnitude, with noise and tracer distributions $\sigma_N(\lambda_c\,, z_o, m)$ and $dn/dm(z_o)$, given the forest length and pixelisation above. In practice, we want noise per pixel, which is equal to $\sqrt{\Delta\lambda}\,\sigma_N(\lambda_c\,, z_o, m)$.

With the noise terms computed, we can get the required covariance matrices, in discreet bins of k an μ. First, we compute the required 3D power spectra: `Ly`α`(LBG)`$^{\mathrm{auto}}$, `Ly`α`(LBG)`$^{\mathrm{cross}}$, `Ly`α`(LBG)xLAE`, `Ly`α`(QSO)xLBG`, `Ly`α`(QSO)LAE`, `LBG`$^{\mathrm{auto}}$ and `LAE`$^{\mathrm{auto}}$. For these we provide a set of bias/RSD parameters taken from DESI DR2 (DESI Collaboration et al., 2025b) and Ruhlmann-Kleider et al., 2024, summarised in table 7.1. We also evolve each bias according to equation 5.15.

The linear matter power spectrum is computed in each redshift bin using `camb`, with Planck 18 cosmology Planck Collaboration et al., 2020. The P1D that features in the Lyα forest covariance is estimated using the empirical model of Palanque-Delabrouille et al., 2013 (equation 7.17). Finally, we perform a sum over k and μ to estimate F_{ij} for each tracer, and derive errors on $\alpha_{\parallel,i}$ and $\alpha_{\perp,i}$ in each redshift bin i, using the method outlined in section 7.1.4. We combine our measurements across redshift as:

$$\sigma_{\alpha_x} = \frac{1}{\sqrt{\sum_i^{15} \frac{1}{\sigma^2_{\alpha_x,i}}}}, \tag{7.51}$$

where $x = \perp, \parallel$.

In the left panel of figure 7.5 we compare the SNR per k mode of `Ly`α`(LBG)`$^{\mathrm{auto}}$ (blue dashed), to the SNR per mode of the quasar Lyα auto-power spectrum (red solid). We see a turn-over point where measuring the Lyα forest auto-power from LBGs becomes more constraining, at $z \sim 2.5$. Noticeably, this lags slightly behind the same turn-over in the tracer SNR (right panel), due to the lower SNR per pixel of LBG forests. The increased SNR for quasar Lyα forests in DESI-II will improve

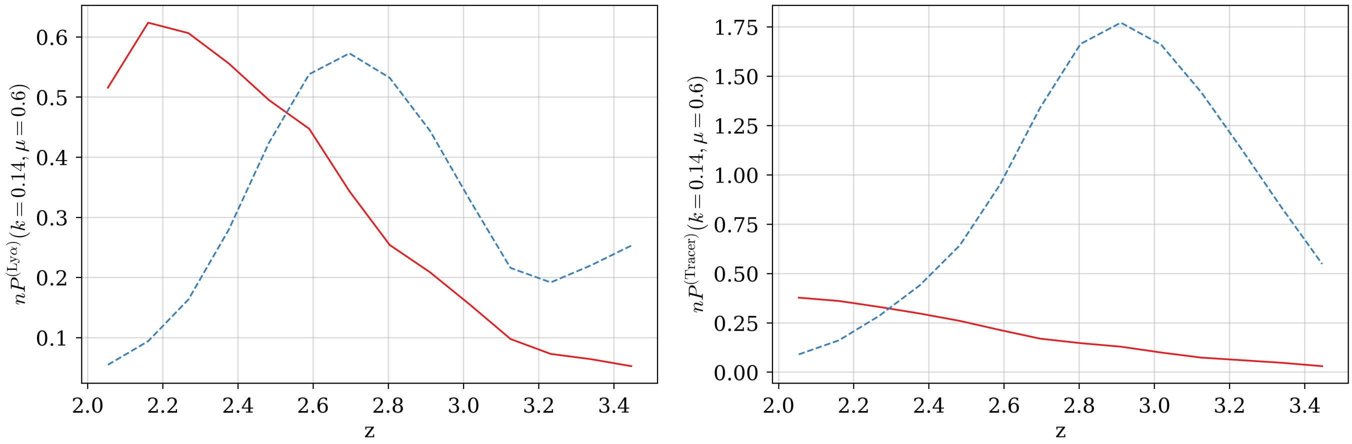

Figure 7.5: (Left) SNR per mode at $(k, \mu) = (0.14\,h\mathrm{Mpc}^{-1}, 0.6)$ in the Lyα forest power spectrum of quasars (red solid) and LBGs (blue dashed). (Right) SNR per mode of the quasar (red solid) and LBG (blue dashed) power spectrum. The large difference on the right is due to the high density of LBG targets in DESI-II (1000 deg^{-2}). In contrast, the lower pixel SNR in LBG forests makes the lines in the left plot more similar.

BAO constraints at $z_{\mathrm{eff}} \sim 2.3$, in combination with the rest of the forests in DESI not included in the DESI-II footprint. Here, we choose to forecast the cross-power `Lyα(QSO)xLBG`, since it is novel, and will benefit from the high target density of LBGs. Looking at the left panel of figure 7.5, the SNR per mode of the quasar power spectrum (red solid) and LBG power spectrum (blue dashed) differ a lot more than the Lyα forests. Here, the increased LBG target density plays a much bigger role, where the spectral noise is not a factor.

Our BAO (percent-)error projections are shown as a function of redshift for each of the forecasted power spectra, in figure 7.6. For each line, the constraints are estimated in 14 redshift bins, of width 0.1, from $z = 2 - 3.5$. The top left panel shows the autocorrelation of Lyα forests from LBGs (blue) and quasars (green). At $z \sim 2.5$ there is a turnover point where LBGs, which are observed at higher redshifts, become more constraining. We also add the cosmic variance floor for both DESI-II (black) and DESI (dark blue). This is the best possible Lyα autocorrelation measurement achievable from each survey, where both the shot noise and spectral noise are completely negligible. DESI has a lower floor only because of its greater survey area.

In the top right panel we compare the LBG and LAE autocorrelations, where the only difference between these tracers is the bias/RSD values we assign them (table 7.1). Note that here we do not include the quasar autocorrelation; because of the relatively low target density, it is not a competitive measurement. The only reason that the two sets of lines in the top left panel are comparable, is that the

SNR of pixels in quasars relatively high.

In the bottom left panel, we include the cross-power measurements: `Ly`α`(LBG)`$^{\mathrm{cross}}$, `Ly`α`(LBG)xLAE`, `Ly`α`(QSO)xLBG`, `Ly`α`(QSO)LAE`. We see (as in figure 7.5) that `Ly`α`(LBG)`$^{\mathrm{cross}}$ becomes more constraining than `Ly`α`(QSO)xLBG` at $z \sim 2.5$, and its LAE counterpart slightly after. We note that the difference is small, given that there are 10× more LBG forests, again reflecting the impact of the increased spectral noise.

The combined `Ly`α`(LBG)`$^{\mathrm{auto}}$ and `Ly`α`(LBG)`$^{\mathrm{cross}}$ measurements (bottom right) seem to provide similar constraints to the LBG auto (top right) across the line-of-sight, outperforming it along the line-of-sight. To combine `Ly`α`(LBG)`$^{\mathrm{auto}}$ and `Ly`α`(LBG)`$^{\mathrm{cross}}$, we performed a simple inverse sum (Font-Ribera et al., 2014a), equivalent to assuming no correlation between the two measurements. This is a high-noise ($nP \ll 1$) approximation, which is not necessarily valid in our regime, where $nP \sim 1$. For this reason we choose not to show other combinations of measurements. A better approach would be to compute the joint Fisher matrix, where the covariance (calculated with equation 7.32) includes all of the relevant cross-terms.

We would of course eventually like to combine all of our correlations together, which would improve our BAO precision and allow us to make better full-shape measurements by breaking parameter degeneracies. However, as shown in figure 7.6, there is a cosmic variance floor which limits measurements with even negligible noise. Thus, there is little benefit of combining tracers for BAO in this regime. This can also be understood looking at effective volume (equations 7.35 and 7.36): as $nP \gg 1$, only increasing survey volume can reduce the covariance of our measurement[6]. Note that even in the cosmic variance limited regime, combining tracers is still beneficial for RSD and growth measurements (McDonald and Seljak, 2009).

In table 7.2, we summarise the forecast results, with the combined (across redshift) precision on BAO for each of our tracers, and the effective redshifts z_{eff}. We confirm that the combination of `Ly`α`(LBG)`$^{\mathrm{auto}}$ and `Ly`α`(LBG)`$^{\mathrm{cross}}$ provide the best measurement of $\alpha_{\parallel}$ (0.70%) and $\alpha_{\perp}$ (0.57%), although the latter is only marginally better than the LBG auto. For comparison, the DESI 5-year survey forecasts (DESI Collaboration et al., 2024b) project high-redshift ($z \sim 2-3.5$) errors of 0.91% and 0.88% on $\alpha_{\perp}$ and $\alpha_{\parallel}$ respectively[7]. Thus, with **only** the combined Lyα measurement, we can expect $\sigma_{\alpha_{\perp}}^{\mathrm{DESI-II}} \approx 0.63\, \sigma_{\alpha_{\perp}}^{\mathrm{DESI}}$ and $\sigma_{\alpha_{\parallel}}^{\mathrm{DESI-II}} \approx 0.80\, \sigma_{\alpha_{\parallel}}^{\mathrm{DESI}}$.

We note that these values will be underestimated, due to the fact that we use

---

[6] This is often referred to as a volume-limited survey.

[7] From the joint Lyα(QSO) autocorrelation and Lyα(QSO)-QSO cross-correlation

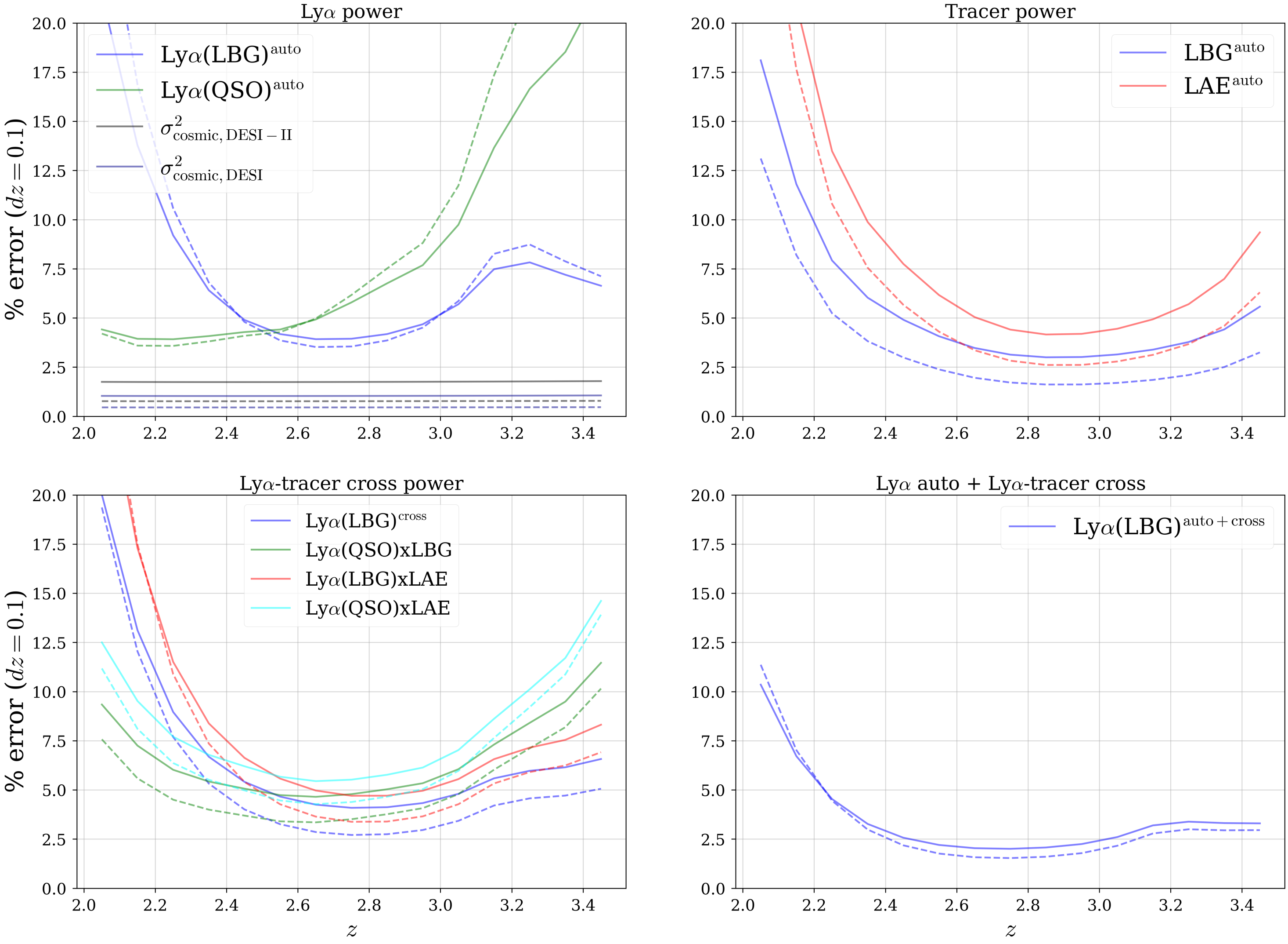

Figure 7.6: (Top left) projected percent error on line-of-sight ($\alpha_{\parallel}$; solid line) and transverse ($\alpha_{\perp}$; dashed line) BAO parameters from the `Ly`α`(LBG)`$^{\rm auto}$ power spectrum, as a function of redshift. Also included here is the cosmic variance limit, or measurement in absence of shot/spectral noise, for DESI and DESI-II. DESI has a lower limit due only to its larger survey area. (Top right) the same projects for the LBG power spectrum. (Bottom left) `Ly`α`(QSO)xLBG` (red) and `Ly`α`(LBG)`$^{\rm cross}$ (blue). (Bottom right) the combined `Ly`α`(LBG)`$^{\rm cross}$ and `LBG`$^{\rm auto}$ projections.

Measurement	$100\times\sigma_{\alpha_\perp}$	$100\times\sigma_{\alpha_\parallel}$	z_{eff}
`Ly`α`(LBG)`$^{\mathrm{auto}}$	1.36	1.41	2.90
`Ly`α`(LBG)`$^{\mathrm{cross}}$	0.99	1.39	2.79
`Ly`α`(LBG)`$^{\mathrm{auto+cross}}$	0.57	0.70	2.82
`Ly`α`(LBG)xLAE`	1.26	1.65	2.80
`Ly`α`(QSO)xLBG`	1.17	1.55	2.61
`Ly`α`(QSO)xLAE`	1.52	1.85	2.64
`LBG`$^{\mathrm{auto}}$	0.59	1.04	2.84
`LAE`$^{\mathrm{auto}}$	1.00	1.5	2.85

Table 7.2: Percent errors on BAO parameters $\alpha_\perp$ and $\alpha_\parallel$ for several measurements in a DESI-II survey forecast.

the high-noise approximation to combine errors. To get a more accurate sense of the improved precision on the multi-tracer BAO measurement, we would have to do a joint Fisher matrix analysis, as mentioned above. One should also note that the effective redshift of the high-redshift measurements in DESI is $z_{\mathrm{eff}} \sim 2.33$ (DESI Collaboration et al., 2025b), making the high effective redshifts of DESI-II very complimentary in a cosmological analysis. Finally, despite being less constraining, the `Ly`α`(QSO)xLBG` measurement has an intermediate effective redshift between DESI and DESI-II, which again is useful for measuring cosmic expansion. The LAE correlations are also competitive, and help to break degeneracies in $f\sigma_8$ by providing independent measurements of RSD.

In future work we would take several steps to improve and expand our forecast. Firstly, we would look to incorporating proper simulations of spectral noise in LBG spectra $\sigma_N(m, z, \lambda)$, rather than assuming a mean SNR per Angstrom from measurements (figure 7.4). In the near future, we would also look to implementing the ability to compute combined tracer Fisher matrices in `lyaforecast`, rather than using low-noise approximations. To expand the capabilities of our forecast, we would look to using the `ForestFlow` emulator (Chaves-Montero et al., 2025), to combine small-scale information from the Lyα P1D with the Lyα P3D. The forecast can also be expanded to include full-shape measurements (see Cuceu et al., 2025a), which are sensitive to growth of structure. Finally, we can use the forecast to measure cosmological parameters from a variety of different models, e.g. w_0w_a dark energy, either in combination with low-redshift galaxy forecasts, or independently.

8

CONCLUSION

For thousands of years the cosmos has been studied, and our understanding of the Universe on the very largest scales has profoundly changed. In only the last 50 years, the era of precision cosmology was conceived, which allowed us for the first time to place tight constraints on the constituent parts of the Universe and its expansion history. With the introduction of DESI, precision cosmology has moved into its next chapter. We are now able to explore the nature of dark energy, determining whether or not it evolves, and validating a wide variety of physical models which try to describe the mysterious substance.

The first part of my thesis (chapter 5) is centred firmly in this objective, where I present the first cosmological analysis of Lyα forest data from DESI. This analysis served as validation of the first DESI data (Early Data Release; EDR), from the spectroscopic pipeline to the methods use to generate the Lyα flux transmission field and the models we use to measure cosmology. The techniques employed in this paper were established as a baseline for future DESI Lyα analyses (Adame et al., 2025b; DESI Collaboration et al., 2025b), which were used to measure the dark energy equation of state. We performed extensive validation of many aspects of our analysis, such as the Lyα catalogue and new methods used to construct it (section 5.2). We also explored the contaminant masking (DLAs, BALs), instrumental systematics (section 5.4.3.2) and the baseline model that we fit with our measured correlations (section 5.5). We showed that our correlations and fits to DESI EDR data were fully consistent with those from the eBOSS DR16 (the previous benchmark), a non-trivial result for two very different instruments.

To extract as much cosmological information as possible from the Lyα forest, we utilise the full shape of our measured Lyα correlation functions, from which we can measure the Alcock-Paczynski (AP) effect - anisotropy in the line-of-sight versus transverse directions - and growth information from RSD. Including AP information from the broadband of the correlation function improves our measurements of Ω_m and $H_0 r_d$ by around 50% with respect to traditional BAO-only analyses, but it is also more susceptible to contaminants. In chapter 6, we study of the impact of continuum redshift errors on our full-shape analysis, using synthetic DESI datasets. Up until now, this was not considered a major contaminant in our analysis, but at the high precision of DESI, we show it to be a potentially large source of bias. Thus, we model the distortion these errors cause in our measured correlation functions, and implement several new free parameters to our baseline full-shape fits. We show that our model successfully recovers unbiased constraints on the broadband AP parameter, and the isotropic/AP BAO parameters. With this work, we have paved the way for the first Lyα full-shape analyses with DESI, and more precise analyses beyond that. Looking forward, we wish to improve our model behaviour at smaller scales along the line-of-sight ($r_\parallel \lesssim 40\,h^{-1}$Mpc), where continuum redshift error contamination is less well captured. This is not particularly impactful now, since our analysis is limited to $r > 25\,h^{-1}$Mpc where our linear model is still suitable. However, in the future where models that successfully describe the non-linear parts of the Lyα correlation function (e.g. using an EFT approach; Ivanov, 2024) are implemented, this will be more important.

In the final part of the thesis (chapter 7), we present projections of potential BAO measurements from the Lyα forest in DESI-II - the successor to DESI due to start in 2029, which will observe 5000deg^2 of sky to much greater depth. We run Fisher forecasts that emulate the high-redshift survey of DESI-II, a programme which will observe targets above $z \gtrsim 2$, increasing the density of quasars (with respect to DESI) from 60 to 100 deg^{-2}. Furthermore, this programme will observe Lyman break galaxies (LBGs) and Lyman-α emitters (LAEs) at densities of ~1000 deg^{-2}, from $z \sim 2-4$. LBG spectra also contain Lyα forest absorption, albeit to a lower SNR than quasars. Thus, in our analysis we forecast several measurements, including the autocorrelation of Lyα forests from LBGs (`Ly`α `(LBG)`$^{\mathrm{auto}}$), the cross-correlation of LBG forests and LBGs (`Ly`α `(LBG)`$^{\mathrm{cross}}$), and the LBG autocorrelation (`LBG`$^{\mathrm{auto}}$). We also make use of the higher SNR quasar forests by taking their cross-correlation with LBGs (`Lya(QSO)xLBG`). Using LAE as tracers, we add a further three correlation functions of similar constraining power: `Ly`α `(LBG)xLAE`, `Ly`α `(QSO)xLAE` and `LAE`$^{\mathrm{auto}}$.

We show that the two best measurements of our BAO parameters $\alpha_\parallel$ and α_perp come from the combined `Ly`α `(LBG)`$^{\mathrm{auto}}$ and `Ly`α `(LBG)`$^{\mathrm{cross}}$ correlations and the LBG autocorrelation, which give ($\sigma_{\alpha_\parallel} = 0.70\%, \sigma_{\alpha_\perp} = 0.57\%$) and ($\sigma_{\alpha_\parallel} = 1.04\%, \sigma_{\alpha_\perp} = 0.59\%$) respectively. The `Lya(QSO)xLBG` correlation is marginally less constraining than its LBG counterpart, but it probes the Universe at a lower effective redshift. The LAE correlations provide slightly weaker constraints then LBGs due to lower tracer bias, but are still competitive.

Comparing only the combined `Ly`α `(LBG)`$^{\text{auto}}$ and `Ly`α `(LBG)`$^{\text{cross}}$ from DESI-II with the combined `Ly`α `(QSO)`$^{\text{auto}}$+`Ly`α `(QSO)`$^{\text{cross}}$ from DESI gives us $\sigma_{\alpha_\parallel}^{\text{DESI-II}} \approx 0.8\,\sigma_{\alpha_\parallel}^{\text{DESI}}$ and $\sigma_{\alpha_\perp}^{\text{DESI-II}} \approx 0.63\,\sigma_{\alpha_\perp}^{\text{DESI}}$. When combining all of our measurements, we expect DESI-II to provide significantly better high-redshift constraints than DESI.

These forecasts are an important first step to developing a science case for DESI-II, where we show that BAO can be measured to better precision than DESI, at a much higher effective redshift thanks to new LBG/LAE targeting. In the near future, to understand how well DESI-II can constrain dark energy parameters, we want to take all of our measured correlations, and combine them the Fisher matrix from the high-density galaxy survey at lower redshift. From there we can transform into a Fisher matrix with dependence on dark energy parameters, and make error projections. We also plan to use a full Fisher matrix in the near future, which takes into account cross-covariances between measurements, allowing us to combine them accurately. A slightly more difficult but important extension of our forecasts would be to include full-shape information, since these improve our constraining power on the expansion rate by $\sim 50\%$. The inclusion of LBGs and LAEs also gives us more tracers with which to measure RSD, helping to break degeneracies between biases and $f\sigma_8$.

BIBLIOGRAPHY

Einstein, A. (1915). “Die Feldgleichungen der Gravitation”. In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp. 844–847 (cit. on p. 1).

Zwicky, F. (1933). “Die Rotverschiebung von extragalaktischen Nebeln”. In: *Helvetica Physica Acta* 6, pp. 110–127 (cit. on p. 8).

Rao, C. R. (1945). “Information and the Accuracy Attainable in the Estimation of Statistical Parameters”. In: *Bulletin of the Calcutta Mathematical Society* 37, pp. 81–91 (cit. on p. 106).

Dicke, R. H. et al. (1965). “Cosmic Black-Body Radiation.” In: *ApJ* 142, pp. 414–419. DOI: `10.1086/148306` (cit. on p. 21).

Gunn, J. E. and B. A. Peterson (1965). “On the Density of Neutral Hydrogen in Intergalactic Space.” In: *ApJ* 142, pp. 1633–1636. DOI: `10.1086/148444` (cit. on p. 25).

Penzias, A. A. and R. W. Wilson (1965). “A Measurement of Excess Antenna Temperature at 4080 Mc/s.” In: *ApJ* 142, pp. 419–421. DOI: `10.1086/148307` (cit. on p. 21).

Lynds, R. (1971). “The Absorption-Line Spectrum of 4c 05.34”. In: Astrophys. J. Let. 164, p. L73. DOI: `10.1086/180695` (cit. on pp. 24, 31).

Meszaros, P. (1974). “The behaviour of point masses in an expanding cosmological substratum.” In: Astron. Astrophys. 37.2, pp. 225–228 (cit. on p. 16).

Guth, A. H. (1981). “Inflationary universe: A possible solution to the horizon and flatness problems”. In: Phys. Rev. ***D*** 23.2, pp. 347–356. DOI: `10.1103/PhysRevD.23.347` (cit. on p. 14).

Kaiser, N. (1987). “Clustering in real space and in redshift space”. In: Mon. Not. Roy. Astron. Soc. 227, pp. 1–21. DOI: `10.1093/mnras/227.1.1` (cit. on pp. 22, 73, 109).

Smoot, G. F. et al. (1992). "Structure in the COBE Differential Microwave Radiometer First-Year Maps". In: Astrophys. J. Let. 396, p. L1. DOI: 10.1086/186504 (cit. on pp. 13, 21).

Cen, R. et al. (1994). "Gravitational Collapse of Small-Scale Structure as the Origin of the Lyman-Alpha Forest". In: Astrophys. J. Let. 437, p. L9. DOI: 10.1086/187670. arXiv: astro-ph/9409017 [astro-ph] (cit. on p. 31).

Miralda-Escudé, J. et al. (1996). "The Ly alpha Forest from Gravitational Collapse in the Cold Dark Matter + Lambda Model". In: *ApJ* 471, p. 582. DOI: 10.1086/177992. arXiv: astro-ph/9511013 [astro-ph] (cit. on p. 31).

Hui, L. and N. Y. Gnedin (1997). "Equation of state of the photoionized intergalactic medium". In: *Monthly Notices of the Royal Astronomical Society* 292.1, pp. 27–42. DOI: 10.1093/mnras/292.1.27. URL: https://doi.org/10.1093\%2Fmnras\%2F292.1.27 (cit. on p. 31).

Tegmark, M. (1997). "Measuring Cosmological Parameters with Galaxy Surveys". In: Phys. Rev. Lett. 79.20, pp. 3806–3809. DOI: 10.1103/PhysRevLett.79.3806. arXiv: astro-ph/9706198 [astro-ph] (cit. on pp. 106, 113).

Croft, R. A. C. et al. (1998). "Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Lyα Forest". In: *ApJ* 495.1, pp. 44–62. DOI: 10.1086/305289. arXiv: astro-ph/9708018 [astro-ph] (cit. on p. 32).

Riess, A. G. et al. (1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant". In: *AJ* 116.3, pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph] (cit. on pp. 9, 65).

Perlmutter, S. et al. (1999). "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: *ApJ* 517.2, pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 [astro-ph] (cit. on pp. 9, 65).

Weinberg, D. and et al. (1999). "Cosmological tests with the Ly-α forest (invited review)". In: *Evolution of Large Scale Structure : From Recombination to Garching*. Ed. by A. J. Banday, R. K. Sheth, and L. N. da Costa, p. 346. DOI: 10.48550/arXiv.astro-ph/9810142. arXiv: astro-ph/9810142 [astro-ph] (cit. on p. 26).

McDonald, P. et al. (2000). "The Observed Probability Distribution Function, Power Spectrum, and Correlation Function of the Transmitted Flux in the Lyα Forest". In: *The Astrophysical Journal* 543, p. 1. DOI: 10.1086/317079 (cit. on p. 32).

Barkana, R. and A. Loeb (2001). "In the beginning: the first sources of light and the reionization of the universe". In: Phys. Rept. 349.2, pp. 125–238. DOI:

10.1016/S0370-1573(01)00019-9. arXiv: astro-ph/0010468 [astro-ph] (cit. on p. 25).

Zehavi, I. et al. (2002). "Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data". In: *ApJ* 571.1, pp. 172–190. DOI: 10.1086/339893. arXiv: astro-ph/0106476 [astro-ph] (cit. on p. 21).

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: *ApJ* 585.1, pp. 34–51. DOI: 10.1086/345945. arXiv: astro-ph/0108064 [astro-ph] (cit. on pp. 26, 31).

Seo, H.-J. and D. J. Eisenstein (2003). "Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys". In: *ApJ* 598.2, pp. 720–740. DOI: 10.1086/379122. arXiv: astro-ph/0307460 [astro-ph] (cit. on p. 107).

Weinberg, D. H. (2003). "The Lyman-α Forest as a Cosmological Tool". In: *AIP Conference Proceedings*. AIP. DOI: 10.1063/1.1581786. URL: https://doi.org/10.1063\%2F1.1581786 (cit. on p. 31).

Cole, S. et al. (2005). "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications". In: Mon. Not. Roy. Astron. Soc. 362.2, pp. 505–534. DOI: 10.1111/j.1365-2966.2005.09318.x. arXiv: astro-ph/0501174 [astro-ph] (cit. on pp. 21, 65).

Eisenstein, D. J. et al. (2005). "Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies". In: *ApJ* 633.2, pp. 560–574. DOI: 10.1086/466512. arXiv: astro-ph/0501171 [astro-ph] (cit. on pp. 21, 23, 65).

Gorski, K. M. et al. (2005). "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". In: *The Astrophysical Journal* 622.2, pp. 759–771. DOI: 10.1086/427976. URL: https://doi.org/10.1086\%2F427976 (cit. on p. 45).

McDonald, P. et al. (2005). "The Linear Theory Power Spectrum from the Lyα Forest in the Sloan Digital Sky Survey". In: *ApJ* 635.2, pp. 761–783. DOI: 10.1086/497563. arXiv: astro-ph/0407377 [astro-ph] (cit. on p. 116).

Fan, X., C. L. Carilli, and B. Keating (2006). "Observational Constraints on Cosmic Reionization". In: *Annual Review of Astron and Astrophys* 44.1, pp. 415–462. DOI: 10.1146/annurev.astro.44.051905.092514. arXiv: astro-ph/0602375 [astro-ph] (cit. on p. 25).

McDonald, P. et al. (2006). "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey". In: *The Astrophysical Journal Supplement Series* 163.1, pp. 80–109.

DOI: 10.1086/444361. URL: https://doi.org/10.1086 (cit. on pp. ix, 32, 41, 47, 59, 61).

Eisenstein, D. J., H.-J. Seo, and M. White (2007). "On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter". In: *The Astrophysical Journal* 664.2, pp. 660–674. DOI: 10.1086/518755. URL: https://doi.org/10.1086 (cit. on p. 47).

Eisenstein, D. J., H.-J. Seo, and M. White (2007). "On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter". In: *ApJ* 664.2, pp. 660–674. DOI: 10.1086/518755. arXiv: astro-ph/0604361 [astro-ph] (cit. on p. 116).

McDonald, P. and D. J. Eisenstein (2007). "Dark energy and curvature from a future baryonic acoustic oscillation survey using the Lyman-α forest". In: Phys. Rev. ***D*** 76.6, 063009, p. 063009. DOI: 10.1103/PhysRevD.76.063009. arXiv: astro-ph/0607122 [astro-ph] (cit. on p. 107).

McDonald, P. and U. Seljak (2009). "How to evade the sample variance limit on measurements of redshift-space distortions". In: *JCAP* 2009.10, 007, p. 007. DOI: 10.1088/1475-7516/2009/10/007. arXiv: 0810.0323 [astro-ph] (cit. on p. 122).

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. DOI: 10.1111/j.1365-2966.2008.14211.x. URL: https://doi.org/10.1111 (cit. on pp. 23, 48, 67, 87).

McQuinn, M. and M. White (2011). "On estimating Lyα forest correlations between multiple sightlines". In: Mon. Not. Roy. Astron. Soc. 415.3, pp. 2257–2269. DOI: 10.1111/j.1365-2966.2011.18855.x. arXiv: 1102.1752 [astro-ph.CO] (cit. on pp. 107, 114, 116).

Slosar, A. et al. (2011). "The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data". In: *JCAP* 2011.9, 001, p. 001. DOI: 10.1088/1475-7516/2011/09/001. arXiv: 1104.5244 [astro-ph.CO] (cit. on pp. 31, 32, 42, 66).

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: *JCAP* 2012.7, 028, p. 028. DOI: 10.1088/1475-7516/2012/07/028. arXiv: 1205.2018 [astro-ph.CO] (cit. on pp. 66, 67).

Font-Ribera, A. et al. (2012). "The large-scale cross-correlation of Damped Lyman alpha systems with the Lyman alpha forest: first measurements from BOSS". In: *Journal of Cosmology and Astroparticle Physics* 2012.11, pp. 059–059. DOI: 10.1088/1475-7516/2012/11/059. URL: https://doi.org/10.1088\%2F1475-7516\%2F2012\%2F11\%2F059 (cit. on pp. 37, 44, 51).

Busca, N. G. et al. (2013). "Baryon acoustic oscillations in the Lyα forest of BOSS quasars". In: *A&A* 552, A96. DOI: 10.1051/0004-6361/201220724. URL: https://doi.org/10.1051/0004-6361/201220724 (cit. on pp. 32, 65).

Font-Ribera, A. et al. (2013). "The large-scale quasar-Lyman α forest cross-correlation from BOSS". In: *JCAP* 2013.5, 018, p. 018. DOI: 10.1088/1475-7516/2013/05/018. arXiv: 1303.1937 [astro-ph.CO] (cit. on pp. 32, 49, 66, 67).

Kirkby, D. et al. (2013). "Fitting methods for baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9". In: *Journal of Cosmology and Astroparticle Physics* 2013.03, pp. 024–024. DOI: 10.1088/1475-7516/2013/03/024. URL: https://doi.org/10.1088 (cit. on pp. 32, 47, 66, 72).

Levi, M. et al. (2013). "The DESI Experiment, a whitepaper for Snowmass 2013". In: *arXiv e-prints*, arXiv:1308.0847, arXiv:1308.0847. arXiv: 1308.0847 [astro-ph.CO] (cit. on pp. 32, 66).

Palanque-Delabrouille, N. et al. (2013). "The one-dimensional Lyα forest power spectrum from BOSS". In: Astron. Astrophys. 559, A85, A85. DOI: 10.1051/0004-6361/201322130. arXiv: 1306.5896 [astro-ph.CO] (cit. on pp. 32, 110, 120).

Rudie, G. C. et al. (2013). "THE COLUMN DENSITY DISTRIBUTION AND CONTINUUM OPACITY OF THE INTERGALACTIC AND CIRCUMGALACTIC MEDIUM AT REDSHIFT $\langle z \rangle = 2.4$". In: *The Astrophysical Journal* 769.2, p. 146. DOI: 10.1088/0004-637x/769/2/146. URL: https://doi.org/10.1088\%2F0004-637x\%2F769\%2F2\%2F146 (cit. on p. 49).

Slosar, A. et al. (2013). "Measurement of baryon acoustic oscillations in the Lyman-α forest fluctuations in BOSS data release 9". In: *JCAP* 2013.4, 026, p. 026. DOI: 10.1088/1475-7516/2013/04/026. arXiv: 1301.3459 [astro-ph.CO] (cit. on pp. 32, 40).

Font-Ribera, A. et al. (2014a). "DESI and other Dark Energy experiments in the era of neutrino mass measurements". In: *JCAP* 2014.5, 023, p. 023. DOI: 10.1088/1475-7516/2014/05/023. arXiv: 1308.4164 [astro-ph.CO] (cit. on p. 122).

Font-Ribera, A. et al. (2014b). "Quasar-Lyman α forest cross-correlation from BOSS DR11: Baryon Acoustic Oscillations". In: *JCAP* 2014.5, 027, p. 027. DOI: 10.1088/1475-7516/2014/05/027. arXiv: 1311.1767 [astro-ph.CO] (cit. on pp. 32, 66).

Pieri, M. M. et al. (2014). "Probing the circumgalactic medium at high-redshift using composite BOSS spectra of strong Lyman α forest absorbers". In: Mon. Not. Roy. Astron. Soc. 441.2, pp. 1718–1740. DOI: 10.1093/mnras/stu577. arXiv: 1309.6768 [astro-ph.CO] (cit. on p. 67).

Arinyo-i-Prats, A. et al. (2015). "The non-linear power spectrum of the Lyman alpha forest". In: *JCAP* 2015.12, pp. 017–017. DOI: 10.1088/1475-7516/2015/12/017. arXiv: 1506.04519 [astro-ph.CO] (cit. on pp. 26, 31, 48).

Delubac, T. et al. (2015). "Baryon acoustic oscillations in the Lyα forest of BOSS DR11 quasars". In: *Astronomy & Astrophysics* 574, A59. DOI: 10.1051/0004-6361/201423969. URL: https://doi.org/10.1051 (cit. on pp. 32, 45).

Handley, W. J., M. P. Hobson, and A. N. Lasenby (2015). "Polychord: nested sampling for cosmology". In: *Monthly Notices of the Royal Astronomical Society: Letters* 450.1, pp. L61–L65. DOI: 10.1093/mnrasl/slv047. URL: https://doi.org/10.1093\%2Fmnrasl\%2Fslv047 (cit. on pp. 54, 61, 89).

DESI Collaboration et al. (2016a). "The DESI Experiment Part I: Science,Targeting, and Survey Design". In: *arXiv e-prints*, arXiv:1611.00036, arXiv:1611.00036. DOI: 10.48550/arXiv.1611.00036. arXiv: 1611.00036 [astro-ph.IM] (cit. on pp. 32, 35, 66, 116).

DESI Collaboration et al. (2016b). "The DESI Experiment Part II: Instrument Design". In: *arXiv e-prints*, arXiv:1611.00037, arXiv:1611.00037. DOI: 10.48550/arXiv.1611.00037. arXiv: 1611.00037 [astro-ph.IM] (cit. on pp. 33, 66).

Harris, D. W. et al. (2016). "The Composite Spectrum of BOSS Quasars Selected for Studies of the Lyα Forest". In: *AJ* 151.6, 155, p. 155. DOI: 10.3847/0004-6256/151/6/155. arXiv: 1603.08626 [astro-ph.GA] (cit. on pp. xv, 68, 69, 84, 94).

Shen, Y. et al. (2016). "The Sloan Digital Sky Survey Reverberation Mapping Project: Velocity Shifts of Quasar Emission Lines". In: *ApJ* 831.1, 7, p. 7. DOI: 10.3847/0004-637X/831/1/7. arXiv: 1602.03894 [astro-ph.GA] (cit. on pp. 67, 94).

Sorini, D. et al. (2016). "Modeling the Lyα Forest in Collisionless Simulations". In: *ApJ* 827.2, 97, p. 97. DOI: 10.3847/0004-637X/827/2/97. arXiv: 1602.08099 [astro-ph.CO] (cit. on p. 26).

Bautista, J. E. et al. (2017). "Measurement of baryon acoustic oscillation correlations at z = 2.3 with SDSS DR12 Lyα-Forests". In: Astron. Astrophys. 603, A12, A12. DOI: 10.1051/0004-6361/201730533. arXiv: 1702.00176 [astro-ph.CO] (cit. on pp. 32, 41, 47, 52, 53, 65, 116).

du Mas des Bourboux, H. et al. (2017). "Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at z = 2.4". In: Astron. Astrophys. 608, A130, A130. DOI: 10.1051/0004-6361/201731731. arXiv: 1708.02225 [astro-ph.CO] (cit. on pp. 32, 41, 49, 52, 116).

Zou, H. et al. (2017). "Project Overview of the Beijing-Arizona Sky Survey". In: *PASP* 129.976, p. 064101. DOI: `10.1088/1538-3873/aa65ba`. arXiv: `1702.03653 [astro-ph.GA]` (cit. on p. 33).

Blomqvist, M. 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, pp. 029–029. DOI: `10.1088/1475-7516/2018/05/029`. URL: `https://doi.org/10.1088\%2F1475-7516\%2F2018\%2F05\%2F029` (cit. on p. 52).

Busca, N. and C. Balland (2018). "QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks". In: *arXiv e-prints*, arXiv:1808.09955, arXiv:1808.09955. DOI: `10.48550/arXiv.1808.09955`. arXiv: `1808.09955 [astro-ph.IM]` (cit. on p. 34).

Font-Ribera, A., P. McDonald, and A. Slosar (2018). "How to estimate the 3D power spectrum of the Lyman-α forest". In: *JCAP* 2018.1, 003, p. 003. DOI: `10.1088/1475-7516/2018/01/003`. arXiv: `1710.11036 [astro-ph.CO]` (cit. on p. 40).

Gontcho A Gontcho, S. et al. (2018). "Quasar - CIV forest cross-correlation with SDSS DR12". In: Mon. Not. Roy. Astron. Soc. 480.1, pp. 610–622. DOI: `10.1093/mnras/sty1817`. arXiv: `1712.09886 [astro-ph.CO]` (cit. on p. 52).

Parks, D. et al. (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. DOI: `10.1093/mnras/sty196`. URL: `https://doi.org/10.1093\%2Fmnras\%2Fsty196` (cit. on p. 38).

Pérez-Ràfols, I. et al. (2018). "The SDSS-DR12 large-scale cross-correlation of damped Lyman alpha systems with the Lyman alpha forest". In: Mon. Not. Roy. Astron. Soc. 473.3, pp. 3019–3038. DOI: `10.1093/mnras/stx2525`. arXiv: `1709.00889 [astro-ph.CO]` (cit. on pp. 37, 88).

Rogers, K. K. et al. (2018). "Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers". In: Mon. Not. Roy. Astron. Soc. 476.3, pp. 3716–3728. DOI: `10.1093/mnras/sty603`. arXiv: `1711.06275 [astro-ph.CO]` (cit. on pp. 51, 66).

Blomqvist, Michael et al. (2019). "Baryon acoustic oscillations from the cross-correlation of Lyorption and quasars in eBOSS DR14". In: *A&A* 629, A86. DOI: `10.1051/0004-6361/201935641`. URL: `https://doi.org/10.1051/0004-6361/201935641` (cit. on p. 32).

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

2019.07, pp. 017–017. DOI: 10.1088/1475-7516/2019/07/017. URL: https://doi.org/10.1088\%2F1475-7516\%2F2019\%2F07\%2F017 (cit. on p. 32).

de Sainte Agathe, V. et al. (2019). "Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14". In: Astron. Astrophys. 629, A85, A85. DOI: 10.1051/0004-6361/201935638. arXiv: 1904.03400 [astro-ph.CO] (cit. on p. 32).

Dey, A. et al. (2019). "Overview of the DESI Legacy Imaging Surveys". In: *AJ* 157.5, 168, p. 168. DOI: 10.3847/1538-3881/ab089d. arXiv: 1804.08657 [astro-ph.IM] (cit. on pp. 29, 33).

du Mas des Bourboux, H. 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. DOI: 10.3847/1538-4357/ab1d49. URL: https://doi.org/10.3847 (cit. on pp. 44, 47, 59, 61).

Guo, Z. and P. Martini (2019). "Classification of Broad Absorption Line Quasars with a Convolutional Neural Network". In: *ApJ* 879.2, 72, p. 72. DOI: 10.3847/1538-4357/ab2590. arXiv: 1901.04506 [astro-ph.GA] (cit. on pp. 37, 67).

Ivezić, Ž. et al. (2019). "LSST: From Science Drivers to Reference Design and Anticipated Data Products". In: *ApJ* 873.2, 111, p. 111. DOI: 10.3847/1538-4357/ab042c. arXiv: 0805.2366 [astro-ph] (cit. on p. 116).

Lewis, A. (2019). "GetDist: a Python package for analysing Monte Carlo samples". In: arXiv: 1910.13970 [astro-ph.IM] (cit. on pp. 53, 61).

Zonca, A. et al. (2019). "healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python". In: *Journal of Open Source Software* 4.35, p. 1298. DOI: 10.21105/joss.01298. URL: https://doi.org/10.21105/joss.01298 (cit. on p. 45).

Bourboux, H. du Mas des et al. (2020). "THE COMPLETED SDSS-IV EXTENDED BARYON OSCILLATION SPECTROSCOPIC SURVEY: BARYON ACOUSTIC OSCILLATIONS WITH LYMAN-α FORESTS". In: *ApJ* 901.2, p. 153. DOI: 10.3847/1538-4357/abb085. arXiv: 2007.08995 (cit. on pp. viii, ix, 32–37, 39–47, 49–52, 54, 57, 59–61, 65, 69, 94).

Cuceu, A., A. Font-Ribera, and B. Joachimi (2020). "Bayesian methods for fitting Baryon Acoustic Oscillations in the Lyman-α forest". In: *JCAP* 2020.7, 035, p. 035. DOI: 10.1088/1475-7516/2020/07/035. arXiv: 2004.02761 [astro-ph.CO] (cit. on p. 51).

Farr, J., A. Font-Ribera, and A. Pontzen (2020a). "Optimal strategies for identifying quasars in DESI". In: *JCAP* 2020.11, 015, p. 015. DOI: 10.1088/1475-7516/2020/11/015. arXiv: 2007.10348 [astro-ph.CO] (cit. on pp. 34, 69).

Farr, J. et al. (2020b). “LyaCoLoRe: synthetic datasets for current and future Lyman-α forest BAO surveys”. In: *JCAP* 2020.3, 068, p. 068. DOI: `10.1088/1475-7516/2020/03/068`. arXiv: `1912.02763 [astro-ph.CO]` (cit. on p. 69).

Planck Collaboration et al. (2020). “Planck 2018 results. VI. Cosmological parameters”. In: Astron. Astrophys. 641, A6, A6. DOI: `10.1051/0004-6361/201833910`. arXiv: `1807.06209 [astro-ph.CO]` (cit. on pp. 32, 47, 53, 65, 71, 87, 89, 109, 120).

Tie, S. S. et al. (2020). “The DESI sky continuum monitor system”. In: *Ground-based and Airborne Instrumentation for Astronomy VIII*. Ed. by C. J. Evans, J. J. Bryant, and K. Motohara. Vol. 11447. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1144785, p. 1144785. DOI: `10.1117/12.2561436`. arXiv: `2101.11736 [astro-ph.IM]` (cit. on p. 30).

Yèche, C. et al. (2020). “Preliminary Target Selection for the DESI Quasar (QSO) Sample”. In: *Research Notes of the American Astronomical Society* 4.10, 179, p. 179. DOI: `10.3847/2515-5172/abc01a`. arXiv: `2010.11280 [astro-ph.CO]` (cit. on p. 117).

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: Phys. Rev. ***D*** 103.8, 083533, p. 083533. DOI: `10.1103/PhysRevD.103.083533`. arXiv: `2007.08991 [astro-ph.CO]` (cit. on p. 31).

Cuceu, A. et al. (2021). “Cosmology beyond BAO from the 3D distribution of the Lyman-α forest”. In: Mon. Not. Roy. Astron. Soc. 506.4, pp. 5439–5450. DOI: `10.1093/mnras/stab1999`. arXiv: `2103.14075 [astro-ph.CO]` (cit. on pp. 66, 72, 86).

Foroozan, S., A. Krolewski, and W. J. Percival (2021). “Testing large-scale structure measurements against Fisher matrix predictions”. In: *JCAP* 2021.10, 044, p. 044. DOI: `10.1088/1475-7516/2021/10/044`. arXiv: `2106.11432 [astro-ph.CO]` (cit. on p. 116).

Ho, M.-F., S. Bird, and 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, pp. 704–719. ISSN: 0035-8711. DOI: `10.1093/mnras/stab2169`. eprint: `https://academic.oup.com/mnras/article-pdf/507/1/704/39771986/stab2169.pdf`. URL: `https://doi.org/10.1093/mnras/stab2169` (cit. on p. 38).

McGreer, I., J. Moustakas, and J. Schindler (2021). *simqso: Simulated quasar spectra generator*. Astrophysics Source Code Library, record ascl:2106.008 (cit. on pp. 69, 94).

Alexander, D. M. et al. (2022). *The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra*. DOI: 10.48550/ARXIV.2208.08517. URL: https://arxiv.org/abs/2208.08517 (cit. on pp. 34, 35).

Cuceu, A. et al. (2022a). "New constraints on the expansion rate at redshift 2.3 from the Lyman-α forest". In: *arXiv e-prints*, arXiv:2209.13942, arXiv:2209.13942. DOI: 10.48550/arXiv.2209.13942. arXiv: 2209.13942 [astro-ph.CO] (cit. on p. 32).

Cuceu, A. et al. (2022b). "The Alcock-Paczyński effect from Lyman-α forest correlations: Analysis validation with synthetic data". In: *arXiv e-prints*, arXiv:2209.12931, arXiv:2209.12931. DOI: 10.48550/arXiv.2209.12931. arXiv: 2209.12931 [astro-ph.CO] (cit. on p. 51).

DESI Collaboration et al. (2022). "Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument". In: *AJ* 164.5, 207, p. 207. DOI: 10.3847/1538-3881/ac882b. arXiv: 2205.10939 [astro-ph.IM] (cit. on pp. 27, 33, 66).

Ennesser, L. et al. (2022). "The impact and mitigation of broad-absorption-line quasars in Lyman α forest correlations". In: Mon. Not. Roy. Astron. Soc. 511.3, pp. 3514–3523. DOI: 10.1093/mnras/stac301. arXiv: 2111.09439 [astro-ph.CO] (cit. on pp. 37, 66).

Givans, J. J. et al. (2022). "Non-linearities in the Lyman-α forest and in its cross-correlation with dark matter halos". In: *JCAP* 2022.9, 070, p. 070. DOI: 10.1088/1475-7516/2022/09/070. arXiv: 2205.00962 [astro-ph.CO] (cit. on p. 31).

Ramírez-Pérez, C. et al. (2022). "CoLoRe: fast cosmological realisations over large volumes with multiple tracers". In: *JCAP* 2022.5, 002, p. 002. DOI: 10.1088/1475-7516/2022/05/002. arXiv: 2111.05069 [astro-ph.CO] (cit. on p. 68).

Wang, B. et al. (2022). "Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems". In: *ApJS* 259.1, 28, p. 28. DOI: 10.3847/1538-4365/ac4504. arXiv: 2201.00827 [astro-ph.GA] (cit. on pp. 38, 67).

Yang, L. et al. (2022). "Metal Lines Associated with the Lyα Forest from eBOSS Data". In: *The Astrophysical Journal* 935.2, p. 121. DOI: 10.3847/1538-4357/ac7b2e. URL: https://doi.org/10.3847\%2F1538-4357\%2Fac7b2e (cit. on p. 51).

Youles, S. et al. (2022). "The effect of quasar redshift errors on Lyman-α forest correlation functions". In: Mon. Not. Roy. Astron. Soc. 516.1, pp. 421–433. DOI: 10.1093/mnras/stac2102. arXiv: 2205.06648 [astro-ph.CO] (cit. on pp. 67, 69, 75, 77, 81, 86, 93, 95, 97).

Alexander, D. M. et al. (2023). "The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra". In: *AJ* 165.3, 124, p. 124. DOI: 10.3847/1538-3881/acacfc. arXiv: 2208.08517 [astro-ph.GA] (cit. on p. 67).

Brodzeller, A. et al. (2023). "Performance of the Quasar Spectral Templates for the Dark Energy Spectroscopic Instrument". In: *AJ* 166.2, 66, p. 66. DOI: 10.3847/1538-3881/ace35d. arXiv: 2305.10426 [astro-ph.IM] (cit. on pp. 34, 63, 68, 94).

Chaussidon, E. et al. (2023). "Target Selection and Validation of DESI Quasars". In: *ApJ* 944.1, 107, p. 107. DOI: 10.3847/1538-4357/acb3c2. arXiv: 2208.08511 [astro-ph.CO] (cit. on pp. 33–35, 66).

Cuceu, A. et al. (2023a). "Constraints on the Cosmic Expansion Rate at Redshift 2.3 from the Lyman-α Forest". In: Phys. Rev. Lett. 130.19, 191003, p. 191003. DOI: 10.1103/PhysRevLett.130.191003. arXiv: 2209.13942 [astro-ph.CO] (cit. on pp. 66, 72).

Cuceu, A. et al. (2023b). "The Alcock-Paczyński effect from Lyman-α forest correlations: analysis validation with synthetic data". In: Mon. Not. Roy. Astron. Soc. 523.3, pp. 3773–3790. DOI: 10.1093/mnras/stad1546. arXiv: 2209.12931 [astro-ph.CO] (cit. on p. 72).

DESI Collaboration et al. (2023a). "The Early Data Release of the Dark Energy Spectroscopic Instrument". In: *arXiv e-prints*, arXiv:2306.06308, arXiv:2306.06308. DOI: 10.48550/arXiv.2306.06308. arXiv: 2306.06308 [astro-ph.CO] (cit. on p. 33).

DESI Collaboration et al. (2023b). "Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument". In: *arXiv e-prints*, arXiv:2306.06307, arXiv:2306.06307. DOI: 10.48550/arXiv.2306.06307. arXiv: 2306.06307 [astro-ph.CO] (cit. on pp. 34, 60).

Gordon, C. et al. (2023). "3D correlations in the Lyman-α forest from early DESI data". In: *JCAP* 2023.11, 045, p. 045. DOI: 10.1088/1475-7516/2023/11/045. arXiv: 2308.10950 [astro-ph.CO] (cit. on pp. 30, 71).

Guy, J. et al. (2023). "The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument". In: *AJ* 165.4, 144, p. 144. DOI: 10.3847/1538-3881/acb212. arXiv: 2209.14482 [astro-ph.IM] (cit. on pp. 29, 33, 34, 36, 49, 66).

Hadzhiyska, B. et al. (2023). "Planting a Lyman alpha forest on ABACUS-SUMMIT". In: Mon. Not. Roy. Astron. Soc. 524.1, pp. 1008–1024. DOI: 10.1093/mnras/stad1920. arXiv: 2305.08899 [astro-ph.CO] (cit. on pp. 56, 57).

Hahn, C. et al. (2023). "The DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation". In: *AJ* 165.6, 253, p. 253. DOI: `10.3847/1538-3881/accff8`. arXiv: `2208.08512 [astro-ph.CO]` (cit. on pp. 33, 66).

Kirkby et al. 2023 (2023). "in preparation". In: (cit. on p. 33).

Myers, A. D. et al. (2023). "The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument". In: *The Astronomical Journal* 165.2, p. 50. DOI: `10.3847/1538-3881/aca5f9`. URL: `https://doi.org/10.3847` (cit. on pp. 33, 35).

Pérez-Ràfols, I. et al. (2023). "The cross-correlation of galaxies in absorption with the Lyman α forest". In: Mon. Not. Roy. Astron. Soc. 524.1, pp. 1464–1477. DOI: `10.1093/mnras/stad1994`. arXiv: `2210.02973 [astro-ph.CO]` (cit. on pp. 37, 66).

Raichoor, A. et al. (2023). "Target Selection and Validation of DESI Emission Line Galaxies". In: *AJ* 165.3, 126, p. 126. DOI: `10.3847/1538-3881/acb213`. arXiv: `2208.08513 [astro-ph.CO]` (cit. on pp. 33, 66).

Schlafly, E. F. et al. (2023). "Survey Operations for the Dark Energy Spectroscopic Instrument". In: *AJ* 166.6, 259, p. 259. DOI: `10.3847/1538-3881/ad0832`. arXiv: `2306.06309 [astro-ph.CO]` (cit. on pp. 27, 33, 66).

Schlegel et al. 2023 (2023). "in preparation". In: (cit. on p. 33).

Silber, J. H. et al. (2023). "The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)". In: *AJ* 165.1, 9, p. 9. DOI: `10.3847/1538-3881/ac9ab1`. arXiv: `2205.09014 [astro-ph.IM]` (cit. on p. 33).

Zhou, R. et al. (2023). "Target Selection and Validation of DESI Luminous Red Galaxies". In: *AJ* 165.2, 58, p. 58. DOI: `10.3847/1538-3881/aca5fb`. arXiv: `2208.08515 [astro-ph.CO]` (cit. on pp. 33, 66).

DESI Collaboration et al. (2024a). "DESI 2024 VII: Cosmological Constraints from the Full-Shape Modeling of Clustering Measurements". In: *arXiv e-prints*, arXiv:2411.12022, arXiv:2411.12022. DOI: `10.48550/arXiv.2411.12022`. arXiv: `2411.12022 [astro-ph.CO]` (cit. on p. 66).

DESI Collaboration et al. (2024b). "Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument". In: *AJ* 167.2, 62, p. 62. DOI: `10.3847/1538-3881/ad0b08`. arXiv: `2306.06307 [astro-ph.CO]` (cit. on pp. 116, 122).

Filbert, S. et al. (2024). "Broad absorption line quasars in the Dark Energy Spectroscopic Instrument Early Data Release". In: Mon. Not. Roy. Astron. Soc. 532.4, pp. 3669–3681. DOI: `10.1093/mnras/stae1610`. arXiv: `2309.03434 [astro-ph.CO]` (cit. on p. 37).

Ivanov, M. M. (2024). "Lyman alpha forest power spectrum in effective field theory". In: Phys. Rev. ***D*** 109.2, 023507, p. 023507. DOI: `10.1103/PhysRevD.109.023507`. arXiv: `2309.10133 [astro-ph.CO]` (cit. on p. 126).

Miller, T. N. et al. (2024). "The Optical Corrector for the Dark Energy Spectroscopic Instrument". In: *AJ* 168.2, 95, p. 95. DOI: `10.3847/1538-3881/ad45fe`. arXiv: `2306.06310 [astro-ph.IM]` (cit. on pp. 33, 66).

Poppett, C. et al. (2024). "Overview of the Fiber System for the Dark Energy Spectroscopic Instrument". In: *AJ* 168.6, 245, p. 245. DOI: `10.3847/1538-3881/ad76a4` (cit. on pp. 27, 33, 66).

Ramírez-Pérez, C. et al. (2024). "The Lyman-α forest catalogue from the Dark Energy Spectroscopic Instrument Early Data Release". In: Mon. Not. Roy. Astron. Soc. 528.4, pp. 6666–6679. DOI: `10.1093/mnras/stad3781`. arXiv: `2306.06312 [astro-ph.CO]` (cit. on pp. 30, 33, 36, 38–41, 57, 61, 70).

Ruhlmann-Kleider, V. et al. (2024). "High redshift LBGs from deep broadband imaging for future spectroscopic surveys". In: *JCAP* 2024.8, 059, p. 059. DOI: `10.1088/1475-7516/2024/08/059`. arXiv: `2404.03569 [astro-ph.CO]` (cit. on pp. 117, 120).

Adame, A. G. et al. (2025a). "DESI 2024 III: baryon acoustic oscillations from galaxies and quasars". In: *JCAP* 2025.4, 012, p. 012. DOI: `10.1088/1475-7516/2025/04/012`. arXiv: `2404.03000 [astro-ph.CO]` (cit. on p. 66).

Adame, A. G. et al. (2025b). "DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman alpha forest". In: *JCAP* 2025.1, 124, p. 124. DOI: `10.1088/1475-7516/2025/01/124`. arXiv: `2404.03001 [astro-ph.CO]` (cit. on pp. 30, 41, 66, 68, 71, 73, 74, 88, 89, 98, 125).

Adame, A. G. et al. (2025c). "DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations". In: *JCAP* 2025.2, 021, p. 021. DOI: `10.1088/1475-7516/2025/02/021`. arXiv: `2404.03002 [astro-ph.CO]` (cit. on pp. 30, 65, 66).

Bailey, S. J. et al. (2025). "Redrock, in prep". In: (cit. on pp. 29, 34, 67).

Bault, A. et al. (2025). "Impact of systematic redshift errors on the cross-correlation of the Lyman-α forest with quasars at small scales using DESI Early Data". In: *JCAP* 2025.1, 130, p. 130. DOI: `10.1088/1475-7516/2025/01/130`. arXiv: `2402.18009 [astro-ph.CO]` (cit. on pp. 33, 44, 49, 60, 68, 94).

Casas, L. et al. (2025). "Validation of the DESI DR2 Lyα BAO analysis using synthetic datasets". In: *arXiv e-prints*, arXiv:2503.14741, arXiv:2503.14741. DOI: `10.48550/arXiv.2503.14741`. arXiv: `2503.14741 [astro-ph.IM]` (cit. on pp. 67, 95).

Chaves-Montero, J. et al. (2025). “ForestFlow: predicting the Lyman-α forest clustering from linear to nonlinear scales”. In: Astron. Astrophys. 694, A187, A187. DOI: 10.1051/0004-6361/202452039. arXiv: 2409.05682 [astro-ph.CO] (cit. on p. 124).

Cuceu, A. et al. (2025a). “DESI DR1 Lyα forest: 3D full-shape analysis and cosmological constraints, in prep.” In: (cit. on pp. 30, 71, 87–89, 91, 92, 94, 124).

Cuceu, A. et al. (2025b). “Validation of the DESI 2024 Lyα forest BAO analysis using synthetic datasets”. In: *JCAP* 2025.1, 148, p. 148. DOI: 10.1088/1475-7516/2025/01/148. arXiv: 2404.03004 [astro-ph.CO] (cit. on pp. 67, 68).

DESI Collaboration et al. (2025a). “Data Release 1 of the Dark Energy Spectroscopic Instrument”. In: *arXiv e-prints*, arXiv:2503.14745, arXiv:2503.14745. DOI: 10.48550/arXiv.2503.14745. arXiv: 2503.14745 [astro-ph.CO] (cit. on pp. 30, 68, 70).

DESI Collaboration et al. (2025b). “DESI DR2 Results I: Baryon Acoustic Oscillations from the Lyman Alpha Forest”. In: *arXiv e-prints*, arXiv:2503.14739, arXiv:2503.14739. DOI: 10.48550/arXiv.2503.14739. arXiv: 2503.14739 [astro-ph.CO] (cit. on pp. 23, 71, 95, 120, 124, 125).

DESI Collaboration et al. (2025c). “DESI DR2 Results II: Measurements of Baryon Acoustic Oscillations and Cosmological Constraints”. In: *arXiv e-prints*, arXiv:2503.14738, arXiv:2503.14738. DOI: 10.48550/arXiv.2503.14738. arXiv: 2503.14738 [astro-ph.CO] (cit. on pp. i, 9, 23).

Gordon, C. et al. (2025). “Modelling the impact of quasar redshift errors on the full-shape analysis of correlations in the Lyman-*a* forest”. In: *arXiv e-prints*, arXiv:2505.08789, arXiv:2505.08789. DOI: 10.48550/arXiv.2505.08789. arXiv: 2505.08789 [astro-ph.CO] (cit. on p. 30).

Guy, J. et al. (2025). “Characterization of contaminants in the Lyman-alpha forest auto-correlation with DESI”. In: *JCAP* 2025.1, 140, p. 140. DOI: 10.1088/1475-7516/2025/01/140. arXiv: 2404.03003 [astro-ph.CO] (cit. on pp. xv, 33, 48, 49, 60, 61).

Herrera-Alcantar, H et al. (2025a). “The Lyman-α forest 3D correlation function from LBG spectra: First measurement with DESI and BAO prospects., in prep”. In: (cit. on pp. xii, 116–119).

Herrera-Alcantar, H. K. et al. (2025b). “Synthetic spectra for Lyman-α forest analysis in the Dark Energy Spectroscopic Instrument”. In: *JCAP* 2025.1, 141, p. 141. DOI: 10.1088/1475-7516/2025/01/141. arXiv: 2401.00303 [astro-ph.CO] (cit. on pp. 33, 39, 68, 69).

J Zou. et al. (2025). “in preparation”. In: (cit. on pp. 38, 59).

Payerne, C. et al. (2025). “Selection of high-redshift Lyman-Break Galaxies from broadband and wide photometric surveys”. In: *JCAP* 2025.5, 031, p. 031. DOI: `10.1088/1475-7516/2025/05/031`. arXiv: `2410.08062 [astro-ph.CO]` (cit. on pp. 116, 117).

2025 Measuring cosmic expansion and growth with the Lyman-α forest in DESI Calum Gordon UAB