UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

TESIS DOCTORAL

Cosmología con la Medida de la Escala de las Oscilaciones Acústicas de Bariones en Grandes Cartografiados de Galaxias

Cosmology from the Measurement of the Baryon Acoustic Oscillations Scale in Large Galaxy Surveys

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juan Mena Fernández

DIRECTORES

Eusebio Sánchez Álvaro

Ignacio Sevilla Noarbe

Universidad Complutense de Madrid
Facultad de Ciencias Físicas
Programa de Doctorado en Física

Tesis Doctoral

Cosmología con la Medida de la Escala de las Oscilaciones Acústicas de Bariones en Grandes Cartografiados de Galaxias

Cosmology from the Measurement of the Baryon Acoustic Oscillations Scale in Large Galaxy Surveys

Memoria para optar al grado de Doctor

presentada por

Juan Mena Fernández

Directores

Dr. Eusebio Sánchez Álvaro
Dr. Ignacio Sevilla Noarbe

A mi familia y amigos

Agradecimientos

Hubo un tiempo en el que me planteé si añadir agradecimientos a la tesis o no. Sin embargo, agradecer a la gente que ha hecho posible este trabajo y con la que he compartido tantas experiencias en los últimos 4 años me parece una parte muy importante del proceso, demasiado como para no hacerlo. Además, para qué vamos a engañarnos, a algunas de las personas que estáis por aquí esta es la parte de la tesis que más os interesa, así que espero que la disfrutéis.

A mis directores de tesis, Eusebio y Nacho. Sin vosotros, esto no habría sido posible. Muchas gracias por la flexibilidad que me habéis dado a la hora de poder participar en diferentes proyectos y por todas las oportunidades para asistir a congresos, reuniones de colaboración y escuelas, ya que todas estas experiencias me han hecho crecer como científico. Gracias, también, por todas las cosas que me habéis enseñado.

Aparte de mis directores, quiero darles las gracias a los demás integrantes del grupo de cosmología del CIEMAT, pasados y presentes: Aurelio, Jacobo, Juan de Vicente y Laura. Gracias por todas nuestras reuniones de los viernes por la mañana. Aprovecho para agradecer a toda la gente que ha hecho posible mi trabajo en los clusters del CIEMAT, en particular a Javier y Juanjo. Muchas gracias, también, a Luis Manuel, mi tutor de la UCM, por ser un gran apoyo durante mi primer año de tesis.

Este párrafo se lo dedico a la gente que conocí y con la que trabajé durante mi estancia en Columbus. Ashley, thank you very much for everything. You are one of the people who helped me the most getting involved in DESI, and I have learnt so many things from you, particularly during the last year. I want to thank John, Klaus and Paul as well, since they made me feel part of their group during my stay. Y, por supuesto, no puedo olvidarme de mi compañero de despacho en Columbus, Walter. Muchas gracias por hacer mi estancia mucho más amena. Por último, muchas gracias también a Andrei, Iván, Kevin, Naim y Sumit, por tantos buenos momentos.

Me gustaría agradecer, también, a todas las personas de las colaboraciones DES y DESI que han hecho posible mi trabajo. Por un lado, en DES me gustaría darles las gracias a Alex, Carles, Giulia, Hugo, Ismael, Jack, Kwan y Noah. I would like to thank Jim Annis as well, since he has been my DES mentor for more than 2 years now. Thank you very much for your advice and all our useful discussions, scientific and non-scientific. Por otro lado, en DESI estoy especialmente agradecido a Alejandro,

Antoine, Boryana, Cristhian, Cullan, Etienne, Hee-Jong, Héctor, Mariana, Misha, Nikhil, Otávio, Samuel, Sandy, Segev, Sesh, Stephen y Uendert. A special thanks to Arnaud for keeping the DESI Python environment up to date, and for always fixing the errors that appear when running codes in NERSC. Finalmente, me gustaría hacerles una especial mención a Anna y Santi: muchísimas gracias por todo lo que me habéis aportado y enseñado durante estos últimos años.

Ahora, les llega el turno a mis compañeros del CIEMAT. Muchísimas gracias a todos por estos últimos 4 años. Sin duda, una de las cosas que más voy a echar de menos es el buen ambiente que tenemos en el departamento. Quiero empezar agradeciendo a Martín, por ser mi padre científico. Me enseñaste lo que era una función de correlación y muchísimas otras cosas más durante mi primer año de tesis. Muchas gracias, también, a los que ya no están en el CIEMAT: Ana, Andrés, Beymar, Chiara, Diana, Édgar, Héctor, Irene, José, Mab y Sergio. Con algunos de vosotros compartí el viaje a Formigal, y lo recuerdo con muchísimo cariño. Alejo, gracias por enseñarme que una salamandra no es lo mismo que una salamanquesa. Antonio, por ser siempre tan alegre y organizar rutas (aunque aún tenga pendiente apuntarme a alguna). Camille, por enseñarme francés. Carmen (¿quién es Carmen?), por ser mi compañera de despacho durante tantos años. Cecilia, por ser una de esas personas con las que se puede hablar de cualquier cosa. Dani, por todas nuestras conversaciones sobre VGC. David, por ser mi hermano pequeño científico estos últimos años y haber compartido tantos viajes conmigo. Diego, por tu gran sentido del humor, aunque a veces haya que llamar a tu madre por tu mal comportamiento. Elvira, por escoger el camino de la mandarina sin dudarlo ni un segundo. Iker, por ser un infiltrado en el edificio 1 y siempre luchar por nuestros derechos. Iñaki, por enseñarme a usar una boca de incendios. Irene, por ser una muy buena compañera de despacho (de esas que no molestan mucho, para compensar conmigo y con Pepe). Jaime, por ser mi amigo durante tantos años (¿10 años ya?). Gracias por ser tan buen compañero de laboratorio durante nuestros años de universidad. Julia, gracias por dejarme tu silla siempre que tomábamos el café en vuestro despacho. Muchas gracias, también, por enseñarnos tantas expresiones andaluzas. Laura, gracias por convertirte en mi compañera de línea 10 y ser siempre tan agradable. Lourdes, por ser un encanto y por tener tan buen gusto para la comida. Ludo, por no dejarte engañar nunca con el español (y mira que lo hemos intentado). Miguel Ángel, por tus consejos e infinita sabiduría. Natalia, por la excusa de "tengo baile", ahora la utilizo para todo. Nicola, por tu entusiasmo con el trabajo desde el primer momento. Probablemente serás la persona que heredará mi sitio (si los teléfonos lo permiten), así que tendrás que cuidar de todos mis objetos personales. Oliver, gracias por aumentar la representación talaverana en el CIEMAT. Pepe, por ser mi compañero de despacho de los viernes por la tarde, aunque siga sin cuadrarme eso de que te guste el fútbol. Muchas gracias por interesarte tanto por mi y preguntarme infinidad de veces cómo es mi firma o si me gusta Frank Ocean. Rodrigo, gracias por acompañarme a las barras en Benasque. Sergio, por ser la mejor persona del CIEMAT, espero que estés disfrutando de tu estancia. Por último, gracias a Yuan, por ser un amante de los videojuegos y un

torturador de kologs.

Este párrafo se lo dedico a mi grupo de amigos de la universidad: Alba, Carol L, Carol M, Jaime, Jesús, María y Roberto, por apoyarme siempre, por nuestros reencuentros y nuestras casas rurales. Se lo dedico, también, a mis amigos de Talavera: Alejandro, André, Antonio, Bea, David, Flor, Inés, Jenni, Jorge, Juanal, Mónica, Sandra y Xu Hao. Muchas gracias por estar ahí siempre, a pesar de los distintos caminos tomados. Muchas gracias, también, a Marc y Raúl, por esas noches de Fortnite y tantos buenos momentos.

Por supuesto, me gustaría darle las gracias, también, a mi familia, por su apoyo incondicional en todo momento. Aunque no sepan exactamente de qué trata mi trabajo, sé que siempre se preocupan por mi. Muchas gracias a mis padres, Juan y Nati; a mis hermanos, Alberto y Marta; a mis abuelas, Concha y Úrsula; a mis primas, Cristina, Mar y Natacha; a mis tíos, Carmen, Conchi, Estrella, Eularico, Fernando, Fide y Jesús; y a mi suegra, Mari. Muchísimas gracias por ser un pilar fundamental en mi vida, y por haber confiado en mi en todo momento. Por último, pero no menos importante, quiero darle las gracias a Carol. Has sido mi mayor apoyo durante los últimos años, la persona con la que más agradecido estoy y, sin duda, la más importante en mi vida. Muchísimas gracias por estar ahí siempre, en los buenos y en los malos momentos.

Contents

List of Figures

List of Tables

List of Abbreviations

2dFGRS	2dF Galaxy Redshift Survey
2LPT	Second-order Lagrangian perturbation theory
6dFGS	6dF Galaxy Survey
BAO	Baryon acoustic oscillations
BBN	Big Bang nucleosynthesis
BGS	Bright galaxy sample
BOSS	Baryon Oscillation Spectroscopic Survey
CAMB	Code for Anisotropies in the Microwave Background
CCD	Charge-coupled device
CDM	Cold dark matter
CFHTLenS	Canada-France Hawaii Telescope Lensing Survey
CMASS	Constant-mass Sample
CMB	Cosmic microwave background
COLA	Comoving Lagrangian acceleration
DECam	Dark Energy Camera
DES	Dark Energy Survey
DESDM	DES Data Management

DESI	Dark Energy Spectroscopic Instrument
DETF	Dark Energy Task Force
DL	Distance ladder
DNF	Directional neighborhood fitting
eBOSS	Extended BOSS
ELG	Emission line galaxies
FLRW	Friedmann-Lemaître-Robertson-Walker
HOD	Halo Occupation Distribution
HSC-SSP	Hyper Suprime-Cam Subaru Strategic Program
ICE	Institut de Ciències de l'Espai
IFFT	Iterative fast Fourier transform
KiDS	Kilo-Degree Survey
LOS	Line of sight
LOWZ	Low-Redshift (sample)
LRG	Luminous red galaxies
LSS	Large-scale structure
LSST	Legacy Survey of Space and Time
MCMC	Markov chain Monte Carlo
MGS	Main Galaxy Sample (SDSS-I and II)
MICE	Marenostrum Institut de Ciències de l'Espai
MLE	Maximum likelihood estimator
MOF	Multi-object fitting
QSO	Quasi-stellar objects, or quasars
RSD	Redshift-space distortions
SDSS	Sloan Digital Sky Survey
SN	Supernovae
SOF	Single-object fitting
SV	Science verification

WL	Weak lensing
WMAP	Wilkinson Microwave Anisotropy Probe
Y1	DES or DESI year 1 observing periods and data analyses
Y3	DES year 3 observing periods and data analyses
Y6	DES year 6 observing periods and data analyses

Abstract

"Cosmology from the Measurement of the Baryon Acoustic Oscillations Scale in Large Galaxy Surveys"

Cosmology has undergone a revolution in the last few decades, driven by advancements in observational techniques, theoretical models, and computational capabilities. Among its most significant achievements is the establishment of ΛCDM as the standard cosmological model, which has transformed our understanding of the composition and evolution of the Universe. ΛCDM describes the Universe as composed mainly of dark energy in the form of a cosmological constant, Λ; cold dark matter (CDM); and ordinary baryonic matter. Each of these constituents account for a 70%, 25% and 5% of the total energy density of the Universe, respectively. Dark energy is the physical entity that drives the accelerated expansion of the Universe. It is, by far, the dominant component. Dark matter, on the other hand, does not interact with electromagnetic radiation, which makes it invisible to telescopes. Its presence is inferred through its gravitational effects on visible matter and the large-scale structure (LSS) of the Universe. Finally, ordinary baryonic matter, which is composed mainly of protons, neutrons, and electrons, makes up the familiar matter we can see and interact with. ΛCDM is based on the cosmological principle, which posits that the Universe is homogeneous and isotropic on large scales, and on general relativity as the theory that describes the gravitational interaction. It successfully explains a wide range of cosmological observations, including the accelerated expansion of the Universe, the cosmic microwave background (CMB) radiation, the abundances of light elements, and the large-scale distribution of galaxies.

Large galaxy surveys have played a pivotal role in increasing our understanding of cosmology within the ΛCDM framework. These surveys observe hundreds of millions of galaxies across a wide range of redshifts, providing valuable information about the LSS of the Universe. Their statistical power has been greatly improved with the development of advanced observational techniques and the use of state-of-the-art telescopes and instruments. One of the key cosmological probes used in large galaxy surveys is the measurement of the baryon acoustic oscillations (BAO) scale, which are imprints of primordial sound waves from the early Universe that left characteristic features in the large-scale distribution of matter. The BAO serves as a "standard ruler" in cosmology, providing precise measurements of cosmic distances that can be

used to constrain cosmological parameters, such as the matter density, the Hubble constant, and the nature of dark energy. It also allows us to perform consistency tests with high-redshift experiments, such as Planck.

This dissertation has been developed within the Dark Energy Survey (DES) and the Dark Energy Spectroscopic Instrument (DESI) collaborations. These projects aim to set tight constraints on the nature of dark energy and, also, to study the LSS of the Universe. DES is a Stage-III photometric galaxy survey that took its data between 2013 and 2019, using the 4-meter Blanco telescope in Chile to observe an area of about 5,000 square degrees of the southern sky. DES studies the distribution and properties of dark energy and dark matter through several complementary techniques, including weak gravitational lensing, galaxy clustering, and supernovae measurements. The analysis of its complete dataset, the Y6 analysis, is still ongoing. On the other hand, DESI is a Stage-IV spectroscopic galaxy survey that started its 5-year survey in 2021, using a new instrument installed on the 4-meter Mayall telescope at the Kitt Peak National Observatory in Arizona, USA, to observe an area of about 16,000 square degrees. Among its main scientific goals are the precise study of BAO and redshift-space distortions.

In this thesis we develop a template-based BAO-fitting pipeline to extract the BAO information from the two-point angular correlation function. This pipeline is tested in Chapter 4 using the Y3 COLA mocks, the fiducial simulations of the DES Y3 analysis. We show that its results are robust against variations in the template cosmology and, also, against variations in the default settings for the BAO fits. Once validated, our pipeline is run on the DES Y3 data, for which we obtain a distance measurement of $d_M(0.835)/r_d = 18.94 \pm 0.48$, which is consistent with Planck at the 2.5σ level. This represents the most precise measurement from a photometric galaxy survey up to date, with a relative error of 2.6%. We finalize Chapter 4 combining our BAO likelihood with the other main DES Y3 likelihoods, namely 3×2pt and SN. We find tighter constraints in h, Ω_m and σ_8 after the combination: the posterior for h is more symmetrical, with a gain in constraining power of ∼20%; the error in Ω_m is reduced by ∼25%; and the constraining power in σ_8 improves by ∼16%. On the other hand, in Chapter 5 we describe the DES Y6 BAO analysis. We first perform an optimization of the Y6 sample, and we forecast an increase in precision of about 25% in the BAO measurement with respect to the Y3 analysis. Then, we run our BAO-fitting pipeline on the Y6 COLA mocks, where we find that the increase in precision with respect to the Y3 ones is consistent with what we expect from forecasts. Since the Y6 analysis is ongoing and it is still on its blinded phase, in this manuscript we do not show the BAO-fit results on the Y6 data. However, we do combine a synthetic 3×2pt datavector with the expected BAO likelihood obtained from the simulations. We find an increase in constraining power of ∼18% in Ω_m and ∼16% in σ_8 with respect to 3×2pt alone.

In this thesis we also study the impact of halo occupation distribution (HOD) systematics for the BAO measurement, i.e., systematics related to the modeling of the galaxy-halo connection. In particular, in Chapter 6 we estimate this effect for

luminous red galaxies (LRG) in the context of the DESI Y1 analysis. For this study, we run BAO fits on a total of 8 sets of 25 AbacusSummit box simulations, being each set generated assuming different HOD models, but with similar a number density to that of the DESI Y1 LRG sample. We find an upper limit for the HOD systematics error budget of $\sigma_{\rm HOD} \leq 0.2\sigma_{\rm stat}$ for the BAO shift parameter, α, in both the parallel and the perpendicular directions with respect to the line of sight. Future work will allow us to estimate the amount of stochasticity in this result, since we want to remove this contribution and leave only the one related to the HOD modeling itself.

Resumen

"Cosmología con la Medida de la Escala de las Oscilaciones Acústicas de Bariones en Grandes Cartografiados de Galaxias"

La cosmología ha experimentado una revolución en las últimas décadas, impulsada por los avances en técnicas de observación, modelos teóricos y capacidades computacionales. Entre sus logros más significativos está el establecimiento de ΛCDM como el modelo cosmológico estándar, el cual ha transformado nuestra comprensión de la composición y evolución del Universo. ΛCDM describe el Universo como compuesto principalmente de energía oscura en forma de una constante cosmológica, Λ; materia oscura fría (CDM, de sus siglas en inglés); y materia bariónica ordinaria. Cada uno de estos constituyentes supone un 70%, 25% y 5% de la densidad energética total del Universo, respectivamente. La energía oscura es la entidad física causante de la expansión acelerada del Universo. Es, de lejos, la componente dominante. La materia oscura, por otro lado, no interactúa con la radiación electromagnética, lo que la hace invisible para los telescopios. Su presencia se infiere a través de sus efectos gravitatorios sobre la materia visible y la estructura a gran escala (LSS, de sus siglas en inglés) del Universo. Finalmente, la materia bariónica ordinaria, que se compone principalmente de protones, neutrones y electrones, constituye la materia familiar que podemos ver y con la que interactuamos. ΛCDM se basa en el principio cosmológico, que postula que el Universo es homogéneo e isótropo a gran escala, y en la relatividad general como la teoría que describe la interacción gravitatoria. Explica con éxito una amplia gama de observaciones cosmológicas, incluida la expansión acelerada del Universo, la radiación del fondo cósmico de microondas (CMB, de sus siglas en inglés), la abundancia de elementos ligeros y la distribución de las galaxias a gran escala.

Los grandes cartografiados de galaxias han representado un papel fundamental en el aumento de nuestra comprensión de la cosmología dentro del marco del modelo ΛCDM. Estos cartografiados observan cientos de millones de galaxias en una amplia gama de *redshifts*, lo que proporciona información valiosa sobre la LSS del Universo. Su poder estadístico ha mejorado mucho con el desarrollo de técnicas de observación avanzadas y el uso de telescopios e instrumentos de última generación. Una de las sondas cosmológicas clave utilizada en los estudios de grandes cartografiados de galaxias es la medición de la escala de las oscilaciones acústicas de bariones (BAO, de sus

siglas en inglés), que son las huellas de las ondas de sonido primordiales del Universo primitivo y dejaron rasgos característicos en la distribución a gran escala de la materia. La escala de las BAO sirve como una "regla estándar" en cosmología, proporcionando mediciones precisas de distancias cósmicas que pueden usarse para medir parámetros cosmológicos, como la densidad de la materia, la constante de Hubble y la naturaleza de la energía oscura. También nos permite realizar pruebas de consistencia con experimentos de alto *redshift*, como Planck.

Esta tesis ha sido desarrollada dentro de las colaboraciones Dark Energy Survey (DES) y Dark Energy Spectroscopic Instrument (DESI). Estos proyectos tienen como objetivo establecer medidas precisas sobre la naturaleza de la energía oscura y, también, estudiar la LSS del Universo. DES es un cartografiado de galaxias fotométrico de tercera generación que tomó sus datos entre 2013 y 2019, utilizando el telescopio Blanco de 4 metros en Chile para observar un área de unos 5.000 grados cuadrados del cielo. DES estudia la distribución y las propiedades de la energía oscura y la materia oscura a través de varias técnicas complementarias, incluidas las mediciones de lentes gravitacionales débiles, agrupamiento de galaxias y supernovas. El análisis de su conjunto de datos completo, el análisis del año 6, aún está en curso. Por otro lado, DESI es un cartografiado de galaxias espectroscópico de cuarta generación que comenzó su toma de datos en 2021, utilizando un nuevo instrumento instalado en el telescopio Mayall de 4 metros en el Kitt Peak National Observatory en Arizona, EE. UU., con el objetivo de observar un área de unos 16.000 grados cuadrados. Entre sus principales objetivos científicos se encuentran el estudio preciso de la escala BAO y las distorsiones en el espacio de *redshift*.

En esta tesis desarrollamos un método de ajuste de la escala BAO basado en plantillas para extraer la información sobre dicha escala de la función de correlación angular. Estudiamos la aplicabilidad y validez de dicho método en el Capítulo 4, usando las simulaciones COLA del año 3. Mostramos que sus resultados son robustos frente a variaciones en la cosmología de la plantilla y, también, frente a variaciones en la configuración predeterminada para los ajustes de la escala BAO. Una vez validado, el método se aplica los datos del año 3 de DES, donde obtenemos una medida de la distancia de $d_M(0,835)/r_d = 18,94 \pm 0,48$, que es consistente con Planck a $2,5\sigma$. Esto representa la medida de la escala BAO más precisa de un cartografiado de galaxias fotométrico hasta la fecha, con un error relativo del 2,6%. Finalizamos el capítulo 4 combinando nuestra función de verosimilitud (*likelihood function*) de BAO con las otras principales del año 3 de DES, a saber, 3×2pt y SN. Encontramos medidas más precisas de h, Ω_m y σ_8 después de la combinación: la distribución de probabilidad de h es más simétrica, con una ganancia en el poder de restricción de ∼20%; el error en Ω_m se reduce en ∼25%; y la determinación de σ_8 mejora en ∼16%. Por otro lado, en el Capítulo 5 describimos el análisis del año 6 de DES. Primero, realizamos una optimización de la muestra del año 6, y prevemos un aumento de precisión de un 25% en la medida de la escala BAO con respecto al análisis del año 3. Luego, utilizamos nuestro método de medida de la escala BAO en las simulaciones COLA del año 6, donde encontramos que el aumento en la precisión con respecto al año 3 es consistente

con lo que esperamos de las previsiones. Dado que el análisis del año 6 está en curso y todavía está en su fase ciega, en este manuscrito no mostramos los resultados del ajuste de BAO en los datos del año 6. Sin embargo, combinamos un vector de datos sintético de 3×2pt con la función de verosimilitud esperada de BAO obtenida de las simulaciones. Encontramos un aumento en el poder de restricción de ∼18% en Ω_m y ∼16% en σ_8 con respecto a 3×2pt solo.

En esta tesis también estudiamos el impacto de las incertidumbres sistemáticas relacionadas con la distribución de ocupación en halos (HOD, de sus siglas en inglés). En particular, en el Capítulo 6 estimamos este efecto para galaxias rojas luminosas (LRG, de sus siglas en inglés) en el contexto del análisis del año 1 de DESI. Para este estudio, hacemos ajustes de la escala BAO en un total de 8 conjuntos de 25 simulaciones AbacusSummit, siendo cada conjunto generado asumiendo diferentes modelos de HOD, pero con una densidad de galaxias similar a la de la muestra de LRGs del año 1 de DESI. Encontramos un límite superior para el error sistemático de HOD de $\sigma_{\rm HOD} \leq 0,2\sigma_{\rm stat}$, tanto para $\alpha_\parallel$ como para $\alpha_\perp$. Futuros trabajos nos permitirán estimar la cantidad de estocasticidad en este resultado, con el fin de eliminar esta contribución y dejar solo la relacionada con el propio modelado del HOD.

Chapter 1

The Standard Cosmological Model

The Universe is, by definition, the ensemble of all existing physical reality. Sometimes we write universe to refer to a specific model of the Universe. The Big Bang is the accepted theory for the evolution of the Universe during its first instants of life. It was proposed by Georges Lemaître in 1927 [1], who realized that an expanding Universe could be traced back in time to a single point from which it originated, which he called the "primeval atom". For several decades, the scientific community was divided between those who defended the Big Bang and those who defended a static Universe. In 1929, Edwin Hubble concluded, by analyzing galactic redshifts, that galaxies are separating from each other, which is an important evidence that the Universe is expanding [2]. In 1964, Arno Pencias and Robert Woodrow Wilson discovered the radiation of the cosmic microwave background (CMB) [3], which was a crucial proof in favor of the Big Bang, since this model predicts a uniform background radiation that fills the entire Universe.

The standard cosmological model is also known as the ΛCDM model, and it is a representation of the Big Bang cosmological model in which the Universe contains three major components: first, a cosmological constant, Λ, which is associated with dark energy; second, the cold dark matter (CDM); and third, ordinary matter (usually referred to as baryonic matter). The Universe also contains radiation, but its contribution to the total energy density is negligible at the present cosmological time (it was, however, the dominant component in the early Universe, and therefore we must also account for it in many situations). The ΛCDM model is based upon the theory of general relativity and the hypothesis that the Universe is homogeneous and isotropic, the so-called cosmological principle; and it is strongly supported by a vast observational basis, such as the CMB radiation, the primordial nucleosynthesis and the large-scale structure (LSS).

Within the Big Bang cosmological framework, observations confirm that the ΛCDM model provides the current best description of the Universe, since it accounts for a

host of observational results. There are, however, several "problems"[1] that arise in the Big Bang theory, such as the flatness and horizon problems and the origin of the large-scale structure. All of these problems can be solved by the so-called theory of inflation, which is a period of exponential expansion that happened during the early stages of the Universe. It was proposed by Alexei Starobinski in 1979 in the Soviet Union [4], and simultaneously by Alan Guth in the United States [5]. These are known as the "old" inflationary models. The so-called "new" inflationary models, or *slow-roll* models, were developed by Andreas Albrecht and Paul Steinhardt [6], and also by Andrei Linde [7]. When the inflationary phase comes to its end, the Universe turns into a "soup" of electrons, protons, photons, neutrinos and many other particles in thermal equilibrium that expands with time. We will go into further detail about inflation in section 1.5, and we will dedicate Chapter 2 to describe the LSS of the Universe.

1.1 Characteristics of the Universe

Observations show that the Universe is expanding in an accelerated way. This accelerated expansion phase is thought to have begun when the Universe entered its dark-energy-dominated era, about 5 billion years ago. The nature of dark energy remains unknown; however, observations indicate that dark energy is the cosmological constant, Λ, a positive vacuum energy that fills up the Universe. Unveiling the nature of dark energy and studying its properties is the main motivation for this thesis.

In Fig. 1.1 we show a diagram of the evolution of the Universe with the main events occurred during its ∼14 billion years of life. From the initial quantum fluctuations, which are the seeds of the structures we see today, to the accelerated expansion caused by dark energy, the Universe has undergone a vast amount of different processes.

1.1.1 Composition

The Universe consists almost entirely of dark energy (∼70%), cold dark matter (∼25%) and baryonic matter (∼5%). Other components, such as electromagnetic radiation, contribute only marginally. Dark energy accounts for most of the energy content of the Universe, and its study is one of the main motivations of this thesis. The following chart summarizes the different contributions to the energy content of

[1] These are not exactly problems with the ΛCDM model, but rather something that we do not know how to explain, particularly why some parameters seem to be fine-tuned.

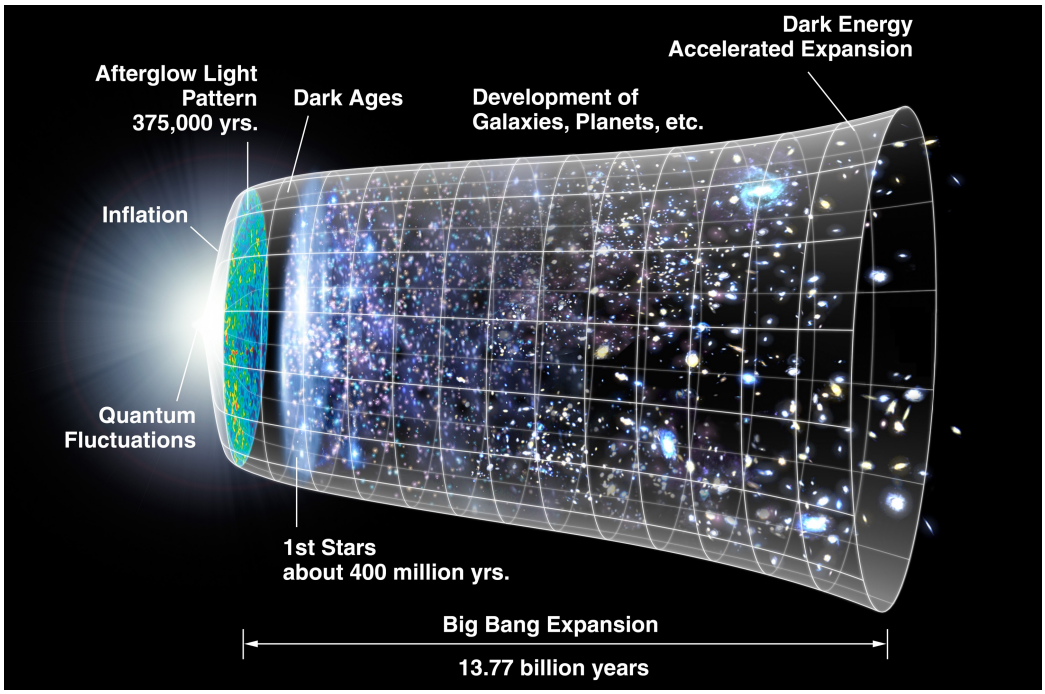

Figure 1.1: Diagram of the evolution of the Universe with the main events occurred during its ∼14 billion years of life. Credits: NASA (https://images.nasa.gov/).

the Universe:

$$\text{Energy content} \begin{cases} \text{dark energy } (\sim 70\%) \\ \text{matter } (\sim 30\%) \begin{cases} \text{cold dark matter } (\sim 25\%) \\ \text{baryonic matter } (\sim 5\%) \\ \text{non-relativistic neutrinos } (< 1\%) \end{cases} \\ \text{radiation } (< 1\%) \begin{cases} \text{photons } (< 1\%) \\ \text{relativistic neutrinos } (< 1\%) \end{cases} \end{cases}$$

1.1.1.1 Dark Energy

The accelerated expansion of the Universe is the observation that the expansion of the Universe is such that the velocity at which a distant galaxy is receding from the observer is continuously increasing with time [8]. This is caused by dark energy, a form of energy that affects the Universe on the largest scales and accounts for ∼70% of its energy density. The first observational evidence for its existence came from supernovae measurements [9]. Understanding the evolution of the Universe requires knowledge of its initial conditions and its composition. Prior to these observations, the only forms of matter-energy known to exist were ordinary baryonic matter, dark matter and radiation.

Dark energy mainly affects the expansion rate of the Universe. In turn, this affects the redshift-distance relationship and the growth of structures. Although dark energy has been important in recent times, we expect its effects at large redshifts to be small, since this would prevent the growth of LSS in most models. Some of the main probes of dark energy are [8]:

1. The number of clusters as a function of redshift. Galaxy clusters are the largest virialized objects in the Universe. In the context of the CDM paradigm, the number density of dark matter halos as a function of the redshift and the mass of the halo can be accurately predicted from N-body simulations. Comparing these predictions with studies of clusters that extend to high redshift, we can obtain precise restrictions on the cosmic expansion of the Universe.

2. The weak lensing (WL) effect in the distribution of galaxies. The gravitational deviation of the light path by structures in the Universe distorts the images of distant galaxies. This distortion allows us to measure the distribution of dark matter and its evolution over time, thus demonstrating the influence of dark energy on the growth of structures.

   The statistical signal due to the gravitational lensing effect of LSS is called the "cosmic shear". The cosmic shear field at a point in the sky is estimated by locally averaging the shapes of large numbers of distant galaxies.

3. The baryon acoustic oscillations (BAO) measurement. The peaks and valleys found in the angular spectrum of temperature fluctuations of the CMB (C_ℓ) arise from acoustic oscillations of the baryon-photon plasma in the early Universe. The scale of these oscillations is given by the sound horizon at the time of the matter-radiation decoupling, i.e., the distance r_d that these sound waves could have traveled until then. This distance r_d (hereafter, the sound horizon scale), can be expressed as

$$r_d = \int_0^{t_d} dt\ c_s(z)(1+z) = \int_{z_d}^{\infty} dz \frac{c_s(z)}{H(z)}, \tag{1.1}$$

   where c_s is the speed of sound in the primordial plasma, $H(z)$ is the Hubble parameter and z_d is the redshift corresponding to the photon decoupling. The speed of sound c_s is determined by the fraction of the energy densities of baryons and photons. Since the scale of the sound horizon provides a standard ruler calibrated by the CMB anisotropy, measurements of the scale of baryon acoustic oscillations in the distribution of galaxies give us information on the cosmic expansion of the Universe. The study of the properties of dark energy via the BAO feature is one of the main goals of this thesis.

4. Redshift-space distortions (RSD). One of the most effective ways to test gravity on large scales is to exploit the apparent anisotropies observed in galaxy maps,

the so-called redshift-space distortions. These anisotropies are due to the peculiar velocities of galaxies, and allow us to constrain the linear growth rate of cosmic structures,

$$\frac{d\log\delta_G}{d\log a}, \tag{1.2}$$

where δ_G is the galaxy overdensity field and a is the scale factor.

5. The Hubble diagram of type Ia supernovae (SN). Type Ia supernovae are "standard candles", which in practice means that, provided their brightness (l), they constrain cosmic acceleration through the Hubble diagram. The first direct evidence of cosmic acceleration came from type Ia supernovae [9], which have also provided the strongest constraints on the parameter of the dark energy equation of state.

1.1.1.2 Dark Matter

Dark matter is hypothesized to be a form of matter that accounts for, approximately, the 80% of the matter in the Universe, and about 25% of its total energy density. Because dark matter has not yet been observed directly, if it exists, it must barely interact with ordinary baryonic matter and radiation, except through gravity. Most dark matter must be non-baryonic in nature; it may be composed of some as-yet undiscovered subatomic particles. The primary candidate for dark matter is some new kind of elementary particle, in particular, weakly interacting massive particles (WIMPs) [10]. The first observational evidences of dark matter were galactic rotation curves and galactic velocity dispersion.

1. Galactic rotation curves. At the scale of a galaxy, Newtonian gravity is suitable to describe gravitational interactions. Therefore, the rotational velocity of an object that is located at a distance r from the center of a galaxy is given by

$$v(r) = \sqrt{\frac{GM(r)}{r}}, \tag{1.3}$$

where $M(r)$ is the mass inside the sphere of radius r, i.e.,

$$M(r) = \int_0^r dr \int_0^\pi d\theta \int_0^{2\pi} d\varphi \; r^2 \sin(\theta)\rho(r,\theta,\varphi). \tag{1.4}$$

When r is small and $\rho \approx$ const, $M(r) \propto r^3$ and $v \propto r$. However, when r is large, $M(r) \approx M$ and $v \propto 1/\sqrt{r}$. We do not observe that galactic rotation curves decrease as $1/\sqrt{r}$ for $r > R_{\text{galaxy}}$, but as $v \approx$ const instead. This means that there is a contribution of matter (which we call dark) such that it extends into a halo with density ρ_{h}.

2. Galactic velocity dispersion. Stars in bound systems must obey the virial theorem. This can be applied to any bound system, from an elliptical galaxy to a globular cluster. The theorem states that

$$\langle E_c \rangle = -\frac{1}{2}\langle E_p \rangle. \tag{1.5}$$

Taking $\langle E_c \rangle = m_g v^2/2$ and $\langle E_p \rangle = -GM_c m_g/r$, we find that

$$\langle v^2 \rangle = -GM_c \left\langle \frac{1}{r} \right\rangle, \tag{1.6}$$

where m_g and M_c are the masses of the galaxy and the cluster, respectively. The mass of the cluster calculated from the previous expression, using the measured velocity distribution, is larger than the measured luminous mass. The difference between these two is the amount of dark matter in the cluster.

1.2 Pillars of the Standard Cosmological Model

In this section we briefly describe the theory of general relativity, which, together with the aforementioned cosmological principle, is one of the main pillars of the ΛCDM model. On the other hand, it is also important to mention that this model has a strong observational basis. Among many other observations, it is supported by the radiation of the CMB, the primordial nucleosynthesis and abundance of light elements and the LSS of the Universe.

1.2.1 The Theory of General Relativity

Gravitation is, by far, the dominant force in the cosmological context. The theory of general relativity is the current description of gravitation in modern physics, which makes it a fundamental ingredient of the standard cosmological model. General relativity provides a unified description of gravity as a geometrical property of spacetime: the curvature of spacetime is directly related to the energy of whatever matter and radiation are present in the Universe. This relation is specified by the Einstein field equations, which are usually expressed as

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.7}$$

where $G_{\mu\nu}$ is the Einstein tensor,

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

and $T_{\mu\nu}$ is the stress-energy tensor [11]. The contribution of the cosmological constant, Λ, can be interpreted as a stress-energy tensor defined as

$$T^{\Lambda}_{\mu\nu} = -\frac{c^4 \Lambda}{8\pi G} g_{\mu\nu}. \tag{1.9}$$

The Einstein field equations can be obtained by finding the extreme of the so-called Hilbert-Einstein action, whose Lagrangian density is given by

$$\mathscr{L}_E = \frac{c^4}{16\pi G}(R - 2\Lambda)\sqrt{-g}. \tag{1.10}$$

1.2.2 The Cosmological Principle

Most modern cosmology is built on the cosmological principle: the hypothesis that all positions and directions in the Universe are, essentially, equivalent [12, 13]. This homogeneity does not apply to the Universe in detail, but to the Universe averaged over regions that include significant numbers of galaxies. At these scales, and at higher, the Universe is homogeneous and isotropic [12].

At small scales, the Universe presents a variety of different structures. From smaller to larger structures we find galaxies (∼10 Kpc), galaxy clusters (∼10 Mpc), walls, filaments and voids (∼100 Mpc). From this point on, the Universe is essentially homogeneous and isotropic, i.e., no structures can be identified. This is the so-called cosmological principle, and it states that, viewed from sufficiently large scales, the properties of the Universe are the same for any comoving observer. The cosmological principle also assumes that the same physical laws apply in all places of our Universe, which implies that the place we occupy is not a special position within its immensity.

The concepts of homogeneity and isotropy may seem equivalent but they represent different properties. Whenever we say that the Universe is homogeneous we mean that it looks the same at any point, i.e., there are no preferred or special observers (invariance of physical laws under translations). On the other hand, whenever we say that the Universe is isotropic we mean that it looks the same in every direction (invariance of physical laws under rotations). A consequence of isotropy is that there is no center of the Universe and of its expansion, since it expands from any point to any other point. In Fig. 1.2 we show an illustration of the concepts of homogeneity and isotropy and how they are not equivalent.

1.2.2.1 Mathematical Formulation of the Cosmological Principle

We have already seen that the cosmological principle assumes that the Universe is homogeneous and isotropic at sufficiently large scales, and that the same physical laws apply to every single part of it. In this section we show how it is formulated mathematically, starting from the definition of the so-called comoving coordinates and ending up with the derivation of the Friedmann-Lemaître-Robertson-Walker metric.

Comoving Coordinates

Observations up to date confirm the applicability of the cosmological principle. The structures observed at large scales and the radiation from the CMB, also at large scales, are almost perfectly isotropic for us. This is, of course, after removing

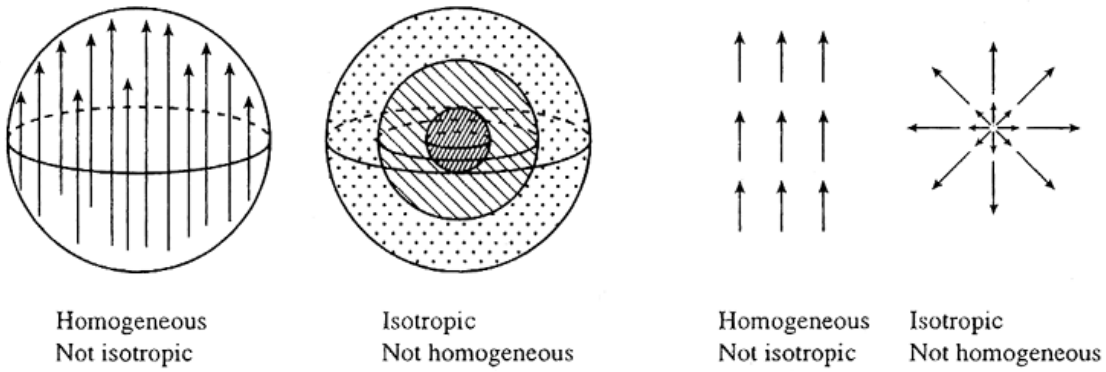

Figure 1.2: Illustration of the concepts of homogeneity and isotropy. The two images on the left represent a homogeneous but anisotropic 3D Universe and an isotropic but inhomogeneous 3D Universe, respectively. The two images on the right represent the same but in 2D. Credits: http://abyss.uoregon.edu/~js/cosmo/lectures/lec05.html.

the contribution from the dipole anisotropy present on the CMB due to the relative velocity between the Earth within the Milky Way and the last scattering surface. Comoving observers are those whose reference system is such that they see a null dipole anisotropy in the CMB. In this privileged system, the matter distribution and the CMB are observed as homogeneous and isotropic. The line of motion of comoving observers is given by

$$x^i = \text{const}, \quad t = \tau, \tag{1.11}$$

where τ is the so-called cosmic time. In order to calculate the line element for these observers, it is necessary to set the different components of the metric. Comoving observers' clocks measure cosmic time, which means that

$$dt = d\tau. \tag{1.12}$$

Since $dx^i = 0$, we easily find that $ds^2 = -c^2 d\tau^2$, which implies

$$g_{00} = -1. \tag{1.13}$$

In order to fix the other components of the metric tensor, we make use of the geodesic equations,

$$\frac{d^2 x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \tag{1.14}$$

1. $\lambda = i$: it is deduced that $\Gamma^i_{00} = 0$ and, from its definition in terms of the metric, it follows that $g_{0j}(x^\mu) = g_{0j}(\boldsymbol{x})$.

2. $\lambda = 0$: it is deduced that $\Gamma^0_{00} = 0$. This does not impose anything new on the metric.

For an isotropic metric it is possible to make $g_{0i} \equiv 0$ (so there is no vector, no privileged direction in space). Thus, an isotropic line element in comoving coordinates can be written as

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -c^2dt^2 + g_{ij}dx^i dx^j. \tag{1.15}$$

The spatial part of the metric, g_{ij}, corresponds to the metric of a maximally symmetric space.

Maximally Symmetric Spaces

Maximally symmetric spaces are defined as those spaces that have the maximum number of Killing vectors possible (or maximum number of symmetries). In a D-dimensional space, a Killing vector $\xi_\mu(x)$ has D independent components, just like any other D-vector. On the other hand, $\nabla_\mu\xi_\nu(x)$ only has $D(D-1)/2$ independent components. This is because of the Killing equation,

$$\nabla_\mu\xi_\nu(x) - \nabla_\nu\xi_\mu(x) = 0, \tag{1.16}$$

which implies $\nabla_\mu\xi_\nu(x)$ is antisymmetric. Therefore, a D-dimensional Killing vector $\xi_\mu(x)$ is defined by

$$N = D + \frac{D(D-1)}{2} = \frac{D(D+1)}{2} \tag{1.17}$$

independent components. Hence, if we take a set of Killing vectors $\{\xi_\mu\}$, the initial values for $\xi_\mu(x)$ and $\nabla_\mu\xi_\nu(x)$ should be thought of as the components of vectors in an $D(D+1)/2$-dimensional vector space, and so the maximum number of linearly-independent Killing vectors is $D(D+1)/2$. A space that has this number of Killing vectors for its metric is known as a maximally symmetric space.

For a maximally symmetric space, it can be shown that

$$R_{\rho\lambda\mu\kappa} = \frac{R}{D(D-1)}(g_{\rho\mu}g_{\lambda\kappa} - g_{\rho\kappa}g_{\lambda\mu}), \tag{1.18}$$

from which it follows that

$$R_{\mu\kappa} = \frac{1}{D}g_{\mu\kappa}R. \tag{1.19}$$

R is constant since the Universe has to be homogeneous, otherwise there would be special locations from the spatial curvature point of view. The Ricci scalar is usually expressed as

$$R = D(D-1)K, \tag{1.20}$$

where K is the curvature of space.

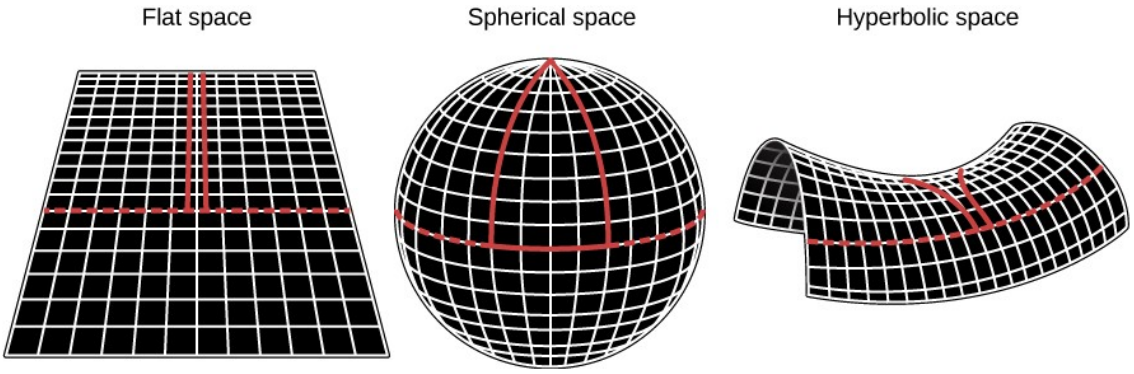

Figure 1.3: Illustration of the concepts of Euclidean or flat space (left), spherical or positively curved space (center) and hyperbolic or negatively curved space (right). The red lines in each diagram are initially parallel but follow different paths depending on the curvature of space. Credits: https://www.coursehero.com/study-guides/astronomy/the-cosmic-microwave-background/.

1.2.2.2 The Friedmann-Lemaître-Robertson-Walker Metric

Maximally symmetric spaces can be constructed for different values of the curvature K. We are interested in three different cases: $K = 0$, which corresponds to an Euclidean space[2]; $K > 0$, which corresponds to a spherically curved space; $K < 0$, which corresponds to a hyperbolic curved space. In Fig. 1.3 we show an illustration for these three cases.

The Friedmann-Lemaître-Robertson-Walker metric (or FLRW metric) is the generalization of these three cases. The line element for the FLRW metric is given by

$$ds^2 = -c^2 dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2 \theta d\varphi^2 \right) \right], \tag{1.21}$$

where

$$\begin{cases} k = 1 & \leftrightarrow \quad K > 0, \\ k = 0 & \leftrightarrow \quad K = 0, \\ k = -1 & \leftrightarrow \quad K < 0. \end{cases} \tag{1.22}$$

A change of variables gives a more compact expression for the FLRW metric,

$$ds^2 = -c^2 dt^2 + a(t)^2 \left[d\chi^2 + S_k(\chi)^2 \left(d\theta^2 + \sin^2 \theta d\varphi^2 \right) \right], \tag{1.23}$$

where

$$d\chi = \frac{dr}{\sqrt{1 - kr^2}}. \tag{1.24}$$

[2] It actually corresponds to a Minkowski spacetime if we also include the time coordinate.

It should be noted that $\chi(0) = S_k^{-1}(0) = 0\ \forall k$, which is relevant for the following calculations. Integrating the previous expression, we find that

$$\chi(r) = \int_0^{\chi} d\chi = \int_0^r \frac{dr'}{\sqrt{1-kr'^2}} = \begin{cases} \dfrac{\sin^{-1}(\sqrt{k}r)}{\sqrt{k}}, & k > 0, \\ r, & k = 0, \\ \dfrac{\sinh^{-1}(\sqrt{|k|}r)}{\sqrt{|k|}}, & k < 0. \end{cases} \tag{1.25}$$

The expression for $S_k(\chi)$ can then be compactly written as

$$r = S_k(\chi) = \chi\ \mathrm{sinc}(\chi\sqrt{k}). \tag{1.26}$$

Let d be the physical or proper distance to an object. We need to relate this physical distance to the radial coordinate r that appears in the FLRW metric. The relation between these two can be directly computed as

$$d = \int_0^r ds = a\int_0^r \frac{dr'}{\sqrt{1-kr'^2}} = a\int_{S_k^{-1}(0)}^{S_k^{-1}(r)} d\chi = aS_k^{-1}(r). \tag{1.27}$$

Therefore,

$$d(r,t) = a(t)\chi(r), \tag{1.28}$$

or, equivalently,

$$S_k(d/a) = r. \tag{1.29}$$

The previous expression allows us to compute the physical distance to an object given its comoving distance χ and the scale factor a. The radial coordinate $r(z)$ is usually referred to as the radial comoving distance or the comoving angular diameter distance, $d_M(z)$.

1.2.3 Observational Basis

The success of the ΛCDM model is due to its precise prediction and explanation of several physical processes of different nature, such as the existence and structure of the CMB, the abundances of light elements and the LSS of the Universe. In this section we summarize some of the most important observational evidences that strongly support the standard cosmological model.

1.2.3.1 The Cosmic Microwave Background

The cosmic microwave background (CMB) is a source of electromagnetic radiation that fills all space. It is a remnant from an early stage of the Universe, and is also known as "relic radiation". Precise measurements of the CMB are critical to

cosmology, since any model that aims to describe the origin and evolution of the Universe must explain this radiation. The CMB corresponds to a thermal black body spectrum at a temperature $T_0 = 2.72548 \pm 0.00057$ K [14].

During the radiation-dominated era that followed inflation, baryonic matter and radiation were tightly coupled through the interactions

$$p^+ + e^- \longleftrightarrow H + \gamma, \tag{1.30}$$

$$e^- + \gamma \longleftrightarrow e^- + \gamma. \tag{1.31}$$

As the Universe expanded, the temperature eventually decreased to the point in which electrons and protons formed hydrogen atoms without photons interfering in the process (recombination), which allowed photons to travel freely throughout the Universe (decoupling)[3]. The decoupling took place about 380,000 years after the Big Bang, corresponding to a redshift of about 1,100. As a result, the Universe became transparent, and we refer to this epoch as the last scattering surface.

The radiation of the CMB is isotropic to roughly one part in 100,000: the rms variations, after subtracting out the dipole anisotropy from the Doppler shift[4] of the background radiation, are only 18 μK [14]. The high degree of isotropy throughout the observable Universe and its small anisotropy lend strong support for the Big Bang theory and, in particular, for the ΛCDM model [15]. In Fig. 1.4 we show the map of the anisotropies of the CMB as obtained by the Planck satellite.

The anisotropy of the CMB is divided into two types: primary anisotropy, which is due to effects that occurred at the last scattering surface and before; and secondary anisotropy, due to interactions of the CMB photons with hot gas and gravitational potentials, which occurred between the last scattering surface and us. The structure of these anisotropies is determined by two effects: baryon acoustic oscillations and diffusion damping, or Silk damping. The acoustic oscillations give the CMB its characteristic peak structure. The peaks in the temperature power spectrum correspond, roughly, to resonances in which photons decouple when a particular mode is at its peak amplitude. In Fig. 1.5 we show the power spectrum of the temperature of the CMB as obtained by Planck [14], together with the best fit model.

1.2.3.2 Primordial Nucleosynthesis and Abundance of Light Elements

Nucleosynthesis is the process that creates new atomic nuclei from pre-existing nucleons (protons and neutrons) and nuclei [16], and is the only window into the conditions of the early Universe before the CMB. The first nuclei were formed a few minutes after the Big Bang, through nuclear reactions in a process called Big Bang nucleosynthesis (BBN). After $\sim$20 min, the Universe had expanded and cooled to a point at which these high energy collisions among nucleons ended, so only the fastest and

[3] Recombination and decoupling are two physical processes strongly linked to each other.

[4] This Doppler shift is due to the peculiar movement of the Earth with respect to comoving observers.

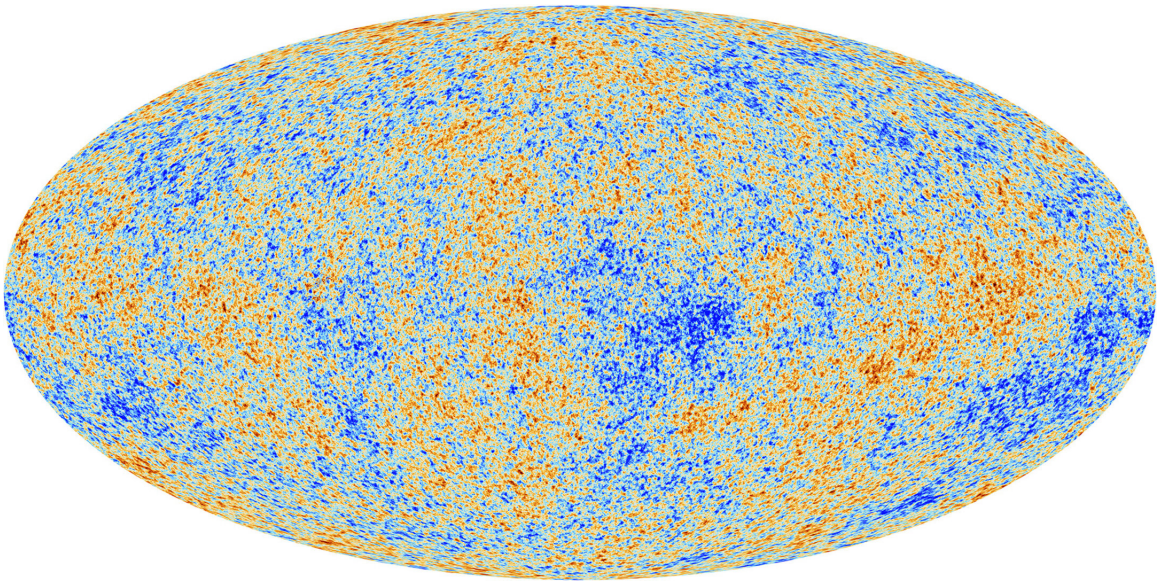

Figure 1.4: Map of the anisotropies of the CMB as measured by Planck. Image taken from https://www.esa.int/var/esa/storage/images/esa_multimedia/images/2018/07/planck_s_view_of_the_cosmic_microwave_background/17601794-1-eng-GB/Planck_s_view_of_the_cosmic_microwave_background_pillars.jpg.

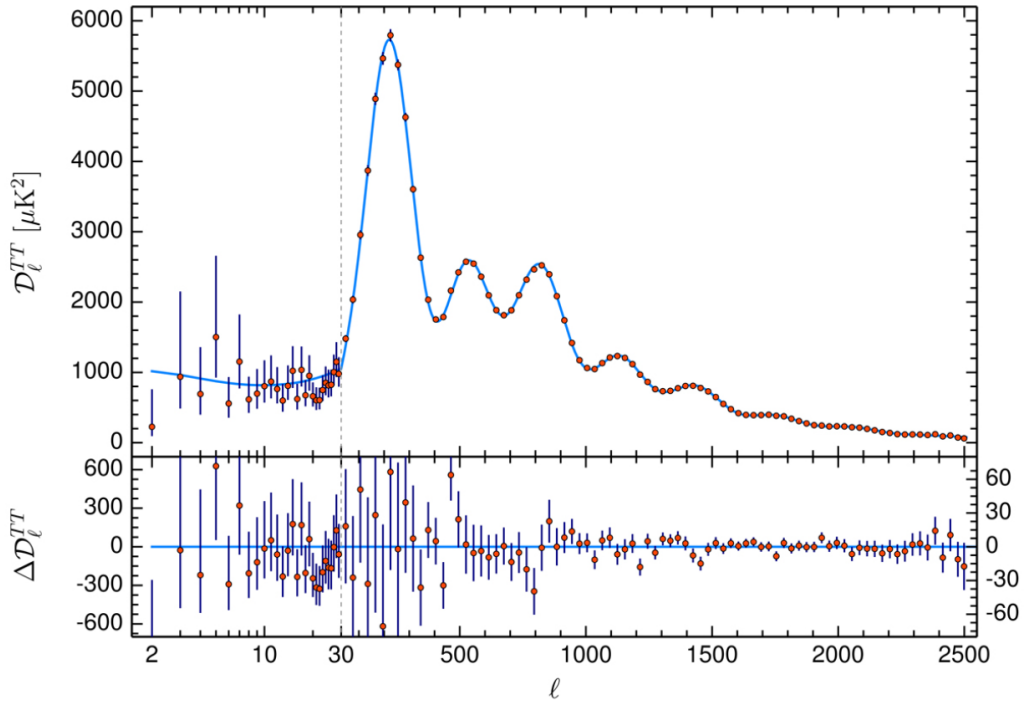

Figure 1.5: Power spectrum of the temperature of the CMB as measured by Planck [14], together with the best fit model.

simplest reactions occurred, leaving our Universe containing ^{1}H (∼75% of mass), ^{2}D, ^{3}He, ^{4}He (∼25% of mass) and ^{7}Li, i.e., no elements heavier than beryllium could be formed. In Fig. 1.6 we show the nuclear reactions that occurred during primordial nucleosynthesis (top), and also the standard Big Bang nucleosynthesis of the most abundant primordial nuclei (bottom).

1.2.3.3 The Large-Scale Structure of the Universe

The large-scale structure of the Universe refers to the patterns that galaxies and matter follow in their distribution on much larger scales than individual galaxies or clusters of galaxies [18]. These structures are created and shaped by the effect of gravity. Just as gravity on smaller scales binds gas particles together to form stars, and these stars join together to form galaxies, it also binds galaxies and matter in patterns on larger scales. These patterns often contain large strands of galaxies and voids in between, which resemble a spider web, so it is often referred to as "the cosmic web". In Chapter 2 we will go into further detail about the LSS of the Universe.

1.2.4 Parameters of the ΛCDM Model

The ΛCDM model can be completely characterized by six independent parameters: the physical baryon density parameter (Ω_b), the physical cold dark matter density parameter (Ω_c), the Hubble constant (H_0), the scalar spectral amplitude (A_s), the scalar spectral index (n_s) and the optical depth at reionization (τ).

Before we continue, it is useful to explicitly define the already-mentioned density parameters and clarify a subtlety regarding their time dependence. These density parameters are given by the ratio of the density of each fluid and the so-called critical density,

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_{\rm cr}(t)}, \tag{1.32}$$

with the critical density given by

$$\rho_{\rm cr}(t) = \frac{3H^2(t)}{4\pi G}. \tag{1.33}$$

$H(t)$ is known as the Hubble parameter, and its value at present is the Hubble constant, i.e., $H_0 \equiv H(t = t_0)$. In general, the density parameters depend on time; however, whenever we drop the time-dependence it means we are referring to their values at present, i.e., $\Omega_i \equiv \Omega_i(t = t_0)$, which are the free parameters of ΛCDM.

The matter content of the Universe can be given either in terms of Ω_b and Ω_c, or in terms of one of them and the sum of both,

$$\Omega_m = \Omega_b + \Omega_c. \tag{1.34}$$

$$
\begin{aligned}
&\mathrm{n}^0 \longrightarrow \mathrm{p}^+ + \mathrm{e}^- + \bar{\nu}_e && \mathrm{p}^+ + \mathrm{n}^0 \longrightarrow {}^2_1\mathrm{D} + \gamma \\
&{}^2_1\mathrm{D} + \mathrm{p}^+ \longrightarrow {}^3_2\mathrm{He} + \gamma && {}^2_1\mathrm{D} + {}^2_1\mathrm{D} \longrightarrow {}^3_2\mathrm{He} + \mathrm{n}^0 \\
&{}^2_1\mathrm{D} + {}^2_1\mathrm{D} \longrightarrow {}^3_1\mathrm{T} + \mathrm{p}^+ && {}^3_1\mathrm{T} + {}^2_1\mathrm{D} \longrightarrow {}^4_2\mathrm{He} + \mathrm{n}^0 \\
&{}^3_2\mathrm{He} + {}^4_2\mathrm{He} \longrightarrow {}^7_3\mathrm{Li} + \gamma && {}^3_2\mathrm{He} + \mathrm{n}^0 \longrightarrow {}^3_1\mathrm{T} + \mathrm{p}^+ \\
&{}^3_2\mathrm{He} + {}^2_1\mathrm{D} \longrightarrow {}^4_2\mathrm{He} + \mathrm{p}^+ && {}^3_2\mathrm{He} + {}^4_2\mathrm{He} \longrightarrow {}^7_4\mathrm{Be} + \gamma \\
&{}^7_3\mathrm{Li} + \mathrm{p}^+ \longrightarrow {}^4_2\mathrm{He} + {}^4_2\mathrm{He} && {}^7_4\mathrm{Be} + \mathrm{n}^0 \longrightarrow {}^7_3\mathrm{Li} + \mathrm{p}^+
\end{aligned}
$$

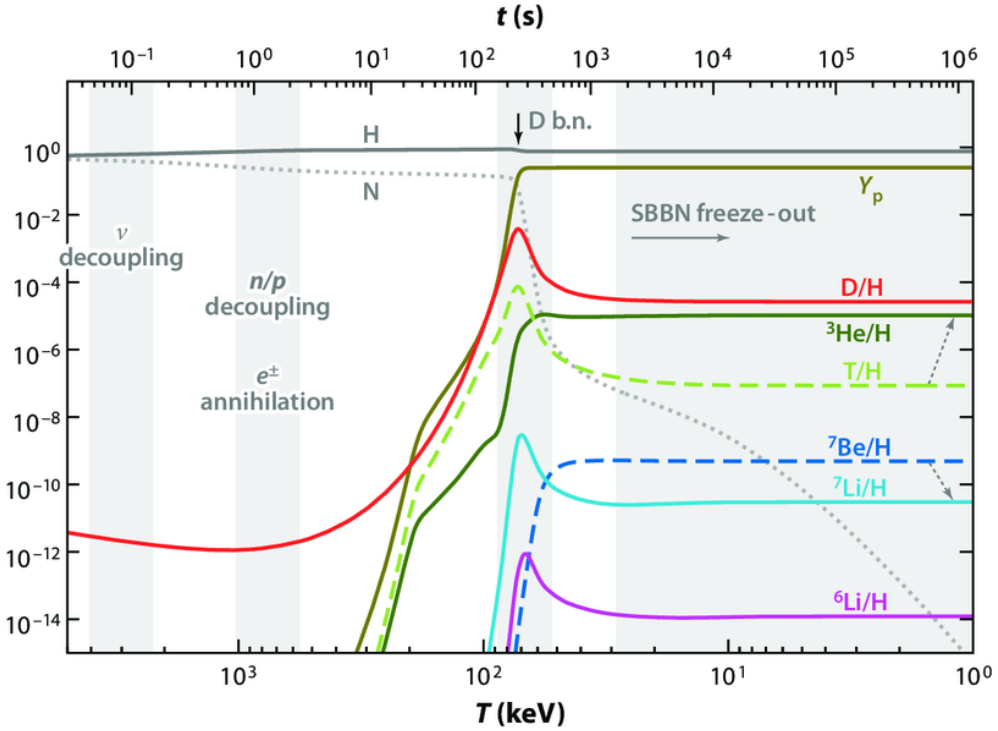

Figure 1.6: **Top:** Relevant nuclear reactions responsible for the relative abundances of light atomic nuclei observed in our Universe. **Bottom:** Standard Big Bang nucleosynthesis of the most abundant primordial nuclei, taken from [17].

Parameter	Planck-2018
Ω_b	0.0493 ± 0.0010
Ω_m	0.3153 ± 0.0073
H_0	(67.36 ± 0.54) km s^{-1} Mpc^{-1}
σ_8	0.8111 ± 0.0060
n_s	0.9649 ± 0.0042
τ	0.0544 ± 0.0073

Table 1.1: Values of some of the cosmological parameters obtained from the latest results from Planck, see [14].

On the other hand, the Hubble constant H_0 is related to the so-called acoustic scale $\theta_{\rm MC}$ through

$$\theta_{\rm MC} = \frac{r_s(z_{\rm d})}{d_M(z_{\rm d})}, \tag{1.35}$$

where $z_{\rm d}$ is the redshift corresponding to the decoupling epoch, $r_s(z)$ is the sound horizon scale and $d_M(z)$ is the comoving angular diameter distance. Using the Hubble constant and Ω_m, the age of the Universe, t_0, can also be computed. Therefore, in practice, we can use either H_0, $\theta_{\rm MC}$ or t_0 as our free parameter, and the other two as derived parameters. Something similar happens with A_s and σ_8[5], since any of them can be derived from the other. In Table 1.1 we display the parameters of the ΛCDM model as obtained by Planck [14].

The dark energy density parameter can then be computed as

$$\Omega_\Lambda = 1 - \Omega_m - \Omega_r. \tag{1.36}$$

The Ω_r appearing in the previous expression is the density parameter for radiation. Its value today is $\Omega_r \sim 10^{-4}$ and, because of this, is usually set to zero when computing quantities at present time, but it must be taken into account when studying the early Universe. Neutrinos are also an important ingredient of the ΛCDM model, and they can be either ultra-relativistic or non-relativistic, depending on their mass. Therefore, they can either contribute to the radiation or the matter energy density of the Universe:

1. Radiation. The photon density is given by

$$\rho_\gamma = \frac{\pi^2}{15}\frac{(k_B T_0)^4}{\hbar^3 c^5}, \tag{1.37}$$

[5] σ_8 is the present root-mean-square matter fluctuation averaged over a sphere of radius 8 Mpc/h. We will describe it in more detail in Chapter 2.

where $T_0 = 2.72548$ K [14]. Therefore, the density parameter of photons at present time is given by

$$\Omega_\gamma = \frac{\rho_\gamma}{\rho_{cr}^0} = \rho_\gamma \left(\frac{3H_0^2}{8\pi G}\right)^{-1}. \tag{1.38}$$

The total radiation density parameter includes the contribution of photons and massless neutrinos (which would be ultra-relativistic today),

$$\Omega_r = \Omega_\gamma + \Omega_\nu^{\text{massless}}. \tag{1.39}$$

The neutrino contribution is given by

$$\Omega_\nu^{\text{massless}} = \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\left(\frac{3 - N_\nu^{\text{massive}}}{3}\right) N_{\text{eff}}\Omega_\gamma, \tag{1.40}$$

where N_ν^{massive} is the number of massive neutrinos (integer number between 0 and 3), and $N_{\text{eff}} = 3.046$. Therefore,

$$\Omega_r = \left[1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\left(\frac{3 - N_\nu^{\text{massive}}}{3}\right) N_{\text{eff}}\right]\Omega_\gamma. \tag{1.41}$$

This is valid for studies of the CMB, when neutrinos were ultra-relativistic.

2. Matter. When the temperature of the Universe falls below the mass of the neutrinos, these particles start behaving like dark matter. Therefore, they no longer contribute to radiation, but rather to the matter energy density,

$$\Omega_m = \Omega_b + \Omega_c + \Omega_\nu^{\text{massive}}. \tag{1.42}$$

The number density for massive neutrinos is given by

$$n_\nu = \frac{3}{2}\frac{\zeta(3)}{\pi^2}\frac{(k_B T_0^\nu)^3}{\hbar^3 c^3}, \tag{1.43}$$

where T_0^ν is the temperature of the neutrino background, which can be written in terms of T_0 as

$$T_0^\nu = \left(\frac{4}{11}\right)^{1/3} T_0. \tag{1.44}$$

Therefore, the density parameter for massive neutrinos is

$$\Omega_\nu^{\text{massive}} = \sum_\nu m_\nu n_\nu \left(\frac{3H_0^2}{8\pi G}\right)^{-1}, \tag{1.45}$$

since $\rho = mn$ for non-relativistic particles. We easily find that

$$\Omega_\nu^{\text{massive}} = \frac{\sum_\nu m_\nu c^2}{94.07h^2 \text{ eV}}. \tag{1.46}$$

It should also be taken into account that each species of neutrino could have a different mass, which would make them become non-relativistic at different epochs.

The sum of the three neutrino masses, $\sum_\nu m_\nu c^2$, is usually set to 0.06 eV (which is the lower limit obtained by neutrino oscillations experiments [19]), and N_ν^{massive} is usually set to 3. There are extensions to ΛCDM, such as the oΛCDM and wCDM models, which are the simplest and most commonly used ones:

- oΛCDM. This model includes Ω_k as an independent parameter,

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1. \tag{1.47}$$

  Unless explicitly stated, we assume $\Omega_k \neq 0$ throughout this chapter, which allows us to obtain equations that depend on Ω_k without loss of generality. However, strictly speaking, the ΛCDM model has $\Omega_k = 0$: observations tell us that our Universe is flat, i.e., $k = 0$, see [14] for the latest results by the Planck mission.

- wCDM model. This model modifies the dark energy equation of state as

$$p = c^2 \left[w_0 + (1-a)w_a\right] \rho. \tag{1.48}$$

  Setting $w_0 = -1$ and $w_a = 0$, we recover the standard ΛCDM model. We will introduce the equation of state and the equation-of-state parameters, w_i, in section 1.4.2.

1.3 Cosmological Distances and Horizons

Measuring distances is one of the most important duties of cosmologists. There are several definitions of distances used in astronomy, and here we describe the most commonly used ones. It is particularly useful to define the so-called conformal time,

$$d\eta = \frac{dt}{a(t)}, \tag{1.49}$$

which allows us to express the FLRW metric, given by eq. (1.21), in a more convenient way,

$$ds^2 = a(\eta)^2 \left[-c^2 d\eta^2 + d\chi^2 + S_k(\chi)^2 \left(d\theta^2 + \sin^2\theta d\varphi^2\right)\right]. \tag{1.50}$$

1.3.1 Hubble's Law and Redshift

Hubble's law was discovered by Edwin Hubble in 1929 during his famous study about the correlation between the distance and radial velocity among extra-galactic nebulae [2]. Let $d(r,t)$ be the physical distance to a galaxy that is at a radial distance r at a certain instant t. As we showed in the previous section, eq. (1.28), it can be written as

$$d(r,t) = a(t)\chi(r). \tag{1.51}$$

Hubble's law can be obtained deriving $d(r,t)$ with respect to time,

$$v(r,t) \equiv \dot{d}(r,t) = H(t)d(r,t), \tag{1.52}$$

where $H(t)$ is the Hubble parameter, which is defined as the derivative of the scale factor divided by the scale factor,

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \tag{1.53}$$

Eq. (1.52) is usually referred to as Hubble's law.

1.3.1.1 The Concept of Redshift

By definition, the redshift of an object is the fractional Doppler shift of its emitted light resulting from radial motion,

$$z \equiv \frac{\nu_{\rm em}}{\nu_{\rm obs}} - 1 \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} - 1. \tag{1.54}$$

The concept of redshift is related to the scale factor and the expansion of the Universe. For a light ray that travels from a source distant galaxy to us, we know that $ds^2 = 0$ (since $ds^2 = 0$ for photons). For a fixed direction, $d\theta = d\varphi = 0$. Therefore, since $c^2 d\eta^2 = d\chi^2$,

$$cd\eta = -d\chi. \tag{1.55}$$

Let's imagine our source emits a photon at $\eta_{\rm em}$ and that we receive it at $\eta_{\rm obs}$. Integrating the previous expression, we find that

$$c(\eta_{\rm obs} - \eta_{\rm em}) = \chi_{\rm em}, \tag{1.56}$$

since $\chi_{\rm obs} = 0$. Let's imagine now that the same source emits another photon at $\eta_{\rm em} + \Delta\eta_{\rm em}$ and that we receive it at $\eta_{\rm obs} + \Delta\eta_{\rm obs}$. If the source does not change its position,

$$c(\eta_{\rm obs} + \Delta\eta_{\rm obs}) - c(\eta_{\rm em} + \Delta\eta_{\rm em}) = \chi_{\rm em}. \tag{1.57}$$

Comparing the two previous equations we find that

$$\Delta\eta_{\rm em} = \Delta\eta_{\rm obs}. \tag{1.58}$$

The two time intervals are exactly the same when measured in conformal time. Transforming the intervals from conformal to cosmic times,

$$\Delta t_{\rm em} = a(t_{\rm em})\Delta\eta_{\rm em}, \quad \Delta t_{\rm obs} = a(t_{\rm obs})\Delta\eta_{\rm obs}, \tag{1.59}$$

and computing the ratio, we find that

$$\frac{\Delta t_{\rm obs}}{\Delta t_{\rm em}} = \frac{\nu_{\rm em}}{\nu_{\rm obs}} = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = \frac{a(t_{\rm obs})}{a(t_{\rm em})}. \tag{1.60}$$

Comparing the previous expression with the definition of redshift, eq. (1.54), we find that the cosmological redshift is given in terms of the scale factor,

$$1 + z = \frac{a(t_{\rm obs})}{a(t_{\rm em})}. \tag{1.61}$$

At short distances, z corresponds to the radial velocity divided by the speed of light. This is easy to demonstrate, since

$$\begin{aligned} z = \frac{a(t_{\rm obs})}{a(t_{\rm em})} - 1 \approx \frac{a(t_{\rm obs})}{a(t_{\rm obs}) + \dot{a}(t_{\rm obs})(t_{\rm em} - t_{\rm obs})} - 1 \\ = \frac{1}{1 + H_{\rm obs}(t_{\rm em} - t_{\rm obs})} - 1 \approx -H_{\rm obs}(t_{\rm em} - t_{\rm obs}) = H_{\rm obs}(t_{\rm obs} - t_{\rm em}). \end{aligned} \tag{1.62}$$

On the other hand,

$$\begin{aligned} c(t_{\rm obs} - t_{\rm em}) = c\left(\frac{dt}{d\eta}\right)_{t_{\rm obs}} (\eta_{\rm obs} - \eta_{\rm em}) = a(t_{\rm obs}) c(\eta_{\rm obs} - \eta_{\rm em}) \\ = a(t_{\rm obs})\chi(r) = d(r, t_{\rm obs}). \end{aligned} \tag{1.63}$$

Therefore,

$$z = H_{\rm obs}\frac{d(r, t_{\rm obs})}{c} = \frac{v(r, t_{\rm obs})}{c}. \tag{1.64}$$

The cosmological redshift is not due to a Doppler effect, but to the growth of the scale factor. However, for nearby galaxies it can be mistaken with a Doppler effect.

In general, velocities are difficult to measure, so at short distances they are replaced by redshift. Distances are also not easy to measure, but there are two types of distances that are particularly useful in cosmology: luminosity distance and angular diameter distance.

1.3.2 Distances

1.3.2.1 Radial Comoving Distance (Line of Sight), χ

A small comoving distance $d\chi$ between two nearby objects in the Universe is the distance between them which remains constant in time if the two objects move with the Hubble flow, i.e., without considering their peculiar velocities. In other words, it is the distance between them which would be measured with rulers at the time they are being observed (the proper distance) divided by the ratio of the scale factor of the Universe then to now. The total line-of-sight (hereafter, LOS) comoving distance $\chi(z)$ from us to a distant object is computed by integrating the infinitesimal $d\chi$ contributions between nearby events along the radial ray from $z = 0$ to the object,

$$\chi(z) = \int_0^z dz \frac{d\chi}{dz} = -c\int_0^z dz \frac{d\eta}{dz} = -c\int_0^z \frac{dz}{a(z)}\frac{dt}{dz}. \tag{1.65}$$

Now, since

$$a(z) = \frac{1}{1+z} \rightarrow \frac{da}{dt} = -a^2 \frac{dz}{dt} \rightarrow dt = -a\frac{dz}{H}, \tag{1.66}$$

we easily find that

$$\chi(z) = c\int_0^z \frac{dz'}{H(z')}. \tag{1.67}$$

1.3.2.2 Transverse Comoving Distance, d_M

The comoving distance between two events at the same redshift or distance but separated on the sky by some angle $\delta\theta$ is $d_M(z)\delta\theta$, being the transverse comoving distance $d_M(z)$ simply related to the LOS comoving $\chi(z)$ as

$$d_M(z) = S_k(\chi). \tag{1.68}$$

Both the radial and the transverse distances are defined as comoving since they have factored out the scale factor, $a(t)$. Therefore, in the absence of peculiar velocities, they remain constant.

1.3.2.3 Luminosity Distance, d_L

The luminosity distance $d_L(z)$ is defined in terms of the relationship between the absolute magnitude M and the apparent magnitude m of an astronomical object. The absolute magnitude is, in turn, related to the absolute luminosity L; whereas the apparent magnitude is related to the apparent luminosity l (also known as observed flux, $F_{\rm obs}$). In an Euclidean space, L and l are related as

$$l = \frac{L}{4\pi d_L^2}, \tag{1.69}$$

where d_L is the luminosity distance to the source. Our objective is to obtain $d_L = d_L(z)$. Let S be the area of the collecting surface,

$$S = 4\pi a_0^2 d_M^2. \tag{1.70}$$

The energy flux of the observed photons is lower than the one they had at the instant of emission by a factor $(1+z)^2$. The first $(1+z)$ factor comes from the expansion of the Universe, whereas the second one comes from the fact that if the source emits a photon every $\Delta t_{\rm em}$, then $\Delta t_{\rm obs} = (1+z)\Delta t_{\rm em}$. Therefore,

$$l = \frac{L}{S}\frac{1}{(1+z)^2} = \frac{L}{4\pi a_0^2 d_M^2}\frac{1}{(1+z)^2}, \tag{1.71}$$

which yields

$$d_L(z) = a_0 d_M(z)(1+z). \tag{1.72}$$

An exact relation between d_L and z can be given in terms of an integral, since

$$c\left[\eta(t_0)-\eta(t_1)\right] = -\int_{t_1}^{t_0} d\chi = \chi(r) = S_k^{-1}(r) \tag{1.73}$$

and, therefore,

$$d_M(z) = S_k\left(c\int_{t_1}^{t_0}\frac{dt}{a(t)}\right). \tag{1.74}$$

Writing the previous expression as an integral in z, we find that

$$d_M(z) = S_k\left(-c\int_z^0 dz'\frac{1}{a_0}\frac{1}{H(z')}\right) = S_k\left(\frac{c}{a_0}\int_0^z dz'\frac{1}{H(z')}\right). \tag{1.75}$$

The luminosity distance is, then, given by

$$d_L(z) = a_0 d_M(z)(1+z) = a_0 S_k\left(\frac{c}{a_0}\int_0^z\frac{dz}{H(z)}\right)(1+z). \tag{1.76}$$

Up to second order in redshift,

$$d_L(z) = cH_0^{-1}\left(z+\frac{1-q_0}{2}z^2+\dots\right), \tag{1.77}$$

where q_0 is the deceleration parameter at present time, with $q(t)$ defined as

$$q(t) = -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)}. \tag{1.78}$$

1.3.2.4 Distance Modulus, $m - M$

The distance modulus describes distances on a logarithmic scale based on the astronomical magnitude system. The history of the astronomical magnitude system starts with the Greek astronomer Hipparchus, who produced a catalogue which noted the apparent brightness of stars in the second century BCE. Claudius Ptolemy, based on Hipparchus' work, classified stars on a six point scale (by eye), and originated the term magnitude. Finally, it was Norman Pogson who proposed, in 1856, that a logarithmic scale of $(100)^{1/5} \approx 2.512$ should be adopted between magnitudes, so that five magnitude steps would correspond, precisely, to a factor of 100 in brightness, i.e.,

$$\frac{l(1)}{l(6)} = 100, \tag{1.79}$$

where l is the brightness or apparent luminosity.

The apparent magnitude is defined as

$$m = -\frac{5}{2}\log_{10}\left(\frac{l}{l_0}\right), \quad l_0 = 2.52\times10^{-8}\mathrm{Wm}^{-2}, \tag{1.80}$$

so that

$$m(6) - m(1) = -\frac{5}{2}\log_{10}\left[\frac{l(6)}{l(1)}\right] = -\frac{5}{2}\log_{10}(10^{-2}) = 5\log_{10}10 = 5. \tag{1.81}$$

On the other hand, the absolute magnitude is defined as the apparent magnitude that a star would have if it was placed at 10 pc in a static Euclidean space,

$$M = -\frac{5}{2}\log_{10}\left[\frac{L/4\pi(10\text{ pc})^2}{l_0}\right], \tag{1.82}$$

where L is the absolute luminosity. Defining $L_0 = 4\pi(10\text{ pc})^2 l_0$,

$$M = -\frac{5}{2}\log_{10}\left(\frac{L}{L_0}\right). \tag{1.83}$$

Objects with fixed absolute magnitude (or, equivalently, fixed absolute luminosity) are known as "standard candles".

We are now in a position to define the distance modulus. Let l and L be the apparent and absolute luminosities of an astronomical object, respectively. We have already defined the apparent and absolute magnitudes as

$$\begin{cases} l = l_0 10^{-\frac{2}{5}m}, \\ L = L_0 10^{-\frac{2}{5}M}. \end{cases} \tag{1.84}$$

The luminosity distance can, then, be expressed in terms of these magnitudes as

$$d_L = \sqrt{\frac{L}{4\pi l}} = 10^{(m-M)/5+1}\text{ pc}. \tag{1.85}$$

The distances modulus, $m - M$, is given by

$$m - M = 5\log_{10}\left(\frac{d_L}{10\text{ pc}}\right). \tag{1.86}$$

Therefore,

$$m - M = 5\log_{10}\left(\frac{d_L}{\text{Mpc}}\right) + 25. \tag{1.87}$$

Up to second order in redshift,

$$m - M = 5\log_{10}\left[\frac{cH_0^{-1}\left(z + \frac{1-q_0}{2}z^2 + \dots\right)}{\text{Mpc}}\right] + 25. \tag{1.88}$$

1.3.2.5 Angular Diameter Distance, d_A

The angular diameter distance $d_A(z)$ is defined as the ratio of an object's physical transverse size, $d_M(z)\delta\theta/(1+z)$, to its angular size, $\delta\theta$. It is used to convert angular separations in telescope images, and is given by

$$d_A(z) = \frac{d_M(z)}{1+z}. \tag{1.89}$$

The angular diameter distance of an object is a good approximation to the real distance, i.e., the proper distance when the light left the object. Note that beyond a certain redshift z_t, the angular diameter distance gets smaller with increasing redshift, i.e.,

$$z_2 > z_1 > z_t \rightarrow d_A(z_2) < d_A(z_1). \tag{1.90}$$

In other words, an object "behind" another of the same size, beyond z_t, appears larger on the sky,

$$\frac{d_A(z_1)}{d_A(z_2)} = \frac{x/\delta\theta_1}{x/\delta\theta_2} = \frac{\delta\theta_2}{\delta\theta_1} > 1 \rightarrow \delta\theta_2 > \delta\theta_1. \tag{1.91}$$

It is related to the luminosity distance as

$$d_A = \frac{d_L}{(1+z)^2}. \tag{1.92}$$

Up to second order in redshift,

$$d_A = cH_0^{-1}\left(z - \frac{3+q_0}{2}z^2 + \dots\right). \tag{1.93}$$

1.3.3 Horizons

The most commonly studied horizons in cosmology are the Hubble and the particle horizons. Here we describe both of them.

1.3.3.1 Hubble Horizon, d_H

The Hubble horizon is a conceptual horizon defining the boundary between particles that are moving slower and faster than the speed of light (relative to an observer at a given time). The Hubble velocity of an object is given by Hubble's law, see eq. (1.52). Replacing v with the speed of light, we can solve for the radius of the Hubble horizon,

$$d_H(t) = cH^{-1}(t). \tag{1.94}$$

The derivative of the previous expression with respect to time gives us the expansion velocity of the Hubble horizon,

$$\frac{1}{c}\dot{d}_H(t) = \frac{1}{c}\frac{d}{dt}d_H(t) = 1 - \frac{\ddot{a}(t)a(t)}{\dot{a}(t)^2}. \tag{1.95}$$

The acceleration of the expansion, $\ddot{a}(t)$, determines the sign of the right hand side of the previous equation. If the Universe is decelerating, i.e., if $\ddot{a}(t) < 0$, then $\dot{d}_H(t) > c$, which means that the Hubble horizon is expanding faster than the speed of light, and galaxies further away than $d_H(t)$ will enter the horizon at some point. On the other hand, if the expansion is accelerated, then the Hubble horizon expands slower than the recession speed of galaxies outside of it.

If we consider physical sizes or scales of structures, the same reasoning applies. During inflation, the expansion is accelerated, and therefore all physical scales are pushed out of the Hubble horizon. When inflation ends and the radiation-dominated and matter-dominated eras successively begin, the expansion is decelerated, which allows larger physical scales to enter the horizon and form structures.

1.3.3.2 Particle Horizon, d_P

The particle horizon is the maximum distance from which light from particles could have traveled to the observer in the age of the Universe. Much like the concept of a terrestrial horizon, it represents the boundary between the observable and the non observable regions of the Universe, so its distance at the present epoch defines the size of the observable universe. Due to the expansion of the Universe, it is not simply the age of the universe times the speed of light, but rather the speed of light times the conformal time. The particle horizon is given by

$$d_P(t) = a(t)c\eta(t) = a(t)\int_0^t \frac{cdt'}{a(t')}. \tag{1.96}$$

1.4 Cosmological Dynamics

The objective in this section is to write down and solve the equations governing the cosmological dynamics. These equations are deduced from the Einstein Field Equations for the FLRW metric, and are known as the Friedmann equations.

1.4.1 Friedmann Equations and Conservation Law

As we saw in section 1.2.2, the FLRW metric, expressed in Cartesian coordinates, is given by

$$ds^2 = -c^2dt^2 + a(t)^2\gamma_{ij}dx^idx^j, \tag{1.97}$$

where

$$\gamma_{ij} = \delta_{ij} + k\frac{x_ix_j}{1-kr^2}. \tag{1.98}$$

From the metric tensor we compute the Ricci scalar, which has the expression

$$R = \frac{6}{c^2a^2}(\ddot{a}a + \dot{a}^2 + c^2k). \tag{1.99}$$

The Einstein tensor is, then, given by

$$G_{\mu\nu}\begin{cases} G_{00} = \dfrac{3}{c^2a^2}(\dot{a}^2 + c^2k), \\ G_{0i} = 0, \\ G_{ij} = -\dfrac{1}{c^2a^2}(2\ddot{a}a + \dot{a}^2 + c^2k)g_{ij}. \end{cases} \tag{1.100}$$

It is also necessary to deduce the stress-energy tensor, $T_{\mu\nu}$. Consider a multi-fluid Universe with N non-interacting components, being each of them labeled with a subscript i, i.e., $T_{\mu\nu} = \sum_i T^i_{\mu\nu}$. From the cosmological principle it follows that

$$T_{\mu\nu}\begin{cases} T_{00} = c^2\rho(t), \\ T_{0i} = 0, \\ T_{ij} = p(t)g_{ij}, \end{cases} \tag{1.101}$$

where $\rho(t) = \sum_i \rho_i(t)$ and $p(t) = \sum_i p_i(t)$. For a comoving fluid we know that $u^\mu = c\delta^\mu_0$. Thus, the stress-energy tensor of the Universe can be written covariantly as

$$T_{\mu\nu} = \left(\rho + \frac{p}{c^2}\right) u_\mu u_\nu + p g_{\mu\nu}. \tag{1.102}$$

This is the stress-energy tensor of a relativistic perfect fluid.

We have already obtained the Einstein tensor and the stress-energy tensor. From the Einstein field equations, eq. (1.7), we obtain a total of two independent equations:

- Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2k}{a^2} - \frac{c^2\Lambda}{3} = \frac{8\pi G}{3}\rho. \tag{1.103}$$

- Acceleration equation:

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2k}{a^2} - c^2\Lambda = -\frac{8\pi G}{c^2}p. \tag{1.104}$$

Combining the acceleration equation with the Friedmann equation, we find

$$\ddot{a} = -\frac{4\pi G}{3}\left(\rho + 3\frac{p}{c^2}\right)a + \frac{c^2\Lambda}{3}a. \tag{1.105}$$

So far we have two equations and $1+2N$ functions to determine (a, ρ_i and p_i). We can obtain another set of N equations by making 0 the 4-divergence of the stress-energy tensor, i.e., $\nabla_\mu T^{\mu\nu}_i = 0$,

$$\dot{\rho}_i + 3\frac{\dot{a}}{a}\left(\rho_i + \frac{p_i}{c^2}\right) = 0, \quad i = 1, \ldots, N. \tag{1.106}$$

These equations are known as the continuity equations. Summing over all the different fluid components, we find that

$$\dot{\rho} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0, \tag{1.107}$$

which can also be deduced from the Friedmann equation when combined with the acceleration one. This means that now we have $1 + N$ independent equations: the Friedmann equation and the N continuity equations. Therefore, the system to be solved is

$$\begin{cases} \left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2 k}{a^2} - \frac{c^2\Lambda}{3} = \frac{8\pi G}{3}\rho, \\ \dot{\rho}_i + 3\frac{\dot{a}}{a}\left(\rho_i + \frac{p_i}{c^2}\right) = 0, \quad i = 1, \ldots, N. \end{cases} \tag{1.108}$$

These are a total of $1 + N$ equations for $1 + 2N$ variables (a, ρ_i, p_i). We need another N independent equations for us to be able to solve the system.

1.4.2 Equation of State

We need N more independent equations to solve our system of differential equations, eq. (1.108). The equation that relates the density ρ of a species i with its pressure p is the so-called equation of state. Adding an equation of state $p = p(\rho)$ for each fluid component would allow us to solve our system of equations. In cosmology, the equations of state are usually chosen to be of the form

$$p_i = c^2 w_i \rho_i. \tag{1.109}$$

The Universe is a multi-fluid, i.e., there are several fluids with different equations of state coexisting. The three most important types of fluids are:

- Dust (cold matter). Matter is considered as dust, and therefore it is pressureless, $p_m = 0$. This automatically implies $w_m = 0$.
- Radiation. The equation of state of radiation is given by $p_r = c^2\rho_r/3$, which means $w_r = 1/3$. This is the equation of state of a photon gas or, in general, of ultra-relativistic particles.
- Cosmological constant. This third component is deduced from the contribution of the vacuum to the stress-energy tensor, i.e., the $T^{\Lambda}_{\mu\nu}$ that we already defined in eq. (1.9). At this point we notice that if we set $p = -c^2\rho$ in the expression for the stress-energy tensor of a relativistic perfect fluid, eq. (1.102), we find that $T_{\mu\nu} = p g_{\mu\nu}$, which has an analogous expression to $T^{\Lambda}_{\mu\nu}$. From this it follows that the density of the cosmological constant is given by

  $$\rho_\Lambda = \frac{c^2\Lambda}{8\pi G}. \tag{1.110}$$

  Therefore, $p_\Lambda = -c^2\rho_\Lambda$, which means $w_\Lambda = -1$.

1.4.3 Solving the Cosmological Equations

The equations that describe the cosmological dynamics are the Friedmann and the continuity equations, eq. (1.108), together with the equations of state of the different fluids, eq. (1.109). Putting all of them together we obtain

$$\begin{cases} \left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2 k}{a^2} - \frac{c^2 \Lambda}{3} = \frac{8\pi G}{3}\rho, \\ \dot{\rho}_i + 3\frac{\dot{a}}{a}\left(\rho_i + \frac{p_i}{c^2}\right) = 0, \quad i = m, r, \\ p_i = c^2 w_i \rho_i, \quad i = m, r. \end{cases} \tag{1.111}$$

Assuming $w_\Lambda = -1$ and defining the Λ component of the density as in eq. (1.110), we can compactly write the equations as

$$\begin{cases} \left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2 k}{a^2} = \frac{8\pi G}{3}\rho, \\ \dot{\rho}_i + 3\frac{\dot{a}}{a}\left(\rho_i + \frac{p_i}{c^2}\right) = 0, \quad i = m, r, \Lambda \\ p_i = c^2 w_i \rho_i, \quad i = m, r, \Lambda. \end{cases} \tag{1.112}$$

Our objective in this section is to try to solve this system of equations analytically. Introducing the equation of state into the continuity equation, we find that

$$\rho_i = C a^{-3(1+w_i)}. \tag{1.113}$$

We can write this result in a more convenient way using the relation between the scale factor and the redshift,

$$\rho_i(z) = \rho_{i,0}(1+z)^{3(1+w_i)}. \tag{1.114}$$

Replacing w_i with its value for each of the fluid components, we obtain

- Dust (cold matter): $\rho = \rho_0(1+z)^3$.
- Radiation: $\rho = \rho_0(1+z)^4$. This extra $(1+z)$ with respect to dust comes from the redshift. For a black body

  $$n \propto T^3, \quad \rho \propto T^4. \tag{1.115}$$

  Since $\rho \propto (1+z)^4$ for radiation,

  $$T = T_0(1+z). \tag{1.116}$$

- Cosmological constant: $\rho = \rho_0 = c^2\Lambda/8\pi G$.

We find that, at short times ($t \to 0,\ z \to \infty$), radiation dominates, even though it is the least important contribution at present.

1.4.4 Friedmann Equation as a Function of the Density Parameters

We already defined the density parameters in section 1.2.4 as

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_{\rm cr}(t)} = \frac{8\pi G}{3H^2(t)}\rho_i(t), \tag{1.117}$$

and we also know that the Friedmann equation is given by

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{c^2 k}{a^2} = \frac{8\pi G}{3}\rho. \tag{1.118}$$

Introducing the definition of the density parameters into the Friedmann equation, we have

$$H^2 + \frac{c^2 k}{a^2} = H^2 \sum_i \Omega_i(t). \tag{1.119}$$

Defining the curvature density parameter as

$$H^2 \Omega_k(t) = -\frac{c^2 k}{a^2}, \tag{1.120}$$

we find that

$$\sum_i \Omega_i(t) + \Omega_k(t) = 1. \tag{1.121}$$

Separating the density in the different components, $\rho = \rho_m + \rho_r + \rho_\Lambda$,

$$\Omega_m(t) + \Omega_r(t) + \Omega_\Lambda(t) + \Omega_k(t) = 1. \tag{1.122}$$

It is easy to check that if $k < 0$ (hyperbolic spaces), then $\Omega_k > 0$ and $\sum_i \Omega_i < 1$. Also, if $k > 0$ (spherically curved spaces), then $\Omega_k < 0$ and $\sum_i \Omega_i > 1$. For the case of flat universes $k = 0$, which implies $\Omega_k = 0$ and $\sum_i \Omega_i = 1$.

From the definition of Ω_k, eq. (1.120), we can find the corresponding equation of state parameter w for the curvature. Since

$$\Omega_k \equiv \frac{\rho_k}{\rho_{\rm cr}} = -\frac{c^2 k}{a^2 H^2}, \tag{1.123}$$

using the definition of the critical density, eq. (1.33), we find that

$$\rho_k \propto a^{-2}. \tag{1.124}$$

Comparing the previous expression with eq. (1.113), we conclude that

$$w_k = -1/3. \tag{1.125}$$

The curvature effectively behaves like a fluid with equation of state

$$p_k = -\frac{c^2}{3}\rho_k. \tag{1.126}$$

It is also particularly useful to express $\rho_i(z) = \rho_{i,0}(1+z)^{3(1+w_i)}$ in terms of these density parameters. We easily find

$$\Omega_i(z) = \frac{H_0^2}{H^2}\Omega_i(1+z)^{3(1+w_i)}. \tag{1.127}$$

Using the Friedmann equation written in terms of the density parameters, $\sum_i \Omega_i(z) = 1$, we can also deduce that

$$H^2 = H_0^2 \sum_i \Omega_i(1+z)^{3(1+w_i)}. \tag{1.128}$$

The two previous equations completely determine the dynamics of the Universe. In Fig. 1.7 we show the evolution of the density parameters of matter, baryons, cold dark matter, dark energy, radiation, curvature, and also the sum of all of them as a function of time, from the early Universe to the present, and also the near future (top panel). We also show the evolution of the scale factor with time (bottom panel).

1.5 Cosmological Inflation

The inflationary epoch is introduced to solve some of the problems[6] of the standard cosmological model. In this section we explain how inflation solves the flatness and the horizon problems, and also introduce several different theoretical aspects of inflation.

1.5.1 Problems with the Standard Cosmological Model

There are two main problems in the standard cosmological model, which are the so-called flatness and horizon problems. Here we briefly discuss both of them, and show how an initial inflationary phase in the evolution of the Universe is able to solve them.

1.5.1.1 Flatness

The latest results by the Planck mission reinforce the idea that our Universe is flat. Explicitly, $\Omega_k = 0.0007 \pm 0.0037$ [14], which, in turn, means that $\Omega \equiv \Omega_m + \Omega_r + \Omega_\Lambda \approx 1$. At short times the curvature does not play any role in the Friedmann equation,

[6] These are not exactly problems with the ΛCDM model, but rather something that we don't know how to explain. Both of them are fine-tuning problems, and we essentially don't have an explanation to the values that these fine-tuned parameters take, but ΛCDM still works.

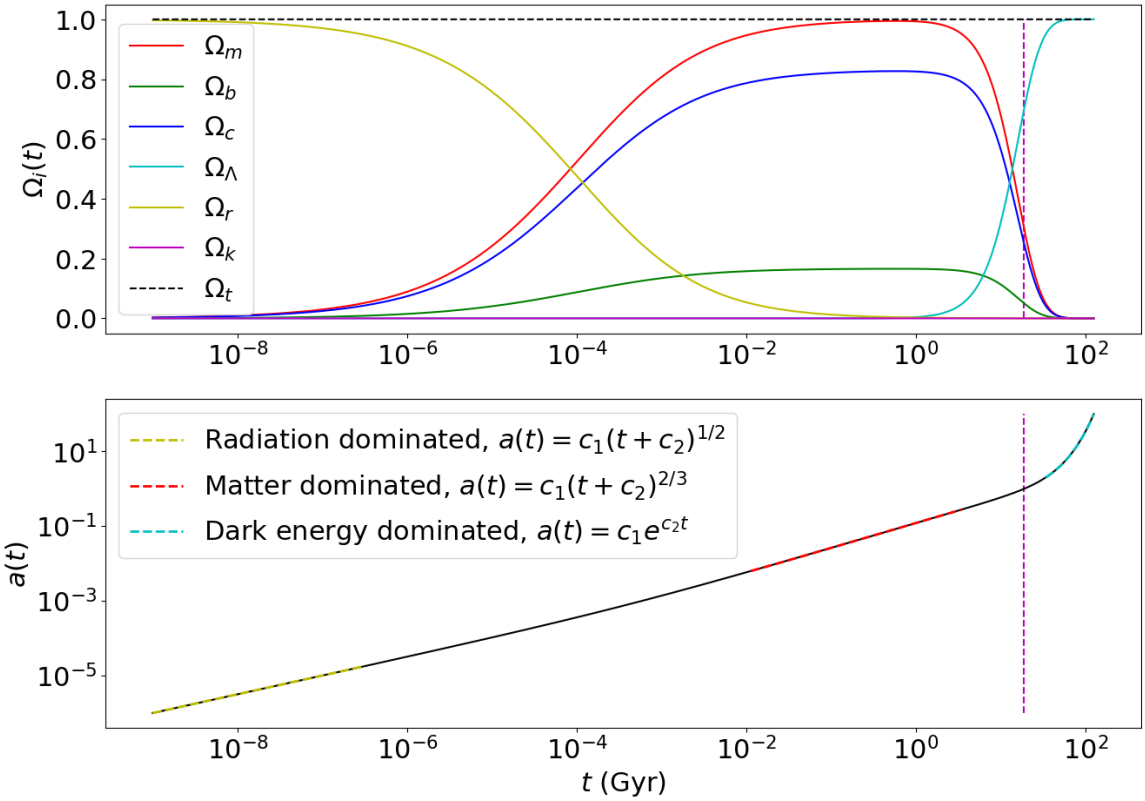

Figure 1.7: **Top panel:** Evolution of the density parameters for the different components that make up the Universe. Red for matter, green for baryonic matter, blue for CDM, cyan for dark energy, yellow for radiation and magenta for curvature. Radiation dominated during the first stages of the Universe, until the epoch of radiation-matter equality (where the yellow and red curves intersect). Then, matter starts dominating, and it's the most important contribution to the energy content until the Universe is around 7 billion years old, when the dark-energy-dominated era begins (together with the accelerated expansion of the Universe). **Bottom panel:** Evolution of the scale factor as a function of time. The different time-dependent expressions shown on the legend were obtained by analytically solving eq. (1.128) for the dominant component whenever the sum of the remaining ones represented less than 10% of the total energy density of the Universe. Results obtained using $\Omega_m = 0.3$ and $\Omega_b = 0.05$.

which could yield to the conclusion that our Universe is still in an early state. This is the reason why the flatness problem is also known as the oldness problem.

Because of how close Ω is to one at present time, the current energy density of the Universe is, by definition, very close to the critical value. Since any deviation of the total energy density from the critical value would increase rapidly over cosmic time, the early Universe must have had a density even closer to the critical density, departing from it by one part in e^N, with $N \sim 62$. N is the so-called number of e-folds the Universe had to expand in order to solve the flatness problem. Assuming a de-Sitter expansion phase, i.e., $a(t) = a(t_i)e^{H_I(t-t_i)}$ with constant $H(t) = H_I$, we find that the number of e-folds can be derived as

$$a(t_f) = a(t_i)e^{H_I(t_f-t_i)} = a(t_i)e^N. \tag{1.129}$$

We can now write the curvature density parameter in terms of N. The complete Friedmann equation is given by

$$\Omega(t) + \Omega_k(t) = 1, \tag{1.130}$$

where, by definition,

$$\Omega_k(t) = -\frac{c^2k}{H^2a^2}. \tag{1.131}$$

Therefore,

$$\Omega_k(t_f) = 1 - \Omega(t_f) = -\frac{c^2k}{H_I^2a_f^2} = -\frac{c^2k}{H_I^2a_i^2}e^{-2N} \propto -e^{-2N}. \tag{1.132}$$

The curvature parameter decays exponentially, so even if it starts with a large value, it is rapidly shrunk. This is how inflation explains how Ω_k had such a small value at early times.

1.5.1.2 Horizon

The horizon problem arises from the difficulty in explaining the observed homogeneity of casually disconnected regions of space in the absence of a mechanism that sets exactly the same initial conditions in the whole Universe.

Assuming a flat cosmology, the comoving angular diameter distance can be computed as

$$d_M(z) = S_{k=0}(\chi) = c\int_0^z \frac{dz'}{H(z')}. \tag{1.133}$$

The epoch of recombination, $z_{\rm rec} \approx 1100$, occurred during the matter-dominated era of the Universe, in which $H^2(z) \approx \Omega_m H_0^2(1+z)^3$. The comoving angular diameter distance, eq. (1.68), can then be written as

$$d_M(z) = \frac{2c}{\sqrt{\Omega_m}H_0}\left(1 - \frac{1}{\sqrt{1+z}}\right). \tag{1.134}$$

Dividing the previous result by $(1+z)$ and evaluating at the epoch of recombination (using $h = 0.7$ and $\Omega_m = 0.3$),

$$d_A(z_{\rm rec}) \approx 14.21 \text{ Mpc}. \tag{1.135}$$

On the other hand, the particle horizon, eq. (1.96), can be computed as

$$d_P(t) = a(t)\int_0^t \frac{cdt'}{a(t')} \approx \frac{2c}{\sqrt{\Omega_m}H_0}\frac{1}{(1+z)^{3/2}}. \tag{1.136}$$

Evaluating at the epoch of recombination,

$$d_P(z_{\rm rec}) \approx 0.43 \text{ Mpc}. \tag{1.137}$$

Dividing the particle horizon by the angular diameter distance, we find that we would expect any region of the CMB within ∼1.73 deg of angular separation to have been in causal contact, bu at any scale larger than that there should have been no exchange of information. The horizon problem describes the fact that we see isotropy in the CMB temperature across the entire sky, despite the entire sky not being in causal contact to establish thermal equilibrium. This problem can be solved by inflation, i.e., introducing a short exponential expansion phase during the early stages of the Universe.

1.5.2 Inflaton Scalar Field

Inflation is a period of exponential expansion in which

$$\ddot{a} > 0. \tag{1.138}$$

Since we know that inflation ends at a certain instant, the inflationary period must be controlled by a dynamic vacuum other than the cosmological constant Λ. Therefore, we need an additional scalar field ϕ, which is commonly known as inflaton. The Lagrangian density of the inflaton is defined as

$$\mathcal{L}_\phi = \left[\frac{1}{2}\nabla^\alpha\phi\nabla_\alpha\phi - V(\phi)\right]\sqrt{-g}. \tag{1.139}$$

The total Lagrangian density is given by the sum of the Hilbert-Einstein contribution, eq. (1.10), plus the one for the inflaton,

$$\mathscr{L} = \frac{c^4}{16\pi G}(R-2\Lambda)\sqrt{-g} + \left[\frac{1}{2}\nabla^\alpha\phi\nabla_\alpha\phi - V(\phi)\right]\sqrt{-g} = L\sqrt{-g}. \tag{1.140}$$

We can now deduce the equations of motion for our full Lagrangian:

- Variation with respect to ϕ.

$$\frac{\partial L}{\partial \phi} - \nabla_\rho \left(\frac{\partial L}{\partial \nabla_\rho \phi} \right) = -V'(\phi) - \nabla^\alpha \nabla_\alpha \phi = 0. \tag{1.141}$$

From now on we will consider that $\phi = \phi(t)$. Hence, since $\Gamma^0{}_{00} = \Gamma^0{}_{0i} = 0$ and $\Gamma^0{}_{ij} = H g_{ij}$,

$$\ddot{\phi} + 3H\dot{\phi} - V'(\phi) = 0. \tag{1.142}$$

This is the equation of motion for the inflaton, and it represents a damped harmonic oscillator. The term $3H\dot{\phi}$ acts as a friction term whose viscosity is given by H. If $3H\dot{\phi} \approx V'(\phi)$, then the motion is similar to that of an overdamped harmonic oscillator, which allows inflation to end.

- Variation with respect to $g^{\mu\nu}$. We calculate the stress-energy tensor of the total action, and we obtain

$$\begin{aligned} \frac{1}{2} T_{\mu\nu} = \frac{\partial L}{\partial g^{\mu\nu}} - \frac{1}{2} g_{\mu\nu} L = \frac{c^4}{16\pi G} \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \right) \\ + \frac{1}{2} \left[\nabla_\mu \phi \nabla_\nu \phi - g_{\mu\nu} \left(\frac{1}{2} \nabla^\alpha \phi \nabla_\alpha \phi - V \right) \right]. \end{aligned} \tag{1.143}$$

Therefore,

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} \\ + \frac{8\pi G}{c^4} \left[\nabla_\mu \phi \nabla_\nu \phi - g_{\mu\nu} \left(\frac{1}{2} \nabla^\alpha \phi \nabla_\alpha \phi - V \right) \right] = \frac{8\pi G}{c^4} T_{\mu\nu}. \end{aligned} \tag{1.144}$$

Let's define

$$T^{\phi\mu}{}_\nu \equiv g^{\mu\rho} \nabla_\rho \phi \nabla_\nu \phi - \delta^\mu{}_\nu \left(\frac{1}{2} g^{\alpha\beta} \nabla_\alpha \phi \nabla_\beta \phi - V \right). \tag{1.145}$$

Since $g_{00} = -1$ and $g^{00} = -1$,

$$T^{\phi\mu}{}_\nu = g^{\mu 0} \dot{\phi} \nabla_\nu \phi - \delta^\mu{}_\nu \left(-\frac{1}{2} \dot{\phi}^2 - V \right). \tag{1.146}$$

If we take $T^\phi_{\mu\nu}$ as the stress-energy tensor of a relativistic perfect fluid,

$$T^\mu{}_\nu = \left(\rho + \frac{p}{c^2} \right) u^\mu u_\nu + p \delta^\mu{}_\nu, \tag{1.147}$$

we find that $T^0{}_0 = -c^2 \rho$, which yields to

$$-\rho = -\dot{\phi}^2 - \left(-\frac{1}{2} \dot{\phi}^2 - V \right) = -\frac{1}{2} \dot{\phi}^2 + V. \tag{1.148}$$

We also find that $T^i{}_j = p\delta^i{}_j$, which implies

$$p\delta^i{}_j = -\delta^i{}_j\left(-\frac{1}{2}\dot{\phi}^2 - V\right) = \left(\frac{1}{2}\dot{\phi}^2 + V\right)\delta^i{}_j. \tag{1.149}$$

Therefore,

$$\begin{cases} \rho = \dfrac{1}{2}\dot{\phi}^2 - V, \\ p = \dfrac{1}{2}\dot{\phi}^2 + V. \end{cases} \tag{1.150}$$

With these expressions for the density and the pressure, we can write the Friedmann and acceleration equations in terms of H instead of a, and we find

$$\begin{cases} H^2 = \dfrac{8\pi G}{3}\rho - \dfrac{c^2k}{a^2} = \dfrac{8\pi G}{3}\left(\dfrac{1}{2}\dot{\phi}^2 - V\right) - \dfrac{c^2k}{a^2}, \\ \dot{H} = -4\pi G\left(\rho + \dfrac{p}{c^2}\right) + \dfrac{c^2k}{a^2} = -4\pi G\dot{\phi}^2 + \dfrac{c^2k}{a^2}. \end{cases} \tag{1.151}$$

These equations are necessary to compute the explicit expressions for the slow-roll parameters, which we will define in the next section.

1.5.3 Slow-Roll Conditions

We already deduced the equation of motion for the inflaton, eq. (1.142). We discussed that, if $3H\dot{\phi} \approx V'(\phi)$, then the motion is similar to that of an overdamped harmonic oscillator, which allows inflation to end. This is an important condition, since we know inflation ends at some point. However, inflation has to last long enough in order to be able to solve the flatness and horizon problems. One of the solutions that incorporates both requirements is the slow-roll approximation, named after the fact that it is analogous to having our field ϕ slowly rolling down its potential, since its kinetic energy is much smaller than its potential, i.e., $\dot{\phi}^2 \ll V$. Therefore, there are two slow-roll conditions. The first one is given by

$$\ddot{\phi} \ll \begin{cases} V'(\phi), \\ 3H\dot{\phi}, \end{cases} \tag{1.152}$$

and it leads to

$$\dot{\phi} \approx \frac{V'}{3H} \tag{1.153}$$

through eq. (1.142). The second slow-roll condition is given by

$$\dot{\phi}^2 \ll V(\phi). \tag{1.154}$$

Under these slow-roll conditions, eq. (1.150) naturally yields

$$p \approx -c^2\rho, \tag{1.155}$$

i.e., the inflaton behaves, effectively, as an early-time cosmological constant. Solving for the number of e-folds in eq. (1.129) and manipulating its expression, we find that

$$\begin{aligned} N = \log\left(\frac{a_f}{a_i}\right) = \int_{a_i}^{a_f} \frac{da}{a} = \int_{t_i}^{t_f} dt H = \int_{\phi_i}^{\phi_f} d\phi \frac{H}{\dot{\phi}} \\ \approx 3\int_{\phi_i}^{\phi_f} d\phi \frac{H^2}{V'} \approx -8\pi G \int_{\phi_i}^{\phi_f} d\phi \frac{V}{V'}. \end{aligned} \tag{1.156}$$

The initial value of the scalar field, ϕ_i, is a parameter of the model. On the other hand, ϕ_f is given by the condition

$$\max\{\epsilon(\phi_f), \eta(\phi_f)\} = 1, \tag{1.157}$$

where ϵ and η, together with δ, are the so-called slow-roll parameters, which are defined as

$$\begin{cases} \epsilon \equiv -\dfrac{\dot{H}}{H^2} \approx \dfrac{1}{16\pi G}\left(\dfrac{V'}{V}\right)^2, \\ \delta \equiv -\dfrac{\ddot{\phi}}{\dot{\phi}H} \approx \dfrac{1}{8\pi G}\dfrac{V''}{V} - \epsilon, \\ \eta \equiv \delta + \epsilon \approx \dfrac{1}{8\pi G}\dfrac{V''}{V}, \end{cases} \tag{1.158}$$

where we used eq. (1.151) to deduce the right-hand side expressions. By definition, ϵ, δ and η are $\ll 1$ during inflation, and of order one whenever inflation ends. The field ϕ is affected by quantum fluctuations of random nature which are assumed to be Gaussian distributed with mean zero. This is due to the uncertainty principle, and is fundamental to understand the formation of the LSS. During the inflationary phase, these fluctuations are magnified and stretched out in every direction beyond the Hubble horizon, where they remain frozen. These are the seeds of the perturbations in the density field which give rise, through gravitational collapse, to the formation of cosmic structures. As we will see in Chapter 2, these fluctuations are described by the so-called primordial power spectrum, $P_R(k)$, which can be fully characterized by its amplitude, A_s, and its spectral index, n_s. These parameters are, in turn, related to the slow-roll parameters ϵ and δ.

Chapter 2

The Large-Scale Structure of the Universe

One of the most important achievements of modern cosmology is the measurement and description of the LSS of the Universe. Galaxies are not evenly distributed in space, they are clustered. On large scales, the Universe shows a very rich structure with a hierarchical order, in such a way that galaxies form clusters, and clusters of galaxies are interconnected in superclusters by "filaments" and "walls", which surround vast empty regions of space, known as voids. However, this hierarchy has a limit: on scales larger than 100 Mpc, the distribution of matter in the Universe becomes smooth and tends to uniformity, in agreement with the homogeneity hypothesis of the cosmological principle. In this chapter we introduce the basics to describe the large-scale structure of the Universe, starting from a theoretical description to then study the matter power spectrum and the two-point correlation function. We also describe the physics of baryon acoustic oscillations (BAO), which is the cosmological probe we focus on throughout this thesis. The BAO feature plays a crucial role as a cosmological probe due to its significance in measuring the expansion history of the Universe and constraining cosmological parameters [20], and it can also be used to test its consistency with the high-redshift results coming from Planck [14].

2.1 Theoretical Description of the LSS

The theoretical description of the LSS of the Universe is relatively simple. It requires a mechanism to create the seeds, in the form of small accumulations of matter, and some physical process that induces the growth of these initial seeds into the hierarchical clustering pattern we see today.

The currently favoured mechanism for the creation of these seeds is cosmic inflation, which explains them as the quantum fluctuations of a field, the "inflaton". This means that we need both general relativity and quantum mechanics to explain the formation of structure in the Universe. The process behind the growth of these

seeds is called gravitational instability. Once the primordial fluctuations appeared, regions with a matter density slightly larger than the average grew due to the purely attractive nature of the gravitational force, resulting in the rich structure we observe today.

The observed LSS cannot be explained if the only matter that exists in the Universe is baryonic matter. In that case, structures as large as those we observe today could not grow. Some additional gravitating matter is needed, i.e., the dark matter. In fact, most of the matter in the Universe must be dark. In addition, the action of a repulsive force at late times is needed to explain the observed sizes and amounts of clusters and superclusters. This is the dark energy. As we can see, the measurement of the properties and evolution of the LSS is extremely sensitive to the dark sector of the Universe.

2.2 The Primordial Power Spectrum

One of the elements of the ΛCDM model is inflation. Inflation requires the existence of a scalar field called inflaton, whose quantum fluctuations, amplified during the inflationary phase, are the seeds of the inhomogeneities of the density field. These inhomogeneities are the cause of the temperature anisotropies found in the CMB and the LSS we observe today.

In order to study the fluctuations of the inflaton field, it is necessary to take perturbations of its action, introducing $\delta\phi$, and then solving the equations of motion. After this, it is also necessary to quantize the perturbation field. The equations of motion are usually written in terms of the so-called Mukhanov-Sasaki variable, which contains the perturbation field, $\delta\phi$. The Mukhanov-Sasaki variable is related to the perturbations in the curvature, R, which are, in turn, related to the Newtonian potential, Φ. The Newtonian potential can be connected with the density contrast[1] of the different components of the Universe, which are the quantities we are actually interested in.

The fluctuations on the inflaton field are of random nature and, therefore, the average of a perturbation variable is always zero. This is the reason why, in order to characterize the fluctuations of a random field R, we evaluate its variance $\langle R^2 \rangle$, which measures the amplitude of the fluctuation. Once the field R is quantized, we can calculate its variance as

$$\langle 0|R^2|0\rangle = \sum_k \sum_{k'} R_k^* R_{k'} \langle 0|a_{k'} a_k^\dagger|0\rangle = \sum_k \sum_{k'} R_k^* R_{k'} \delta_{kk'} = \sum_k |R_k|^2, \tag{2.1}$$

where R_k are the Fourier coefficients[2] of R, $a^\dagger$ is the creation operator and a is the

[1] The density contrast, δ, is given by eq. (2.28).

[2] We assumed the Fourier coefficients $R_{\boldsymbol{k}}$ depend only on the modulus of $\boldsymbol{k}$, which is a consequence of isotropy.

annihilation one. These operators satisfy

$$[a, a^\dagger] = 1 \quad \text{and} \quad a|0\rangle = 0. \tag{2.2}$$

Taking now the continuum limit and integrating over the solid angle, we find

$$\langle 0|R^2|0\rangle = \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}|R_k|^2 = \int_0^\infty \frac{dk}{k} P_R(k), \tag{2.3}$$

where $P_R(k)$ is the power spectrum of our random field R, which is defined as

$$P_R(k) = \frac{k^3|R_k|^2}{2\pi^2}. \tag{2.4}$$

The power spectrum describes the contribution of each mode or scale to the quadratic dispersion of the field. The mode k is related to the physical size of the structures that form the observed inhomogeneities λ by $k = 1/\lambda$. Solving the equation of motion for R_k, it can be found that

$$P_R(k) = A_s \left(\frac{ck}{aH}\right)^{n_s - 1}, \tag{2.5}$$

where A_s is the scalar amplitude of the fluctuations and n_s is the spectral index of the power spectrum. They can be expressed in terms of the slow-roll parameters as

$$\begin{cases} A_s = \dfrac{4\pi G}{\varepsilon}\left(\dfrac{H}{2\pi}\right)^2, \\ n_s = 4 - 2\nu = 1 - 4\varepsilon + 2\delta. \end{cases} \tag{2.6}$$

The scalar amplitude of the fluctuations has a small dependence on time. However, H and ϵ are approximately constant during inflation, so a pivot scale, k_0, is usually taken as a reference,

$$A_s = \frac{4\pi G}{\varepsilon}\left(\frac{H}{2\pi}\right)^2\Bigg|_{aH=ck_0}. \tag{2.7}$$

The power spectrum then takes the expression

$$P_R(k) = A_s \left(\frac{k}{k_0}\right)^{n_s - 1}. \tag{2.8}$$

The slow-roll conditions require that ϵ, δ, $\eta \ll 1$, which implies $n_s \approx 1$. Therefore, the primordial power spectrum $P_R(k)$ depends very weakly on the scale, k, which means that all modes contribute to the power spectrum with approximately the same amplitude.

During inflation, the perturbations of the inflaton field are stretched out until their length is larger than the Hubble horizon, i.e., $\lambda > c/H$, and then they remain frozen. The power spectrum connects the perturbations of the inflaton field with the curvature ones, which remain the same as long as they are beyond the Hubble horizon, carrying useful information about the inflationary epoch.

2.3 Matter Density Fluctuations and Matter Power Spectrum

The matter power spectrum is related to the primordial power spectrum of the fluctuations of the inflaton field, $P_R(k)$. These fluctuations of the inflaton field are the origin of the temperature and density inhomogeneities that we observe on the CMB and the LSS, respectively. The modes that describe these fluctuations, R_k, are of random nature, and we assumed they are drawn from a Gaussian distribution of mean zero. In this section we deduce how these fluctuations evolve with time, and what statistics describe them. We are particularly interested in describing the density fluctuations of matter, which are the ones related to galaxy clustering, one of the main topics of this thesis.

2.3.1 Evolution of Matter Density Fluctuations

Let us define the matter density field, $\rho = \rho(t, \boldsymbol{r})$, and the velocity field, $\boldsymbol{u} = \boldsymbol{u}(t, \boldsymbol{r})$. For a perfect fluid, using the Newtonian approximation[3], it is widely known that the equations governing the evolution of the fluid are given by

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla}_r(\rho \boldsymbol{u}) = 0 \quad \text{(Continuity equation)}, \tag{2.9}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \boldsymbol{\nabla}_r)\, \boldsymbol{u} = -\frac{1}{\rho}\boldsymbol{\nabla}_r p - \boldsymbol{\nabla}_r \Phi \quad \text{(Euler equation)}, \tag{2.10}$$

$$\boldsymbol{\nabla}_r^2 \Phi = 4\pi G \rho \quad \text{(Poisson equation)}, \tag{2.11}$$

$$p = p(\rho) \quad \text{(Equation of state)}. \tag{2.12}$$

The operator $\boldsymbol{\nabla}$ is the spatial gradient with respect to the proper coordinates $\boldsymbol{r}$, and $\boldsymbol{u}$ is the proper velocity of the fluid,

$$\boldsymbol{u}(t, \boldsymbol{r}) \equiv \frac{d\boldsymbol{r}}{dt}. \tag{2.13}$$

Let us first consider a static non-expanding universe, assuming a homogeneous isotropic background with constant matter density, $\rho(t, \boldsymbol{r}) = \rho_0$. Slightly perturbing the matter distribution, we have

$$\begin{cases} \rho(t, \boldsymbol{r}) = \rho_0 + \delta\rho(t, \boldsymbol{r}), \\ \boldsymbol{u}(t, \boldsymbol{r}) = \delta\boldsymbol{u}(t, \boldsymbol{r}), \\ \Phi(t, \boldsymbol{r}) = \Phi_0 + \delta\Phi(t, \boldsymbol{r}), \\ p(t, \boldsymbol{r}) = p_0 + \delta p(t, \boldsymbol{r}). \end{cases} \tag{2.14}$$

[3] The Newtonian approximation accurately describes the matter density fluctuations, which are the ones we are interested in. However, it is not suitable for describing, e.g., radiation density fluctuations.

In linear approximation,

$$\delta p = c_s^2 \delta\rho, \tag{2.15}$$

where c_s is the speed of sound of the fluid. Substituting the perturbed quantities into eq. (2.9) and keeping terms up to first order in the perturbations, we obtain

$$\frac{\partial \delta\rho}{\partial t} + \rho \boldsymbol{\nabla}_r \cdot \delta\boldsymbol{u} = 0, \tag{2.16}$$

$$\frac{\partial \delta\boldsymbol{u}}{\partial t} = -\frac{c_s^2}{\rho}\boldsymbol{\nabla}_r \delta\rho - \boldsymbol{\nabla}_r \delta\Phi, \tag{2.17}$$

$$\boldsymbol{\nabla}_r^2 \delta\Phi = 4\pi G \delta\rho. \tag{2.18}$$

Let us now consider an expanding homogeneous and isotropic universe. The background matter density is a function of time, and the background velocities obey the Hubble law, which means

$$\rho_0 = \rho_0(t), \quad \boldsymbol{u}_0 = \boldsymbol{u}_0(t, \boldsymbol{r}) = H(t)\boldsymbol{r}. \tag{2.19}$$

Slightly perturbing the matter distribution, we have

$$\begin{cases} \rho(t, \boldsymbol{r}) = \rho_0(t) + \delta\rho(t, \boldsymbol{r}), \\ \boldsymbol{u}(t, \boldsymbol{r}) = \boldsymbol{u}_0(t, \boldsymbol{r}) + \delta\boldsymbol{u}(t, \boldsymbol{r}), \\ \Phi(t, \boldsymbol{r}) = \Phi_0 + \delta\Phi(t, \boldsymbol{r}), \\ p(t, \boldsymbol{r}) = p_0(t) + \delta p(t, \boldsymbol{r}). \end{cases} \tag{2.20}$$

Substituting the previous expressions into eq. (2.9) and keeping terms up to first order in the perturbations, we obtain

$$\frac{\partial \delta\rho}{\partial t} + \rho_0 \boldsymbol{\nabla}_r \cdot \delta\boldsymbol{u} + \boldsymbol{\nabla}_r(\delta\rho \cdot \boldsymbol{u}_0) = 0, \tag{2.21}$$

$$\frac{\partial \delta\boldsymbol{u}}{\partial t} + (\boldsymbol{u}_0 \cdot \boldsymbol{\nabla}_r)\delta\boldsymbol{u} + (\delta\boldsymbol{u} \cdot \boldsymbol{\nabla}_r)\boldsymbol{u}_0 = -\frac{c_s^2}{\rho}\boldsymbol{\nabla}_r \delta\rho - \boldsymbol{\nabla}_r \delta\Phi, \tag{2.22}$$

$$\boldsymbol{\nabla}_r^2 \delta\Phi = 4\pi G \delta\rho. \tag{2.23}$$

The Hubble velocity $\boldsymbol{u}_0$ depends explicitly on $\boldsymbol{r}$ and, therefore, its Fourier transform does not reduce these equations to a decoupled set of ordinary differential equations. This is why it is more convenient to use Lagrangian or comoving coordinates $\boldsymbol{x}$, which are related to the Eulerian ones as

$$\boldsymbol{r} = a(t)\boldsymbol{x}. \tag{2.24}$$

We, then, have

$$\boldsymbol{\nabla}_r \to \frac{1}{a}\boldsymbol{\nabla}_x \tag{2.25}$$

and

$$\left(\frac{\partial}{\partial_t}\right)\Big|_{\boldsymbol{r}} \rightarrow \left(\frac{\partial}{\partial_t}\right)\Big|_{\boldsymbol{x}} - H\boldsymbol{x}\cdot\boldsymbol{\nabla}_x, \tag{2.26}$$

since

$$\left(\frac{\partial f(t,\boldsymbol{r}=a\boldsymbol{x})}{\partial_t}\right)\Big|_{\boldsymbol{x}} = \left(\frac{\partial f}{\partial t}\right)\Big|_{\boldsymbol{r}} + \dot{a}\boldsymbol{x}\cdot\left(\boldsymbol{\nabla}_r f\right)_t. \tag{2.27}$$

In cosmology, it is very common to work with the density contrast field, or over-density field, which is defined as

$$\delta(\boldsymbol{x}) \equiv \frac{\delta\rho(\boldsymbol{x})}{\bar{\rho}} = \frac{\rho(\boldsymbol{x})-\bar{\rho}}{\bar{\rho}}, \tag{2.28}$$

where $\rho(\boldsymbol{x})$ is the matter density field and $\bar{\rho}$ is the average density of the Universe. By definition, δ can take values in the interval $[-1,\infty)$. Hereafter we assume that the fluctuations of the density contrast field are small, i.e., $|\delta| \ll 1$, and so we work in the linear regime. $\delta(\boldsymbol{x})$ can then be thought of as a random field that follows a Gaussian distribution. We describe the properties of such a field in section Appendix A.1. On the other hand, the average density $\bar{\rho}$ can be interpreted as the background density, ρ_0. Equations (2.21) can now be written, in comoving coordinates, as

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\boldsymbol{\nabla}_x\cdot\delta\boldsymbol{u} = 0, \tag{2.29}$$

$$\frac{\partial\delta\boldsymbol{u}}{\partial t} + H\delta\boldsymbol{u} = -\frac{c_s^2}{a}\boldsymbol{\nabla}_x\delta - \frac{1}{a}\boldsymbol{\nabla}_x\delta\Phi, \tag{2.30}$$

$$\boldsymbol{\nabla}_x^2\delta\Phi = 4\pi G a^2\delta\rho. \tag{2.31}$$

Combining the three equations, and then moving to Fourier space, we obtain

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \left(\frac{c_s^2}{a^2}k^2 - 4\pi G\bar{\rho}\right)\delta_k = 0, \tag{2.32}$$

where δ_k are the Fourier modes of the density contrast, $\delta(\boldsymbol{x})$. The third term in the previous equation defines a characteristic proper length, the so-called Jeans length, which is given by

$$\lambda_J \equiv \frac{2\pi a}{k_J} = c_s\sqrt{\frac{\pi}{G\bar{\rho}}}. \tag{2.33}$$

Making a change of variables in eq. (2.32) from cosmic to conformal time ($' \equiv d/d\eta$) and expressing it in terms of the Jeans length, we easily find

$$\delta_k'' + aH\delta_k' + c_s^2\left(k^2 - k_J^2\right)\delta_k = 0. \tag{2.34}$$

For perturbations with $k \ll k_J$, or $\lambda \gg \lambda_J$, the pressure term of eq. (2.34) can be neglected. In this situation, matter behaves as a pressureless fluid, and fluctuations grow. On the other hand, for perturbations with $k \gg k_J$, or $\lambda \ll \lambda_J$, the gravity

term can be neglected, and eq. (2.32) turns into a damped oscillator. This means that only perturbations with $k \ll k_J$ can grow with time. Since the sound speed of CDM is small, fluctuations grow efficiently over a large range of scales (since, as we already deduced, fluctuations grow when $k \ll k_J \propto c_s^{-1}$).

Up to this point we have only considered the evolution of matter density perturbations in a universe in which there is only matter. For the evolution of matter fluctuations in the primordial photon gas, pressure plays an important role: fluctuations only grow on very large scales and oscillate otherwise. These acoustic oscillations are observed in the CMB and in the distribution of galaxies. Therefore, it is also interesting to study the evolution of matter perturbations in a universe in which there are also radiation and dark energy acting in the background and modifying the expansion rate. In this case the equation to be solved is the so-called Mészáros equation, which is given by [21]

$$y^2(1+y^{-3w})\frac{d^2\delta_k}{dy^2} + \frac{3}{2}y\left[1+(1-w)y^{-3w}\right]\frac{\delta_k}{dy} - \frac{3}{2}\delta_k = 0, \tag{2.35}$$

where w is the parameter of the equation of state of the dominant species and the variable y is given by

$$y = \frac{a}{a_{\rm eq}}, \tag{2.36}$$

with $a_{\rm eq}$ the scale factor at the epoch of matter-radiation equality.

- At early times, the Universe was dominated by a mixture of radiation and matter. In the particular case in which radiation is the dominant species, the Mészáros equation can be written as

$$y(1+y)\frac{d^2\delta_k}{dy^2} + \frac{2+3y}{2}\frac{d\delta_k}{dy} - \frac{3}{2}\delta_k = 0. \tag{2.37}$$

  The solutions to this equation take the form

$$\delta_k \propto \begin{cases} 2+3y, \\ (2+3y)\log\left(\dfrac{\sqrt{1+y}+1}{\sqrt{1+y}-1}\right) - 6\sqrt{1+y}. \end{cases} \tag{2.38}$$

  In the radiation-dominated limit, $y \ll 1$, the growing-mode solution is $\delta_k \propto \log y \propto \log a$, i.e., matter fluctuations only grow logarithmically in the radiation era. In the matter-dominated limit, $y \gg 1$, the growing-mode solution is $\delta_k \propto y \propto a$.

- At late times, the Universe is a mixture of matter and dark energy. Since dark energy doesn't have fluctuations, eq. (2.34) still applies. For sub-horizon scales,

$$\delta_k'' + aH\delta_k' - c_s^2 k_J^2 \delta_k \approx 0. \tag{2.39}$$

In the dark-energy-dominated era we have $(aH)^2 = (-\eta^{-1})^2 \gg c_s^2 k_J^2$, which means we can drop the last term in the previous equation. Therefore,

$$\delta_k'' - \frac{1}{\eta}\delta_k' \approx 0, \tag{2.40}$$

which has the following solution

$$\delta_k \propto \begin{cases} \text{const}, \\ \eta^2 \propto a^{-2}. \end{cases} \tag{2.41}$$

Matter fluctuations stop growing once dark energy starts to dominate.

The general solution to the Mészáros equation, independently of the dominant species, i.e., independently of w, is given by

$$\delta(t, \boldsymbol{x}) = f_D(\boldsymbol{x}) D\left(a(t)\right) + f_E(\boldsymbol{x}) E\left(a(t)\right). \tag{2.42}$$

The two additive terms in the previous expression, $f_D(\boldsymbol{x})D\left(a(t)\right)$ and $f_E(\boldsymbol{x})E\left(a(t)\right)$, are the growing and the decaying modes, respectively. $D(a)$ is known as the linear growth factor, and has the form [22]

$$D(a) = \frac{H(a)}{H_0} \int_0^a \frac{da'}{\left(a' H(a')\right)^3} \left[\int_0^1 \frac{da'}{\left(a' H(a')\right)^3} \right]^{-1}. \tag{2.43}$$

It is normalized so that

$$D(a = 1) = 1. \tag{2.44}$$

The growth factor is particularly useful in cosmology, since it describes the growth of linear perturbations in δ within the Hubble horizon with an arbitrary background (radiation or dark energy). However, the so-called growth index γ is the parameter usually used in order to establish the connection between the effect of dark energy on the growth history of LSS while trying to be as model independent as possible [23]. The growth index γ is introduced as

$$g(a) = \exp\left(\int_0^a \frac{da}{a} \left(\Omega_m^\gamma - 1 \right) \right). \tag{2.45}$$

Cosmological models with the same expansion history but with a different underlying gravity have different values of γ. Therefore, fitting this $g(a)$ to the data can help us ruling out gravitation models. In particular, for the ΛCDM model it is found that $\gamma = 0.55$ to better than 0.05% in the range $\Omega_m \in [0.22, 2]$ [24].

In order to relate all of the previous results to the primordial power spectrum, it is necessary to extend out analysis to a general relativity framework. It is then necessary to solve the Einstein field equations for the FLRW metric. It can be found that

$$\delta_c(k, a) = -\frac{3}{5} \frac{c^2 k^2}{\Omega_m H_0^2} \Phi_k^0 T(k) D(a), \tag{2.46}$$

where Φ_k^0 is the scalar perturbation of the FLRW metric, which can be interpreted as the Newtonian potential, and $T(k)$ is the so-called transfer function. Computing the transfer function requires solving the Boltzmann equations in a perturbed FLRW metric, which is a formidable task. These Boltzmann equations encompass all interactions and couplings between all the different particles that make up the Universe. Under certain approximations, the transfer function can be written as

$$T(k) \propto \begin{cases} 1, & k < k_{\rm eq}, \\ \dfrac{1}{k^2}\log k, & k > k_{\rm eq}, \end{cases} \tag{2.47}$$

where $k_{\rm eq}$ represents the scale entering the horizon at the matter-radiation equality epoch. For a more complete and precise computation of the transfer function, specific codes such as CAMB [25] are usually used by the cosmological community. Also, in order to introduce the small scale non-linear effects in the power spectrum, i.e., in the largest k values, additional codes such as the HALOFIT model [26] are particularly useful.

Since we are particularly interested in studying baryon acoustic oscillations in the following chapters of this thesis, it is also necessary to describe the evolution of baryon density perturbations. For baryons coupled with the photons of the primordial plasma, the density contrast for modes within the Hubble horizon is given by

$$\delta_b(k,a) = -4\left[T(k) - \frac{5}{3}\cos(kr_s)\exp\left(-k^2/k_D^2\right)\right]\frac{9}{10}\Phi_k^0, \tag{2.48}$$

where k_D is the Silk damping scale and r_s is the sound horizon scale in the fluid, which can be computed directly integrating the speed of sound,

$$r_s(\eta) = \int_0^\eta d\eta' c_s(\eta'). \tag{2.49}$$

The oscillatory behaviour of the cosine in eq. (2.48) gives rise to the baryon acoustic oscillations, which we describe in section 2.5.

2.3.2 The Matter Power Spectrum

Inflationary models predict density perturbations that are generated by Gaussian quantum fluctuations of a scalar field. Therefore, the density perturbations are random fields with Gaussian distribution, and they are completely described by the matter power spectrum, $P_m(k)$. The variance of a Gaussian random variable g can be written as

$$\langle g(\boldsymbol{x})^2\rangle = \int \frac{dk}{k} V \frac{k^3}{2\pi^2}\langle |g_{\boldsymbol{k}}|^2\rangle = \int \frac{dk}{k} P_g(k), \tag{2.50}$$

where V is the volume, g_k are its Fourier modes and

$$P_g(k) \equiv V\frac{k^3}{2\pi^2}\langle |g_{\boldsymbol{k}}|^2\rangle \tag{2.51}$$

is its power spectrum (see section Appendix A.1 for further details on these expressions). Since we have already determined the evolution of the matter density fluctuations (for both cold dark matter and baryons), we can define a power spectrum for them, study its evolution with time and also connect it with the primordial Universe.

As we already deduced in eqs. (2.46) and (2.48), both the CDM and the baryon density fluctuations depend on the primordial fluctuations of the Newtonian potential Φ_k^0,

$$\begin{cases} \delta_c(k,a) = -\dfrac{3}{5}\dfrac{c^2k^2}{\Omega_m H_0^2}\Phi_k^0 T(k) D(a), \\ \delta_b(k,a) = -4\left[T(k) - \dfrac{5}{3}\cos(kr_s)\exp\left(-k^2/k_D^2\right)\right]\dfrac{9}{10}\Phi_k^0. \end{cases} \tag{2.52}$$

These perturbation modes are frozen as long as they are out of the Hubble horizon, and therefore they have not evolved since the inflationary epoch. Hence, they carry very valuable information. They can be related to the curvature perturbation modes, R_k, when they left the horizon during inflation via

$$R_k = \frac{5+3w}{3+3w}\Phi_k^0, \tag{2.53}$$

where w is the equation-of-state parameter of the species dominating at the moment these modes re-enter the Hubble horizon. The power spectrum of Φ_k^0 can be written as

$$P_\Phi(k) = \frac{k^3}{2\pi}\langle|\Phi_k^0|^2\rangle = \frac{4}{9}P_R(k), \tag{2.54}$$

where we assumed $w = 1/3$ because, at the end of inflation, the Universe was dominated by radiation. Following a similar procedure, we can define the power spectrum of CDM density fluctuations as

$$P_{\delta_c}(k) = \frac{k^3}{2\pi^2}\langle|\delta_c(\boldsymbol{k},a)|^2\rangle = \frac{4}{25}\frac{c^4k^4}{\Omega_m^2 H_0^4}T^2(k)D^2(a)P_R(k). \tag{2.55}$$

The so-called CDM power spectrum is usually defined as

$$P_c(k) \equiv \frac{2\pi^2}{k^3}P_{\delta_c}(k), \tag{2.56}$$

and has dimensions of L^3. Using the expression for the transfer function $T(k)$, eq. (2.47), we find that

$$P_c(k) \propto \begin{cases} k, & k < k_{\rm eq}, \\ \dfrac{1}{k^3}(\log k)^2, & k > k_{\rm eq}. \end{cases} \tag{2.57}$$

On large scales (low k values), the CDM power spectrum grows as k, whereas it decreases as $\sim k^{-3}$ at small scales (large k values).

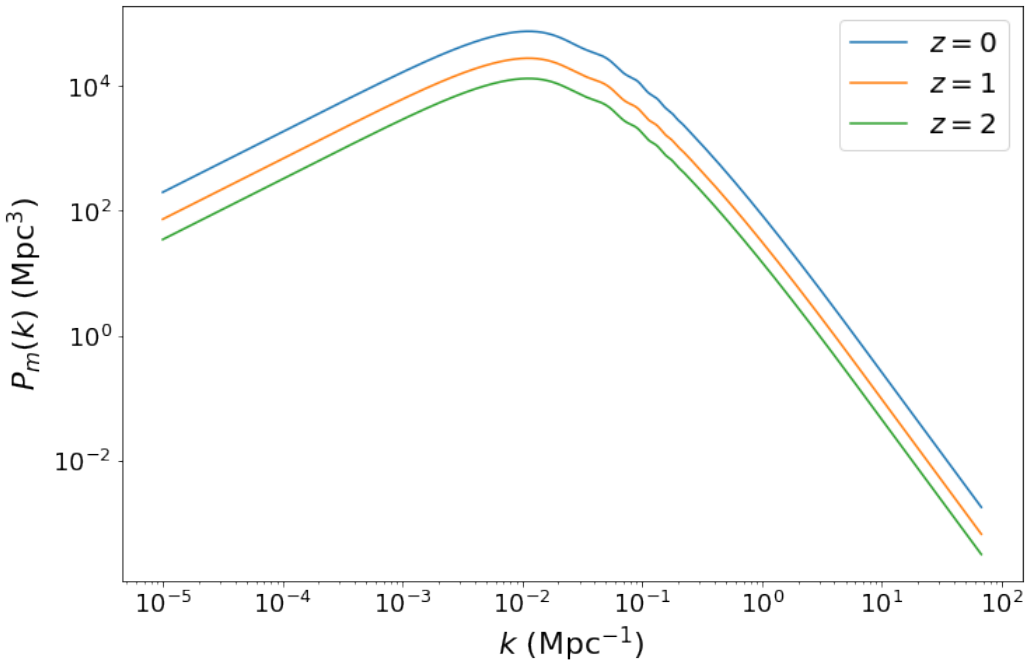

Figure 2.1: Matter power spectrum for different redshifts, computed using CAMB [25] and the cosmological parameters of the MICE "Grand Challenge" simulation, see Table 4.2.

Our main goal now is to compute the total matter power spectrum, i.e., including not only CDM, but also the contribution of baryons,

$$\delta_m = \delta_c + \delta_b. \tag{2.58}$$

The total matter power spectrum, $P_m(k)$, can be obtained following the same steps as for the CDM power spectrum. The contribution of baryons effectively results in oscillations in the matter power spectrum due to their coupling with photons in the very early Universe. These are the so-called baryon acoustic oscillations and, as we already mentioned, their study is one of the main objectives of this thesis.

In Fig. 2.1 we show the matter power spectrum for different redshifts, computed using CAMB [25] and the cosmological parameters of the MICE (Marenostrum Institut de Ciències de l'Espai) "Grand Challenge" simulation [27], listed in Table 4.2. It can be seen that the lower the redshift is, i.e., the later in time, the higher the amplitude of each mode k is. The turning point corresponds to $k_{\rm eq}$, and the wiggles at around $k \sim 10^{-1}$ Mpc^{-1} are due to the oscillatory behaviour of δ_b, corresponding to the BAO feature in real space.

The normalization of the matter power spectrum is also of special interest in cosmology. It is usually normalized as a function of the so-called σ_8 parameter, which

is defined as

$$\sigma_8^2 = \int_0^\infty dk \frac{k^2}{2\pi^2} \left[3\frac{j_1(8k)}{8k} \right]^2 P_m(k, z=0), \tag{2.59}$$

where k has units of h/Mpc. Physically, σ_8 is the typical amplitude or standard deviation of the matter fluctuations measured within a sphere of radius 8 Mpc/h at present time. In other words, σ_8 describes the clumpiness of the cosmos across vast distances. The value of 8 Mpc/h is chosen because it is typically accessible to large galaxy surveys, and also because it has a value of the order of unity. Since the amplitude of primordial fluctuations, A_s, is also proportional to the power spectrum, both σ_8 and A_s are strongly correlated. Not only that, but σ_8 is also correlated with Ω_m, see [28] for reference. This is the reason why S_8, defined as

$$S_8 \equiv \sigma_8 \left(\frac{\Omega_m}{0.3} \right)^{1/2}, \tag{2.60}$$

is commonly used to measure the clumpiness of the Universe, instead of σ_8. In Appendix A.2 we further describe the σ_8 parameter and its relationship with the concept of smoothing.

2.3.3 The Large-Scale Galaxy Bias

It is important to realize that galaxy surveys do not measure the matter density field itself, but rather the distribution of galaxies or other tracers[4], i.e., the distribution of highly nonlinear objects which underwent complex formation processes. The so-called galaxy bias gives use the statistical relation between the distribution of galaxies and matter. The observed galaxies are a biased tracer of the underlying matter density field. The relation between the density contrast of galaxies and that of matter is given by

$$\delta_G(\boldsymbol{x}) = b(z)\delta_m(\boldsymbol{x}), \tag{2.61}$$

where $b(z)$ is the linear galaxy bias. In principle, it can also depend on the scale, but in the linear regime it is assumed to be constant. The linear galaxy bias can be seen as a parameter that measures the efficiency in tracing the underlying matter density contrast, since it has different values depending on the tracer observed. For example, luminous red galaxies (LRG) are highly biased tracers of the large-scale structure, yielding a higher per-object signal-to-noise ratio for the BAO measurement compared to other types of galaxies. LRGs are galaxies that merge with each other as they fall into clusters and lose their gas due to various processes, which results in redder[5] massive elliptical galaxies with low star-formation rates. One of the main features of LRGs is their characteristic 4000 Å break, which makes them particularly useful

[4] Galaxies are not the only tracers of the large-scale structure. Other important examples are clusters of galaxies, voids, quasars, the Lyman-α forest and even the 21 cm hydrogen hyperfine structure transition, among others.

[5] The blue light is absorbed by the surrounding dust.

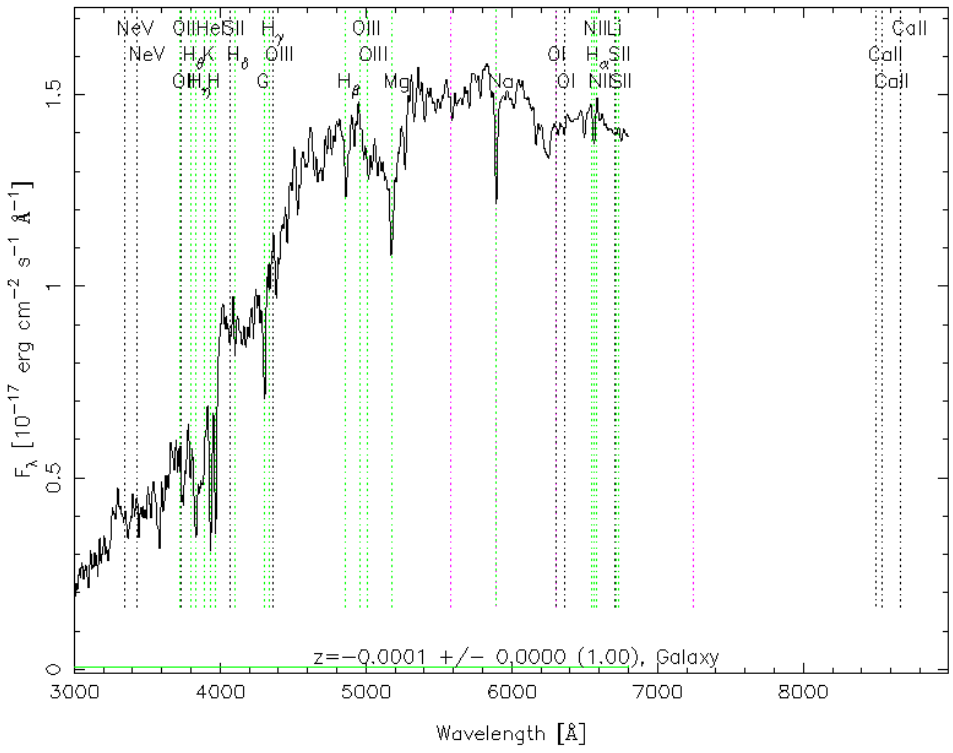

Figure 2.2: Spectral cross-correlation template for LRGs used in SDSS. Cross-correlation redshift was one of the redshift measurements of SDSS, and consisted of cross-correlating the spectra of galaxies with that of a template obtained from SDSS commissioning spectra of high signal-to-noise ratio. Credits: https://classic.sdss.org/dr5/algorithms/spectemplates/.

to find galaxy clusters and to determine their redshifts. Searching for the redshifted 4000 Å break in the different bands of a photometric redshift survey (such as DES) allows us to estimate their redshift. In Fig. 2.2 we show the spectral cross-correlation template for LRGs used in the Sloan Digital Sky Survey (SDSS), in which the 4000 Å break can be observed.

Finally, we can also compute the galactic power spectrum as a function of the matter power spectrum as

$$P_G(k) = b^2(z)P_m(k). \tag{2.62}$$

The amplitude of the galactic power spectrum depends on the product of b^2 and σ_8, which means these two parameters are degenerate.

2.4 Two-Point Statistics

The initial density perturbations that acted as seeds for the formation of structure in the early Universe were quantum fluctuations of the primordial inflaton field. Hence, they cannot be predicted from theory. The precise locations of galaxies, clusters, superclusters and voids are the result of a random process and can only be studied statistically. Huge maps of the Universe must be constructed, where as many galaxies as possible are located, and the statistical properties of the galaxy distribution are studied on these maps. What is measured in practice is not the matter power spectrum, $P_m(k)$, that we defined in the previous section, but rather the galactic power spectrum, $P_G(k)$, or its Fourier transform, the two-point correlation function, $\xi_G(r)$.

The two-point correlation function is the excess of probability of finding a pair of galaxies at a given distance when compared to the probability calculated using a fully uniform random distribution. Different amounts of dark matter and dark energy, or different properties of dark energy, predict different correlation functions and different evolution of the correlation functions with time. Measuring the correlation function at different epochs in the Universe allows us to infer its matter–energy content.

The two-point correlation function of a random fluctuation $g(\boldsymbol{x})$ is mathematically defined as

$$\xi_g(\boldsymbol{x}_1, \boldsymbol{x}_2) \equiv \langle g(\boldsymbol{x}_1) g(\boldsymbol{x}_2) \rangle = \frac{1}{(2\pi)^3} \int d^3\boldsymbol{k} d^3\boldsymbol{k}' \langle g(\boldsymbol{k}) g^*(\boldsymbol{k}') \rangle e^{i\boldsymbol{k}\cdot\boldsymbol{x}_1} e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}_2}. \tag{2.63}$$

Expressing the mean value of $g(\boldsymbol{k})g^*(\boldsymbol{k}')$ as

$$\begin{aligned}
\langle g(\boldsymbol{k}) g^*(\boldsymbol{k}') \rangle &= (2\pi)^3 \left(\frac{L}{2\pi}\right)^3 \left(\frac{L}{2\pi}\right)^3 \langle g_{\boldsymbol{k}} g^*_{\boldsymbol{k}'} \rangle \\
&= (2\pi)^3 \left(\frac{L}{2\pi}\right)^3 \left(\frac{L}{2\pi}\right)^3 \frac{1}{V} \frac{2\pi^2}{k^3} P_g(k) \delta_{\boldsymbol{k}\boldsymbol{k}'} \\
= (2\pi)^3 \left(\frac{L}{2\pi}\right)^3 \frac{1}{V} \frac{2\pi^2}{k^3} P_g(k) \delta^3(\boldsymbol{k} - \boldsymbol{k}') &= \frac{2\pi^2}{k^3} P_g(k) \delta^3(\boldsymbol{k} - \boldsymbol{k}'),
\end{aligned} \tag{2.64}$$

we find that

$$\begin{aligned}
\xi_g(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \frac{1}{(2\pi)^3} \int d^3\boldsymbol{k} d^3\boldsymbol{k}' \frac{2\pi^2}{k^3} P_g(k) \delta^3(\boldsymbol{k} - \boldsymbol{k}') e^{i\boldsymbol{k}\cdot\boldsymbol{x}_1} e^{-i\boldsymbol{k}'\cdot\boldsymbol{x}_2} \\
&= \frac{1}{(2\pi)^3} \int d^3\boldsymbol{k} \frac{2\pi^2}{k^3} P_g(k) e^{i\boldsymbol{k}\cdot(\boldsymbol{x}_1 - \boldsymbol{x}_2)}.
\end{aligned} \tag{2.65}$$

Assuming our random fluctuation is the matter density contrast, i.e., $g(\boldsymbol{x}) = \delta_m(\boldsymbol{x})$,

$$\xi_m(\boldsymbol{s}) \equiv \langle \delta_m(\boldsymbol{x}_1) \delta_m(\boldsymbol{x}_2) \rangle = \frac{1}{(2\pi)^3} \int d^3\boldsymbol{k} P_m(k) e^{i\boldsymbol{k}\cdot\boldsymbol{s}}, \tag{2.66}$$

where we defined

$$\boldsymbol{s} = \boldsymbol{x}_1 - \boldsymbol{x}_2. \tag{2.67}$$

Integrating in φ and θ, and using $\boldsymbol{k}\cdot\boldsymbol{s} = ks\cos\theta$, we find that

$$\xi_m(s) = \int dk \frac{\sin(ks)}{ks}\frac{k^2}{2\pi^2}P_m(k) = \int dk\, j_0(ks)\frac{k^2}{2\pi^2}P_m(k). \tag{2.68}$$

Evaluating at $s = 0$, and taking into account that $\lim_{x\to 0} j_0(x) = 1$, we easily find that

$$\xi_m(s=0) = \int dk \frac{k^2}{2\pi^2}P_m(k). \tag{2.69}$$

Using eq. (2.50), we conclude the previous quantity is none other than the variance of δ_m at a fixed position,

$$\xi_m(s=0) = \langle \delta_m(\boldsymbol{x})^2\rangle. \tag{2.70}$$

Because of this, the variance at a fixed position is also known as the zero-lag correlation function.

Finally, the galaxy two-point correlation function can be computed as

$$\xi_G(s) = b^2(z)\xi_m(s). \tag{2.71}$$

Similarly to the galactic power spectrum, the amplitude of the galactic correlation function is proportional to $b^2\sigma_8$, which establishes a degeneration between these two parameters. It is particularly useful to write the variable s as a function of the angular separations and radial distances to galaxies. Since we defined $\boldsymbol{s} = \boldsymbol{x}_1 - \boldsymbol{x}_2$ and, in spherical coordinates, $x_i = r_i(\sin\theta_i\cos\varphi_i, \sin\theta_i\sin\varphi_i, \cos\theta_i)$, where $r(z) = d_M(z)$ is the comoving angular diameter distance, we easily find that

$$s^2 = |\boldsymbol{x}_1 - \boldsymbol{x}_2|^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\theta, \tag{2.72}$$

where

$$\cos\theta = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\varphi_1 - \varphi_2). \tag{2.73}$$

Therefore,

$$\xi_G(s) = \xi_G(r_1, r_2, \theta) = \xi_G(z_1, z_2, \theta). \tag{2.74}$$

The previous expression connects the two-point correlation function to observable quantities, such as redshifts and angles.

2.4.1 The Angular Correlation Function

The different power spectra and correlation functions defined previously are defined in three-dimensional space. The position of each object in the Universe can be specified by two angular coordinates, θ and φ, and one radial coordinate, $r(z)$, which gives its distance to us along the LOS via its redshift, z. In order to compute distances to objects from their redshifts, we must assume a cosmological model, e.g., the ΛCDM model, and also some cosmological parameters, e.g., Ω_m and H_0. Because of this, uncertainties in the values of z directly translates into uncertainties in the value of r,

and ultimately into uncertainties in $\xi_m(r)$. This is the case of photometric redshift surveys, such as DES, in which we rely on photometric estimates of our redshifts. This is the reason why it is common to consider objects projected in two-dimensional shells encompassing a large redshift range, and writing the different statistics on the two-dimensional sphere.

The angular correlation function as a function of the separation between two points in the sky whose positions are given by $\mathbf{\Omega}_i = (\theta_i, \varphi_i)$ can be defined as

$$w(\theta) = w(\mathbf{\Omega}_1, \mathbf{\Omega}_2) \equiv \langle \tilde{\delta}_G(\mathbf{\Omega}_1)\tilde{\delta}_G(\mathbf{\Omega}_2)\rangle, \tag{2.75}$$

where θ is given by eq. (2.73).

The density contrast we measure with photometric redshift surveys is not $\delta_G(\boldsymbol{x}) = \delta_G(\mathbf{\Omega}, z)$, but rather its corresponding redshift-projected $\tilde{\delta}_G(\mathbf{\Omega})$. Therefore, they are related to each other by

$$\tilde{\delta}_G(\mathbf{\Omega}) = \int dz\, \phi(z)\delta_G(\mathbf{\Omega}, z), \tag{2.76}$$

where $\phi(z)$ is the radial selection function, which incorporates all the uncertainties in the redshift. In general, it can be expressed as

$$\phi(z) = \frac{dN_G}{dz}(z) \int dz_{\text{survey}} P(z|z_{\text{survey}}) W(z_{\text{survey}}), \tag{2.77}$$

where dN_G/dz is the galaxy redshift distribution and $W(z_{\text{survey}})$ is a window function that defines the redshift bin, usually a top-hat function,

$$W(z_{\text{survey}}) = \begin{cases} 1, & z_{\text{survey}} \in \text{bin}, \\ 0, & z_{\text{survey}} \notin \text{bin}. \end{cases} \tag{2.78}$$

Here, z_{survey} is the redshift (either spectroscopic or photometric) measured by the survey, and $P(z|z_{\text{survey}})$ gives the probability for the true redshift to be z when the redshift measured by the survey observations is z_{survey}. In the particular case of a spectroscopic redshift survey, the probability for z_{survey} to be the true redshift is one, which means

$$P(z|z_{\text{survey}}) = \delta(z - z_{\text{survey}}). \tag{2.79}$$

Therefore, for spectroscopic surveys we find that

$$\phi(z) = \frac{dN_G}{dz}(z) W(z). \tag{2.80}$$

In general, the two-point angular correlation function can be defined as an integral of the two-point correlation function, given by eq. (2.71), as

$$w(\theta) = \int dz_1\, \phi(z_1) \int dz_2\, \phi(z_2) \xi_G(r(z_1), r(z_2), \theta), \tag{2.81}$$

where the radial selection function $\phi(z)$ is given by eq. (2.77). It should be noted that the angular correlation function depends on the redshift since $\phi(z)$ is different for each redshift bin. In Chapters 4 and 5 we will use the previous expression to compute the theory template $w(\theta)$ from $\xi_G(z_1, z_2, \theta)$ in order to perform the BAO fits, both in simulations and data.

2.4.2 The Angular Power Spectrum

Similarly to how we defined the angular correlation function in the previous section, here we define the angular or projected power spectrum. Assuming our projected density contrast, $\tilde{\delta}_G(\mathbf{\Omega})$, lies on $\mathcal{S}^2$, it can be expanded in spherical harmonics,

$$\tilde{\delta}_G(\mathbf{\Omega}) = \sum_{\ell} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\mathbf{\Omega}). \tag{2.82}$$

Assuming the $a_{\ell m}$ modes are Gaussian random variables of mean zero and statistically independent, the angular power spectrum can be defined as

$$\langle a_{\ell m} a^*_{\ell' m'} \rangle = \langle |a_{\ell m}|^2 \rangle \delta_{\ell \ell'} \delta_{m m'} \equiv C_\ell \delta_{\ell \ell'} \delta_{m m'}. \tag{2.83}$$

Therefore,

$$C_\ell = \langle |a_{\ell m}|^2 \rangle. \tag{2.84}$$

The addition theorem states that

$$\sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{\Omega}_1) Y^*_{\ell m}(\mathbf{\Omega}_2) = \left(\frac{2\ell + 1}{4\pi} \right) \mathcal{L}_\ell(\cos\theta), \tag{2.85}$$

where $\mathcal{L}_\ell(\cos\theta)$ are the Legendre polynomials, with θ given by eq. (2.73). The angular power spectrum, C_ℓ, can, then, be easily related to the angular correlation function, $w(\theta)$,

$$w(\theta) = \langle \tilde{\delta}_G(\mathbf{\Omega}_1) \tilde{\delta}_G(\mathbf{\Omega}_2) \rangle = \sum_{\ell \ell'} \sum_{m m'} \langle a_{\ell m} a^*_{\ell' m'} \rangle Y_{\ell m}(\mathbf{\Omega}_1) Y^*_{\ell' m'}(\mathbf{\Omega}_2) \tag{2.86}$$

$$= \sum_{\ell} \left(\frac{2\ell + 1}{4\pi} \right) \mathcal{L}_\ell(\cos\theta) C_\ell. \tag{2.87}$$

Equivalently,

$$C_\ell = 2\pi \int_{-1}^{1} d(\cos\theta) \mathcal{L}_\ell(\cos\theta) w(\theta). \tag{2.88}$$

This expression will be particularly useful in Chapters 4 and 5.

2.4.3 Covariance for Projected Observables

In order to compute an analytical estimation of the covariance for the angular correlation function, it is necessary to obtain the one for the angular power spectrum first, and then compute its Fourier transform [29]. However, we have to firstly overcome the so-called "cosmic variance" problem. The expected values that have appeared previously, denoted by $\langle\rangle$, correspond to the ensemble average over infinite realizations of our Universe. Therefore, they are quantities that we do not have access to: it is necessary to define estimators for both $w(\theta)$ and C_ℓ. Assuming our survey observes the full sky,

$$\hat{C}_\ell \equiv \frac{1}{2\ell+1}\sum_{m=-\ell}^{\ell}|a_{\ell m}|^2, \tag{2.89}$$

where $\hat{C}_\ell$ is the estimator for C_ℓ, and

$$a_{\ell m} = \int d\mathbf{\Omega}\ \tilde{\delta}_G(\mathbf{\Omega})Y^*_{\ell m}(\mathbf{\Omega}). \tag{2.90}$$

Is is easy to check that $\langle\hat{C}_\ell\rangle = C_\ell$, which makes this estimator unbiased when computed over the full sky. The variance of $\hat{C}_\ell$ is given by

$$\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'}) = \langle\hat{C}_\ell\hat{C}_{\ell'}\rangle - \langle\hat{C}_\ell\rangle\langle\hat{C}_{\ell'}\rangle = \frac{2\delta_{\ell\ell'}}{2\ell+1}C_\ell^2. \tag{2.91}$$

In the case that our survey does not observe over the full sky, the Gaussian term for the covariance, i.e., the one that does not include higher-order moments of the density field, is given by

$$\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'}) = \frac{2\delta_{\ell\ell'}}{f_{\mathrm{sky}}(2\ell+1)}\left(C_\ell + \frac{1}{\bar{n}}\right)^2, \tag{2.92}$$

where f_{sky} is the fraction of the sky that our survey observed and $\bar{n}$ is the mean number density of galaxies.

Similarly, we now define our estimator for $w(\theta)$. Once the density field is pixelized of the sky, the standard estimator to measure the angular correlation function is given by [29]

$$\hat{w}(\theta) \equiv \frac{1}{N_{\mathrm{pair}}}\sum_{ij}\delta_i\delta_j, \tag{2.93}$$

where N_{pair} is the number of pixel pairs and

$$\delta_i = \frac{n_i - \bar{n}}{\bar{n}} \tag{2.94}$$

is the overdensity density field in the $i-$th pixel, with $\bar{n}$ being the mean number density of galaxies (in objects per steradian).

Before we continue, it is particularly useful to explicitly write the expression for the Landy-Szalay estimator [30]. The Landy-Szalay estimator is the one we use throughout this thesis to estimate the angular correlation function (on both simulations and data), and has the form

$$\hat{w}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \tag{2.95}$$

where $DD(\theta)$ and $RR(\theta)$ are the number of pairs of objects separated by an angular distance θ from the galaxy catalog and from the random catalog, respectively. $DR(\theta)$ corresponds to the number of galaxy-random pairs of objects separated by an angle θ. This estimator is unbiased and minimizes the variance to the Poisson level [30]. The Landy-Szalay estimator was the one used in DES during the Y3 analysis in order to measure the correlation function for galaxy clustering, and it is also the one being used for the Y6 analysis.

Following [31, 29], we find that the covariance for the estimator of $w(\theta)$ can be written as

$$\mathrm{cov}(\hat{w}(\theta), \hat{w}(\theta')) = \langle \hat{w}(\theta)\hat{w}(\theta')\rangle - \langle \hat{w}(\theta)\rangle\langle \hat{w}(\theta')\rangle \tag{2.96}$$

$$= \sum_{\ell\ell'} \frac{(2\ell+1)(2\ell'+1)}{(4\pi)^2} \mathcal{L}_\ell(\cos\theta)\mathcal{L}_{\ell'}(\cos\theta')\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'}). \tag{2.97}$$

For the Gaussian case, i.e., using eq. (2.92), we find

$$\mathrm{cov}(\hat{w}(\theta), \hat{w}(\theta')) = \frac{2}{f_{\mathrm{sky}}} \sum_\ell \frac{2\ell+1}{(4\pi)^2} \mathcal{L}_\ell(\cos\theta)\mathcal{L}_\ell(\cos\theta') \left(C_\ell + \frac{1}{\bar{n}}\right)^2. \tag{2.98}$$

The previous expression will be particularly useful in Chapters 4 and 5, since it allows to compute analytical covariance matrices in order to perform the BAO fits.

2.5 Baryon Acoustic Oscillations

The Universe has been expanding, which implies that, billions of years ago, it was much smaller, denser and hotter than it is today. In this early Universe, electrons interacted with photons via the Compton effect, and also with protons via Coulomb scattering. Therefore, the three components behaved as a mixed fluid, i.e., they were in equilibrium due to the gravity exerted by baryons and the radiation pressure exerted by photons, and they oscillated as sound modes. These oscillations are known as baryon acoustic oscillations (BAO).

2.5.1 The Physics of BAO

Let's imagine an overdense region of the primordial plasma. This overdense region contained dark matter, baryons and photons. The pressure results in spherical sound waves of both baryons and photons moving outwards from the origin of the overdensity. The dark matter interacts only gravitationally, and so it stays at the center of the sound wave. Before decoupling, photons and baryons moved outwards together. After decoupling, photons were no longer interacting with baryonic matter and, therefore, they diffused away. That relieved the pressure on the system, leaving behind shells of baryonic matter. Out of all those shells, representing different sound waves' wavelengths, the resonant shell corresponds to the first one, since it is the one that travels the same distance for all overdensities before decoupling. This radius is given eq. (1.1) evaluated at the photon decoupling epoch

$$r_d \equiv r_s(z_d) = \int_{z_d}^{\infty} dz \frac{c_s(z)}{H(z)}. \tag{2.99}$$

The speed of sound, $c_s(z)$, is given by

$$c_s(z) = c\left[3(1+R(z))\right]^{-1/2}, \tag{2.100}$$

where

$$R(z) = \frac{1+w_b}{1+w_\gamma}\left[\frac{\rho_b(z)}{\rho_\gamma(z)}\right] = \frac{3}{4}\left(\frac{\Omega_b}{\Omega_\gamma}\right)\frac{1}{1+z}. \tag{2.101}$$

Without the radiation pressure driving the system outwards, the only remaining force on the baryons was gravitational. Therefore, baryons and dark matter (which was left behind at the center of the perturbation) formed a configuration which included overdensities of matter both at the origin of the anisotropy and in the shell located at the sound horizon. In Fig. 2.3 we show snapshots of an N-body simulation representing the evolution of the mass profile of an overdense region in the early Universe.

These anisotropies became the ripples in matter density that eventually formed galaxies. Therefore, one would expect to see a larger number of galaxy pairs separated by the sound horizon scale than by any other length scale, besides small scales. This particular configuration of matter occurred at each overdense region in the early Universe, which means that our Universe is composed of many overlapping ripples. It is not possible to observe this preferred separation of galaxies by eye, but one can statistically measure this artifact by computing the two-point correlation function of galaxies, which we already introduced in 2.4.

Since the BAO produces an accumulation of galaxies at the sound horizon scale, this translates into a peak in the correlation function at this scale, which is fixed at all epochs (in comoving coordinates). If we are able to observe this peak at different epochs in the Universe, we can map the expansion history by measuring the distance to the galaxy distribution, knowing the angles and redshifts, and also infer its matter–energy content. The critical issue here is our understanding of how exactly the

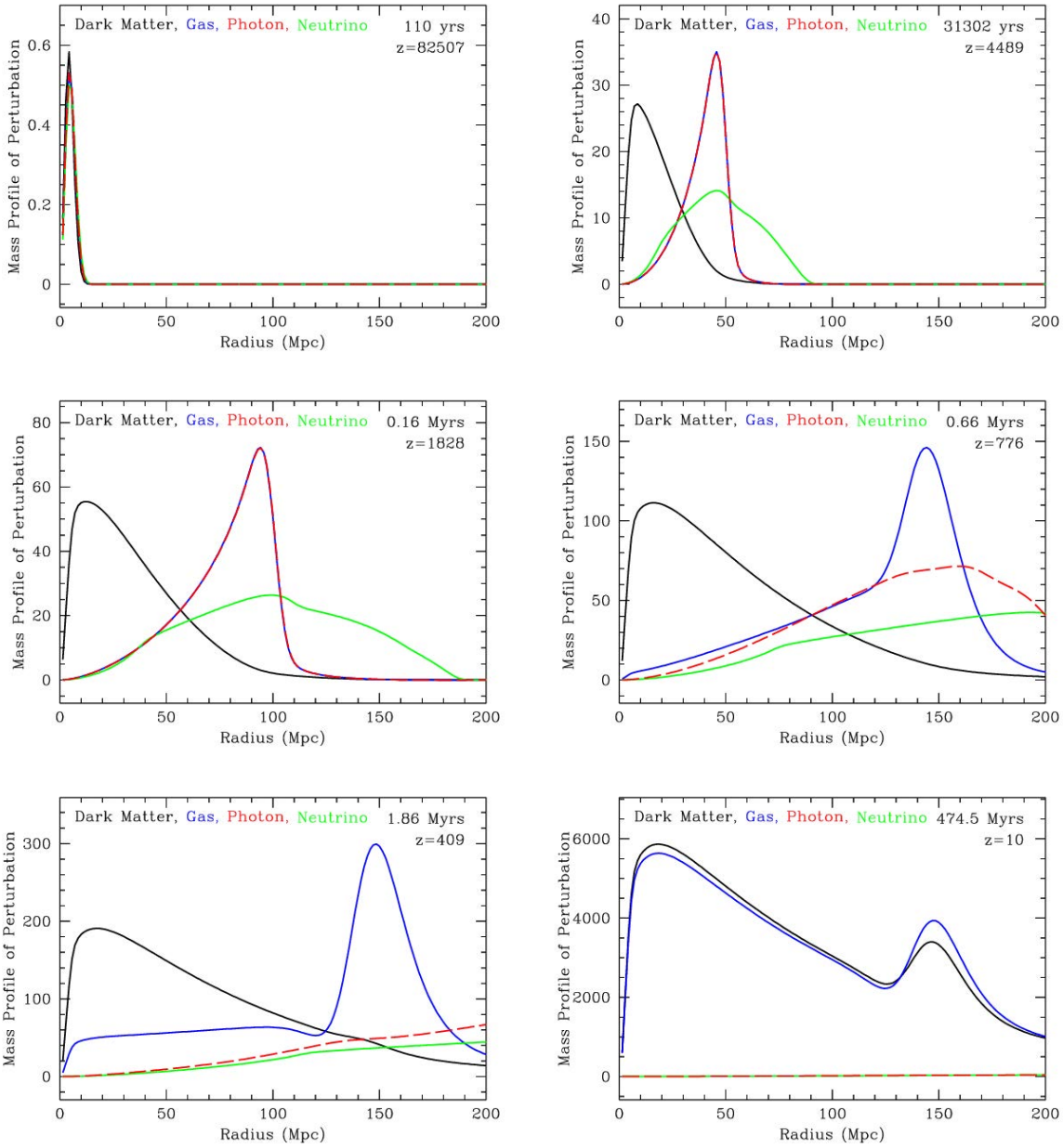

Figure 2.3: Evolution of the mass profile of an overdense region in the early Universe. At the very beginning, all components share the same mass profile (top-left). Photons start dragging baryons away (top-right and middle-left) until the decoupling era (middle-right). At that point, neutrinos and photons start travelling freely and the mass profile of baryons freezes (bottom-left). Finally, baryons fall into the potential wells of dark matter (bottom-right). Credits: https://scholar.harvard.edu/deisenstein/book/export/html/28634.

visible galaxies trace the dark-matter distribution. We can only observe the visible light, but we know that galaxies formed in the high density regions of dark-matter halos. Therefore using galaxies to trace the underlying dark matter introduces an inherent bias, by over-weighting the overdense regions and under-weighting the underdense regions, as we already explained in 2.3.3. Therefore, to map the correlation function at different epochs, we need to select galaxy samples at different redshifts and compute the correlation function for each sample. In this way, we map the evolution of LSS and BAO.

2.5.2 The BAO in the Galaxy Distribution

After recombination, baryons become non-relativistic. The baryon perturbation then begins to grow at their locations and interact with the dark matter perturbation, as shown in Fig. 2.3. Therefore, the BAO should be imprinted on the large-scale structure of the Universe. Theoretically, the BAO should produce an overdensity at about ∼150 Mpc. However, this is not easy to study because the observation of the BAO signal requires a number of tracers of the overdensity field of matter large enough to exceed the cosmic variance, and also an area of the sky large enough to map separations of ∼150 Mpc between galaxies (unlike the CMB, large galaxy surveys observe in all three dimensions of space.). In 2005, the BAO signal in the galaxy distribution was detected by two groups almost simultaneously: one using SDSS [32] and the other using the 2dF Galaxy Redshift Survey (2dFGRS) [33]. In Fig. 2.4 we show the large-scale two-point correlation function as measured by SDSS (first ever detection, see [32]) in which the BAO peak can be clearly observed at around $100/h$ Mpc.

What makes the BAO particularly useful for cosmology is the fact that it represents a "standard ruler"[6], which allows us to accurately measure the geometry of the Universe. A standard ruler is an astronomical object or phenomenon whose true size is known (in this case, the sound horizon scale), which allows $d_A(z)$ and $H(z)$ to be measured. By measuring the angle $d\theta$ subtended by this ruler of size $dr_\perp$ (perpendicular to the LOS) as a function of redshift, we map out the angular diameter distance,

$$dr_\perp = d_A(z)d\theta. \tag{2.102}$$

On the other hand, by measuring the redshift interval dz associated with this ruler of size $dr_\parallel$ (along the LOS) as a function of redshift, we map out the Hubble parameter,

$$dr_\parallel = cH^{-1}(z)dz = d_H(z)dz. \tag{2.103}$$

In principle, $dr_\perp = dr_\parallel$ because the Universe is isotropic, but there are situations in

[6] Standard rulers should not be mistaken with standard candles, which are astronomical objects whose absolute magnitude is known, and for which $d_L(z)$ can be obtained via eq. (1.87) by measuring their apparent magnitude. The BAO standard ruler is particular in the sense that it is actually a comoving standard ruler.

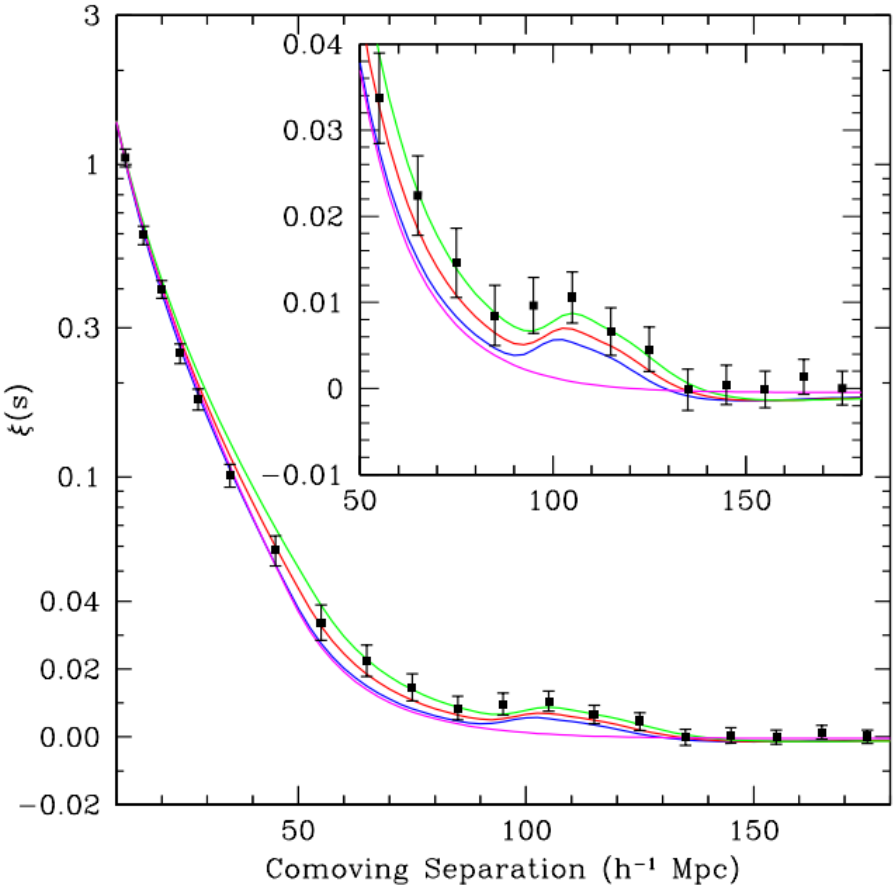

Figure 2.4: The two-point correlation function of the SDSS LRG sample (plot taken from [32]). The black points represent the measured correlation function, whose error-bars come from the covariance matrix estimated from mock catalogs. The red, green and blue solid lines represent models with different sets of cosmological parameters. The magenta one represents a model with no BAO feature (a pure CDM model).

which the BAO signal is anisotropic. There are two sources of anisotropies: the so-called redshift-space distortions and the Alcock-Paczynski effect [34]. We will explain these effects in section 2.5.3.

As we have already mentioned, the scale of the BAO perpendicular to the LOS depends on $d_A(z)$, whereas the scale of the BAO along the LOS depends on $d_H(z)$. Therefore, the BAO scale probed by the correlation function has a dependence on the cosmology given by

$$d_V(z) \equiv \left[z d_M^2(z) d_H(z)\right]^{1/3}, \tag{2.104}$$

with $d_M(z) = (1+z)d_A(z)$. The powers of $d_M(z)$ and $d_H(z)$ in the previous expression account for the fact that we have two perpendicular directions and one parallel direction with respect to the LOS. Spectroscopic redshift surveys almost always specify the transverse and radial BAO as two independent measurements with correlated errors [35], whereas photometric redshift surveys only specify the transverse BAO measurement. The BAO scale obtained from the high redshift CMB anisotropies should be the same as the one obtained from the low redshift distribution of galaxies.

2.5.3 Anisotropic BAO

There are two main sources for anisotropies in the clustering signal: the redshift-space distortions and the Alcock-Paczynski effect. Here we briefly describe both of them.

2.5.3.1 Redshift-Space Distortions and Reconstruction

Driven by dark matter overdensities, peculiar velocities of galaxies induce redshift-space distortions (RSD) in galaxy clustering [36], which yield to an anisotropic clustering signal. When making a 3D map of the Universe, the third coordinate (radial distance) is typically obtained from the redshift using Hubble's law or its generalization. However, the measured redshift of a galaxy comes from a superposition of its Hubble flow recession velocity due to the expansion of the Universe, $\boldsymbol{v}_r$, and its peculiar velocity, $\boldsymbol{v}_p$. Therefore, the measured redshift-space position $\boldsymbol{r}$ (found by transforming the observed galaxy redshift to a distance using the reference cosmology) is displaced from the real-space position $\boldsymbol{x}$ by its peculiar velocity projected along the LOS, i.e., $\boldsymbol{v}_p \cdot \hat{\boldsymbol{x}}$, where $\hat{\boldsymbol{x}}$ is the unit vector in LOS direction. In general, the mapping from $\boldsymbol{x}$ to $\boldsymbol{r}$ is given by

$$\boldsymbol{r}(\boldsymbol{x}) = \boldsymbol{x} + \frac{\left[\boldsymbol{v}_p(\boldsymbol{x}) \cdot \hat{\boldsymbol{x}}\right] \hat{\boldsymbol{x}}}{aH(a)}. \tag{2.105}$$

A particular manifestation of redshift-space distortions is the Fingers-of-God effect [37], in which the galaxy distribution appears elongated and distorted along the line of sight due to the random, virialized motions of galaxies within massive galaxy clusters. This effect is only observed in redshift space, and it results in "fingers" or elongated structures pointing towards the center of the cluster, which can make it difficult to accurately measure the clustering properties of galaxies in these regions.

In a Lagrangian perturbation theory framework, the real-space or Eulerian position $\boldsymbol{x}$ of a particle can be written as

$$\boldsymbol{x}(\boldsymbol{q}, t) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q}, t), \tag{2.106}$$

where $\boldsymbol{q}$ is the initial position and $\boldsymbol{\Psi}$ is the displacement field. At first order, the peculiar velocity field can be expressed as

$$\boldsymbol{v}_p(\boldsymbol{x}(\boldsymbol{q}, t)) = aHf\boldsymbol{\Psi}(\boldsymbol{q}, t), \tag{2.107}$$

where

$$f \equiv \frac{d\log D(a)}{d\log a} \tag{2.108}$$

is the logarithmic growth rate, which depends on the scale-independent linear growth factor $D(a)$. The logarithmic growth rate can be written in terms of the matter density parameter as

$$f = \left[\Omega_m(a)\right]^{\gamma}, \tag{2.109}$$

where γ is the growth rate exponent, which has an approximate value of $\gamma = 0.55$ for general relativity and $w = -1$.

Combining the previous expressions and using the fact that $\hat{\boldsymbol{x}} = \hat{\boldsymbol{r}}$, we find that

$$\boldsymbol{q}(\boldsymbol{r}) = \boldsymbol{r} - \boldsymbol{\Psi} - f(\boldsymbol{\Psi} \cdot \hat{\boldsymbol{r}})\hat{\boldsymbol{r}}. \tag{2.110}$$

Using the previous expression together with the fact that the number of galaxies is conserved when transforming redshift-space to real-space coordinates, we find that

$$\boldsymbol{\nabla} \cdot \boldsymbol{\Psi} + \frac{f}{b}\boldsymbol{\nabla}(\boldsymbol{\Psi} \cdot \hat{\boldsymbol{r}})\hat{\boldsymbol{r}} = -\frac{\delta_G^{\mathrm{red}}}{b}, \tag{2.111}$$

where δ_G^{red} is the smoothed redshift-space galaxy overdensity field and b is the linear galaxy bias. Solving the previous equation to obtain $\boldsymbol{\Psi}$ is the basis of the so-called density-field reconstruction procedure, for which reference values of b and f are used. Galaxies can then be moved back to their initial Lagrangian positions $\boldsymbol{q}$ removing non-linear effects, which sharpens the BAO feature and increases its detection significance. This technique is widely used in spectroscopic redshift surveys such as DESI, since it requires an accurate measurement of the three-dimensional position of galaxies, which implies an accurate measurement of redshift. For photometric redshift surveys such as DES, density-field reconstruction cannot be applied.

Kaiser Formula for the Matter Power Spectrum

It is also particularly useful to study how redshift-space distortions modify the power spectrum, which is no longer isotropic because of them. A correction must be

applied to $P_m(k)$ in order to account for these RSD, and at very good approximation it is given by the Kaiser formula [36],

$$P_m(k) \to P_m(\boldsymbol{k}) = P_m(k,\mu) = (1+f\mu^2)^2 P_m(k), \tag{2.112}$$

where μ is the cosine of the angle of the wave vector $\boldsymbol{k}$ with respect to the LOS. The function f that appears in the previous formula is given by

$$f = \frac{d\log D}{d\log a}, \tag{2.113}$$

where D is the growth factor, which we already defined in eq. (2.43). In order to be able to compute $P_m(k,\mu)$, it is convenient to expand in Legendre polynomials,

$$P_m(k,\mu) = \sum_\ell P_{m,\ell}(k)\mathcal{L}_\ell(\mu). \tag{2.114}$$

Solving for $P_{m,\ell}(k)$, we find that

$$\begin{aligned} P_{m,\ell}(k) &= \frac{2\ell+1}{2}\int_{-1}^{1} d\mu P_m(k,\mu)\mathcal{L}_\ell(\mu) \\ &= \frac{2\ell+1}{2}\left[\int_{-1}^{1} d\mu (1+f\mu^2)^2\mathcal{L}_\ell(\mu)\right] P_m(k). \end{aligned} \tag{2.115}$$

The only non-zero multipoles are the monopole, the quadrupole and the hexadecapole,

$$\begin{cases} P_{m,0}(k) = \left(1+\frac{2}{3}f+\frac{1}{5}f^2\right)P_m(k), \\ P_{m,2}(k) = \left(\frac{4}{3}f+\frac{4}{7}f^2\right)P_m(k), \\ P_{m,4}(k) = \frac{8}{35}f^2 P_m(k). \end{cases} \tag{2.116}$$

On the other hand, the galactic power spectrum can be written as

$$P_G(k,\mu) = (b+f\mu^2)^2 P_m(k) = b^2(1+\beta\mu^2)^2 P_m(k), \tag{2.117}$$

where we defined

$$\beta = \frac{f}{b}. \tag{2.118}$$

Therefore, the multipoles for the galactic power spectrum are given by

$$\begin{cases} P_{G,0}(k) = \left(1+\frac{2}{3}\beta+\frac{1}{5}\beta^2\right)b^2 P_m(k), \\ P_{G,2}(k) = \left(\frac{4}{3}\beta+\frac{4}{7}\beta^2\right)b^2 P_m(k), \\ P_{G,4}(k) = \frac{8}{35}\beta^2 b^2 P_m(k). \end{cases} \tag{2.119}$$

Finally, the matter two-point correlation function can be computed as

$$\xi_m(s,\mu) = \sum_\ell \xi_{m,\ell}(s)\mathcal{L}_\ell(\mu), \tag{2.120}$$

where

$$\xi_{m,\ell}(s) = i^\ell \int dk\, j_\ell(ks)\frac{k^2}{2\pi^2}P_{m,\ell}(k). \tag{2.121}$$

Analogously, the galactic correlation function, $\xi_G(s,\mu)$, can be obtained replacing $P_{m,\ell}(k)$ with $P_{G,\ell}(k)$. In the absence of redshift-space distortions, i.e., if $f = 0$, the only non-zero multipole is the monopole, and we recover the same expression for the correlation function that we deduced in eq. (2.68). As we did in section 2.4, it is particularly useful to write s and μ as a function of z_1, z_2 and θ. We already computed the expression for s, which is simply given by $s^2 = r_1^2 + r_2^2 - 2r_1r_2\cos\theta$. On the other hand, the expression for μ can be written as

$$\mu(z_1,z_2,\theta) = \frac{r(z_1) - r(z_2)}{s(z_1,z_2,\theta)}. \tag{2.122}$$

Therefore, the galactic correlation function can be explicitly written as

$$\xi_G(z_1,z_2,\theta) = \sum_\ell \xi_{G,\ell}\left(\sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos\theta}\right)\mathcal{L}_\ell\left(\frac{r(z_1) - r(z_2)}{\sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos\theta}}\right). \tag{2.123}$$

2.5.3.2 Alcock-Paczynski Effect and Scaling Parameters

In order to extract as much cosmological information as possible, we need to measure the correlation function as a function of the perpendicular and the parallel separations between galaxies ($r_\perp$ and $r_\parallel$) with respect to the LOS, i.e., $\xi(s,\mu) = \xi(r_\perp, r_\parallel)$ [38]. In the previous section we derived that

$$\xi_G(s,\mu) = \sum_\ell \xi_{G,\ell}(s)\mathcal{L}_\ell(\mu), \tag{2.124}$$

where s is the observed separation between two galaxies and μ is the observed cosine of the angle that the pair of galaxies makes with respect to the LOS,

$$\mu = \frac{r_\parallel}{s}. \tag{2.125}$$

In addition to the above equation, it must be satisfied that $s^2 = r_\perp^2 + r_\parallel^2$. Therefore,

$$r_\perp = s\sqrt{1-\mu^2}, \quad r_\parallel = s\mu. \tag{2.126}$$

The observed separation between the two galaxies, s, is computed from their positions in the sky and their redshift assuming a fiducial or reference set of cosmological

parameters, $\mathbf{\Omega}_{\rm ref}$. This reference cosmology is necessary to transform redshifts into distances. The deviation of the reference cosmology ($\mathbf{\Omega}_{\rm ref}$) from the underlying cosmology ($\mathbf{\Omega}$) leads to an anisotropic clustering signal known as Alcock-Paczynski effect [34]. This effect, together with the variation of the acoustic scale with respect to the fiducial template, is usually accounted through scaling $r_\perp$ and $r_\parallel$ as

$$r_\perp \to r'_\perp = \alpha_\perp r_\perp, \quad r_\parallel \to r'_\parallel = \alpha_\parallel r_\parallel, \tag{2.127}$$

where the scaling parameters are defined as

$$\alpha_\perp(z_{\rm eff}) = \left[\frac{d_M(z_{\rm eff})}{r_d}\right]\left[\frac{d_M^{\rm ref}(z_{\rm eff})}{r_d^{\rm ref}}\right]^{-1} \tag{2.128}$$

and

$$\alpha_\parallel(z_{\rm eff}) = \left[H(z_{\rm eff})r_d\right]^{-1}\left[H^{\rm ref}(z_{\rm eff})r_d^{\rm ref}\right], \tag{2.129}$$

respectively (see [38] for reference). The redshift $z_{\rm eff}$ is the effective redshift of the sample of galaxies we use to measure these scaling parameters. The true separation between galaxies can then be computed as

$$\begin{aligned} s'(s,\mu) &= \sqrt{r'^2_\perp + r'^2_\parallel} = \sqrt{\alpha_\perp^2 r_\perp^2 + \alpha_\parallel^2 r_\parallel^2} \\ &= \sqrt{\alpha_\perp^2 s^2(1-\mu^2) + \alpha_\parallel^2 s^2\mu^2} = s\sqrt{(1-\mu^2)\alpha_\perp^2 + \mu^2\alpha_\parallel^2}, \end{aligned} \tag{2.130}$$

where we can define

$$\alpha(\mu) \equiv \sqrt{(1-\mu^2)\alpha_\perp^2 + \mu^2\alpha_\parallel^2} \tag{2.131}$$

so that $s' = \alpha s$. On the other hand, the true cosine of the angle with respect to the LOS is given by

$$\mu'(\mu) = \frac{r'_\parallel}{s'(s,\mu)} = \frac{\alpha_\parallel \mu}{\sqrt{(1-\mu^2)\alpha_\perp^2 + \mu^2\alpha_\parallel^2}}. \tag{2.132}$$

For the power spectrum we proceed similarly to what we did in the correlation function case. We need to measure the power spectrum as a function of $k_\perp$ and $k_\parallel$, i.e., $P_G(k,\mu) = P_G(k_\perp, k_\parallel)$, with

$$P_G(k,\mu) = \sum_\ell P_{G,\ell}\mathcal{L}_\ell(\mu), \tag{2.133}$$

where k is the observed wave number and μ is the observed cosine of the angle of $\boldsymbol{k}$ with respect to the LOS,

$$\mu = \frac{k_\parallel}{k}. \tag{2.134}$$

It must be satisfied that $k^2 = k_\perp^2 + k_\parallel^2$, which implies

$$k_\perp = k\sqrt{1-\mu^2}, \quad k_\parallel = k\mu. \tag{2.135}$$

The scaling of $k_\perp$ and $k_\parallel$ because of the Alcock-Paczynski effect is given by

$$k_\perp \to k'_\perp = k_\perp/\alpha_\perp, \quad k_\parallel \to k'_\parallel = k_\parallel/\alpha_\parallel, \tag{2.136}$$

and the combination of both yields

$$k'(k,\mu) = \sqrt{k'^2_\perp + k'^2_\parallel} = \frac{k}{\alpha_\perp}\sqrt{1+\mu^2\left(\frac{\alpha^2_\perp}{\alpha^2_\parallel}-1\right)}. \tag{2.137}$$

On the other hand, the true cosine of the angle with respect to the LOS is given by

$$\mu'(\mu) = \frac{k'_\parallel}{k'(k,\mu)} = \frac{\mu}{\frac{\alpha_\parallel}{\alpha_\perp}\sqrt{1+\mu^2\left(\frac{\alpha^2_\perp}{\alpha^2_\parallel}-1\right)}}. \tag{2.138}$$

For the isotropic BAO case we usually refer to the scaling parameter as $\alpha_{\rm iso}$ [39], which can be written as

$$\alpha_{\rm iso}(z_{\rm eff}) = \left[\frac{d_V(z_{\rm eff})}{r_d}\right]\left[\frac{d_V^{\rm ref}(z_{\rm eff})}{r_d^{\rm ref}}\right]^{-1}. \tag{2.139}$$

This is the scaling parameter we would obtain if we only took into account the monopole term ($\ell = 0$) of the correlation function. The monopole is an average over all directions without weighting as a function of μ, i.e., every direction has the same weight [39]. Therefore, $\alpha_{\rm iso}$ is related to the perpendicular and the parallel components of α at the same time. Taking a look at the definition of $d_V(z)$, eq. (2.104), we easily find

$$\alpha_{\rm iso} = \left(\alpha_\parallel \alpha^2_\perp\right)^{1/3}. \tag{2.140}$$

2.6 State of the Art of the BAO

The accuracy of BAO measurements has steadily improved over the years, thanks to advances in observational techniques and data analysis. In 2005, the original SDSS (SDSS-I and II) made the first ever BAO measurement [32], using a sample of 46,748 luminous red galaxies (known as the "main galaxy sample", or MGS), almost simultaneously to the detection by the 2dFGRS [33]. These detections provided one of the first direct evidences for the existence of dark energy. In 2011, the 6dF Galaxy Survey (6dFGS) measured the BAO scale at low redshifts ($z \sim 0.1$) [40] by studying the distribution of galaxies in the nearby Universe. Other surveys, such as the WiggleZ Dark Energy Survey [41], also made significant contributions to the study of BAO over this time period. Over the next decade, SDSS continued to make more BAO measurements using larger and larger galaxy samples, with progressively higher

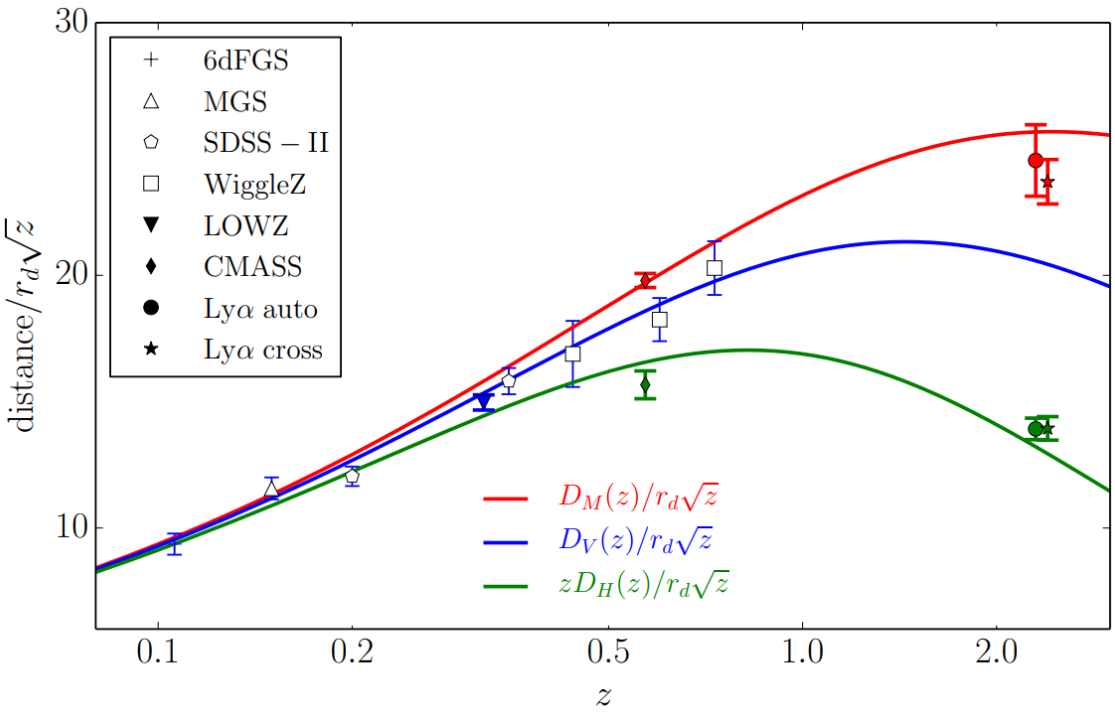

Figure 2.5: The BAO Hubble diagram constructed from several detections. Blue, red and green points show measurements of $d_V(z)/r_d$, $d_M(z)/r_d$ and $zd_H(z)/r_d$, respectively, and the colored lines represent the fiducial Planck ΛCDM model (with $\Omega_m = 0.3183$, $h = 0.6704$). Plot taken from [48].

precision, which provided important constraints on different cosmological parameters. The SDSS-III's Baryon Oscillation Spectroscopic Survey (BOSS) [42, 43] and the completed SDSS-III's BOSS [44, 45, 46] achieved a 1% distance measurement in two independent redshift bins (LOWZ and CMASS galaxy samples[7]). These surveys used different techniques and observed different portions of the sky, which helped to provide complementary measurements of the large-scale structure of the Universe. The BAO has also been studied in other cosmic tracers, such as the Lyman-α forest of quasars [47]. These observations, combined with BAO measurements from galaxy surveys, have allowed for more accurate constraints on cosmological parameters and the evolution of the Universe over time. In Fig. 2.5 we show the BAO Hubble diagram constructed from several detections, as taken from [48].

In recent years, new surveys such as DES and the Taipan Galaxy Survey [49] have continued to make important BAO measurements using even larger and more precise datasets. These measurements have provided new constraints on the nature of dark energy and the history of the Universe, and have opened up new avenues of research

[7] LOWZ stands for low-redshift sample, whereas CMASS stands for constant-mass sample

in the study of cosmology. The latest BAO measurements from DES [50], SDSS-IV's extended Baryon Oscillation Spectroscopic Survey (eBOSS) [51, 38, 52, 53, 54, 47, 35], and other surveys have reached accuracies of around 1-2%, allowing for highly precise constraints on the properties of dark energy and several cosmological parameters.

The DES Y6 BAO analysis, i.e., the final BAO analysis of DES, is still ongoing, and its results will be the highest-precision measurements from a photometric-redshift survey ever. We expect an accuracy of about 2%, with a red-galaxy sample in the redshift range $0.6 < z < 1.2$. On the other hand, the DESI Y1 BAO analysis is also ongoing, and its results will be released by the end of 2023. We expect a combined accuracy of about 0.5% for the DESI Y1 results, already smaller than that of eBOSS. The European Space Agency's Euclid mission [55], which is set to launch in the mid-2020s, is also expected to provide precise measurements of BAO and improve our understanding of the nature of dark energy. Overall, the BAO probe has played an important role in cosmology over the last few decades and continues to be a key tool for understanding the Universe.

Chapter 3

Large Galaxy Surveys For Cosmology

For nearly half a century, photometric (or imaging) and spectroscopic galaxy surveys have given us information about the contents of the Universe. We obtain this cosmological information statistically analyzing huge amounts of galaxies, e.g., through the two-point correlation function. Current galaxy surveys allow us to study massive numbers of galaxies, which in turn have made cosmology grow as a field. These large amounts of scientific information have to be stored and analyzed, which has taken cosmology to the big data era.

As we already mentioned, there are mainly two types of galaxy surveys: photometric and spectroscopic. Photometric galaxy surveys represent a way of covering huge areas of the sky, since they make it possible to obtain information of thousands of objects at the same time. Therefore, the statistical power of these surveys is huge, making it possible to obtain reliable cosmological information from the statistical analysis of their data. Photometric galaxy surveys measure the fluxes of the observed objects through ultraviolet, optical or infrared filters by means of images taken with charge-coupled devices (CCDs). These filters can be either broad of narrow-band filters, and they allow the pass of light only if it is within a certain wavelength range. Therefore, photometry does not allow us to measure the redshift of these objects because of the lack of spectral features, which means we cannot access the fully 3D positions of these objects. This is the reason why in photometric surveys it is necessary the measurement of photometric redshifts, which are an estimation of the redshift of galaxies using photometry. However, photometric surveys have the advantage to provide information about the shapes of the observed objects, which is essential for WL analyses. Some examples of photometric galaxy surveys are the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) [56], the Kilo-Degree Survey (KiDS) [57] the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) [58] and the Dark Energy Survey (DES) [59].

On the other hand, spectroscopic galaxy surveys measure the redshift of galaxies through their spectra, making it possible to obtain accurate 3D information for these

galaxies. This is possible because they are sensitive to some kind of spectroscopic feature, such as emission or absorption lines. However, these spectroscopic observations are very time consuming, which makes complicated the balance between obtaining the spectra and covering wide areas of the sky. Some examples of past spectroscopic galaxy surveys are the Sloan Digital Sky Survey (SDSS) [60], the Baryon Oscillation Spectroscopic Survey (BOSS) [42], the extended Baryon Oscillation Spectroscopic Survey (eBOSS) [61] and the 2dF Galaxy Redshift Survey (2dFGRS) [62]. Ongoing (and certainly future) spectroscopic galaxy surveys such as the Dark Energy Spectroscopic Instrument (DESI) [63] represent a step forward in obtaining a balance between the measurement of the spectra of galaxies and covering wide areas of the sky.

In this chapter we briefly describe the two experiments in which we were involved during the development of this thesis: the Dark Energy Survey and the Dark Energy Spectroscopic Instrument.

3.1 The Dark Energy Survey

The Dark Energy Survey (DES) is a visible and near-infrared survey that aims to probe the physical nature of dark energy with several methods. Among them, the dynamics of the expansion of the Universe and the growth of large-scale structure. In order to achieve this, it has imaged about 5,000 square degrees of the southern sky in a 6-year photometric survey from Cerro Tololo (Chile).

Organizationally, the Dark Energy Survey is an international collaboration composed of more than 650 scientists from more than 20 institutions in the United States, Spain, the United Kingdom, Brazil, Germany, Switzerland and Australia. The collaboration built and used an extremely sensitive 570-megapixel digital camera, the Dark Energy Camera (DECam) [64], mounted on the Víctor M. Blanco Telescope at Cerro Tololo Inter-American Observatory, high in the Chilean Andes. Over six years (2013-2019), the DES collaboration used 758 nights of observation to carry out a deep, wide-area survey to record information from ∼600 million galaxies that are billions of light-years from Earth. A fraction of the survey time was used to observe smaller patches of sky roughly once a week to discover and study thousands of supernovae and other astrophysical transients.

As an instrument, the Dark Energy Survey is a Stage-III photometric galaxy survey. In Fig. 3.1 we show the DES footprint, which covers, approximately, 5,000 deg^2. The final shape of the footprint is a consequence of avoiding the Milky Way galactic plane and trying to overlap with the South Pole Telescope and other important telescopes, while covering the largest possible area to study LSS in a fully-connected region to allow full calibration of photometry.

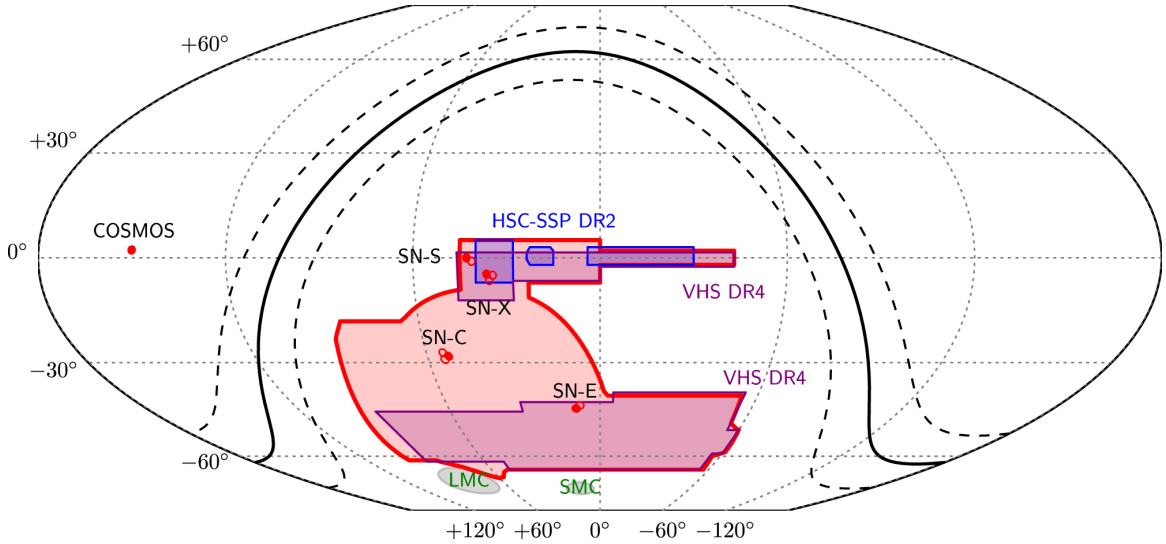

Figure 3.1: The DES footprint in equatorial coordinates, including the Wide Survey, the supernovae Survey fields (labeled as "SN", see [65]), the COSMOS field and the overlapping regions of HSC-SSP DR2 [66] and VHS DR4 [67]. Image taken from [68].

3.1.1 Scientific Goals

The primary mission of DES is to try to unveil the nature of dark energy. To achieve this mission, DES studies a variety of cosmological probes, whose combination allow us to probe the large-scale structure of the Universe at different distances and times, and to study the properties of dark energy and dark matter with unprecedented precision. The main 4 cosmological probes that DES studies are the number of clusters as a function of redshift, the WL effect in the distribution of galaxies, the BAO measurement and the Hubble diagram of type Ia supernovae. Combining these 4 cosmological probes, which were explained in more detail in section 1.1.1.1, DES aims to measure the dark energy equation of state parameter, w_0, with a precision better than 5%, and also its time evolution, w_a, which is given by

$$p = [w_0 + (1 - a)w_a]\,\rho, \tag{3.1}$$

with a precision around 30%.

3.1.2 The Instrument: the DECam

The DECam is the main instrument of DES [64]. This camera has five main systems: a 570 megapixel CCD imager with a cooling system, a low-noise electronic readout system, a wide-field optical corrector, a system of five filters (g, r, i, z, Y or $grizY$) in the optical and near-infrared range and a hexapod adjustor to provide real-time focus and alignment system to maintain high image quality. In Fig. 3.2 it is shown the standard bandpasses for the DECam $grizY$ filters.

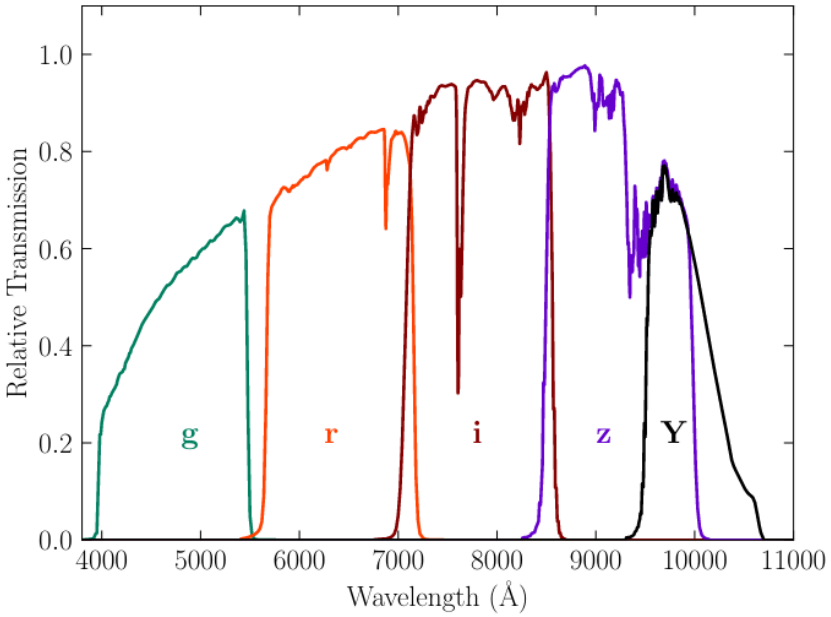

Figure 3.2: Standard bandpasses for the DECam *grizY* filters. The bandpasses represent the total system throughput, including atmospheric transmission (airmass=1.2) and the average instrumental response across the science CCDs.

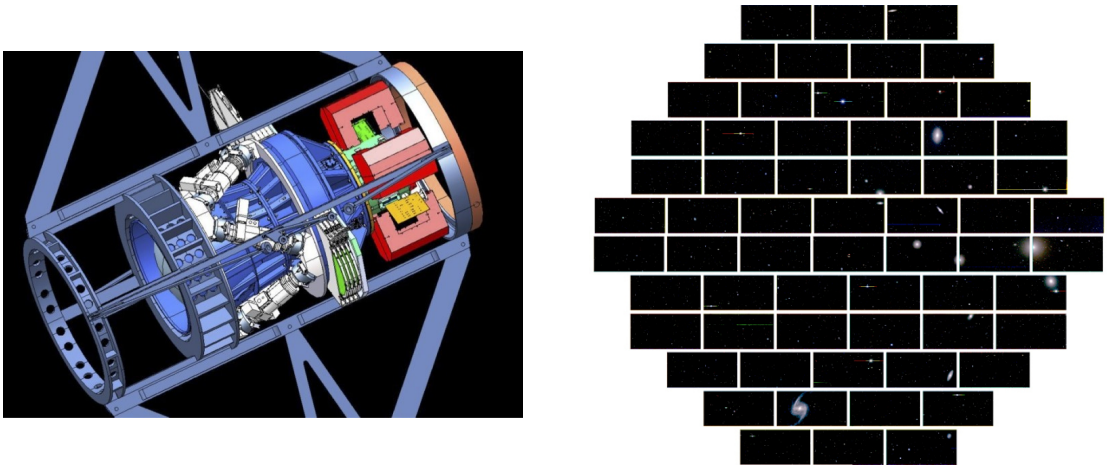

Figure 3.3: **Left panel:** Model of the DECam showing the most important components: Optics cage (blue rings at lower left), hexapod (white, diagonally oriented tubular structures near center), shutter (white, flat element near top right and above hexapod), filters (green, to right of hexapod), imager (between red boxes at right). **Right panel:** Image from the First Light of the DECam, where CCDs are apparent (https://www.darkenergysurvey.org/the-des-project/instrument/).

The focal plane of the DECam includes CCDs designed by engineers from the Lawrence Berkeley National Laboratory (LBNL). These CCDs were specifically designed to be more sensitive to red light from distant galaxies. The DECam focal plane consists of an array of 62 $2,048 \times 4,096$ CCDs placed in the shape of a hexagon, covering 3 deg^2. Besides these 62 CCDs, there are also 4 $2,048 \times 2,048$ guider CCDs and 8 $2,048 \times 2,048$ focus and alignment CCDs (all of them placed in the border of the focal plane). The quantum efficiency of these CCDs was optimized to be $> 90\%$ at 900 nm and over 60% in the 400-1,000 nm range. In Fig. 3.3 we show a model of the camera (left panel) and also an image from its first light (right panel).

The electronic system recorded the data in the form of a multi-extension fits file and also provided real-time quality and instrument health checks. Its design allowed the images to be read in less time than it took for the telescope to move to the next observation position.

3.1.3 The Víctor M. Blanco 4-Meter Telescope

The DECam was mounted on the Víctor M. Blanco Telescope (at a height of 2,200 meters). The Víctor M. Blanco Telescope, also known as Blanco 4m, is a 4-meter aperture telescope located at 460 km north of Santiago de Chile, and at 80 km away from La Serena. The telescope is part of the Cerro Tololo Inter-American Observatory (CTIO), which is operated by the Association of Universities for Research in Astronomy (AURA) and was funded by the U.S. National Science Foundation. This telescope hosted in the past some of the observations of type Ia supernovae that led

Figure 3.4: Image of the Cerro Tololo Inter-American Observatory summit taken from above (https://www.darkenergysurvey.org/the-des-project/instrument/).

to the discovery of the accelerated expansion of the Universe by Riess, Perlmutter and Schmidt ([69, 9]). In Fig. 3.4 we show an image of the CTIO summit taken from above.

3.1.4 Observations

The Dark Energy Survey had 758 nights of observation in 6 years. The wide-area survey covers a large, contiguous area of sky (5,000 deg^2), each part of which is observed ten times in each *grizY* filter over the course of the survey; the long cumulative exposure time is needed to detect very faint galaxies, and the wide survey area enables us to detect a very large number of them. The deep or time-domain survey covers a smaller area on the sky (27 deg^2), but exposures are repeated a large number of times at regular, six-night intervals over the course of the survey. To carry out a complete mapping on all the filters, it was necessary to decide on an observation strategy. This was designed optimizing three variables:

- the possibility of studying the large-scale structure and the number of clusters from initial stages.
- the best observation of WL.
- the collection of a significant number of type Ia supernovae.

Name	Period	Area (deg^2)	Depth (i band)	Objects
SV	Nov. 2012 - Feb. 2013	250	23.68	25M
Y1	Aug. 2013 - Feb. 2014	1,800	23.29	137M
Y3	Aug. 2013 - Feb. 2016	5,000	23.44	399M
Y6	Aug. 2013 - Jan. 2019	5,000	23.80	691M

Table 3.1: Periods of time spanned by the data used in each DES analysis (SV, Y1, Y3 and Y6), area covered, magnitude reached in the i band and number of objects contained in the corresponding galaxy catalogs.

Nights with good seeing were called photometric nights (whenever seeing < 1.1"). In these nights, the observations were focused on the main survey, otherwise the DES observed the supernovae fields. If there were 7 consecutive photometric nights, the observations focused in supernovae fields for that night.

The tiling is the coverage of the entire DES area observing with a given filter. Successive tilings increase the exposure time in each filter. The 5,000 deg^2 were covered with 1,650 non-overlapping images from the DECam camera. Each year, DES had 105 nights of observation. which were between September and February (approximately 21 nights per month). The rest of the nights, the DECam was available for the scientific community. In Table 3.1 we display the periods of time spanned by the data used in each DES analysis (science verification or SV, Y1, Y3 and Y6) and the main properties of each galaxy sample.

3.1.5 The Dark Energy Survey Data Management

The Dark Energy Survey took about 400 images per night, being each image about a gigabyte in size. This was a very high data collection rate for an astronomy experiment by the time DES was taking its data. These images were sent to La Serena via a microwave link, and from there were sent to the National Center for Supercomputer Applications (NCSA) in Illinois if order to be stored and reduced. The Dark Energy Survey Data Management (DESDM) is in charge of processing and reducing the raw images taken by the DECam. Reduction consists of standard image corrections of the raw CCD information to remove instrumental effects and the joining of these images into 0.5 deg^2 single-epoch images. After this, galaxies and stars in the images are identified and cataloged, and their measured properties are stored in a database.

The reduction process includes over-scan removal, crosstalk and non-linearity corrections, bias subtraction, brighter-fatter effect correction, bad-pixel masking, astrometric matching, flagging of saturated pixels and bleed trails, principal-components background subtraction, secondary flat-field correction, and the masking of cosmic rays and other artifacts [70]. These reduced images are then used to generate the *multi-epoch* or *coadd* images, which are the combinations of the *single-epoch* ones. This procedure not only allows us to reduce the number of images, but also to detect

fainter objects and to mitigate the remaining spurious effects. The software used for the source detection is SExtractor [71], and is applied to the r, i and z band coadded images. Then, once the sources are detected, SExtractor is used on the coadded, individual band images to obtain the fluxes and the shapes of the galaxies. At the end of this process, a catalog in which each object has associated the information obtained during the reduction is produced. This procedure does not remove the impact of other systematic effects, such as the atmosphere conditions, dust and survey depth among others, which must be mitigated using other methodologies [72].

3.1.6 Photometric Redshifts

A photometric redshift is an estimate for the redshift of an astronomical object, such as a galaxy or quasar, without measuring its spectrum. The technique uses photometry (i.e., the brightness of the object in different standard filters) to determine the redshift.

In order to achieve its scientific objectives, DES needs to measure the redshift of hundreds of millions of galaxies, using images in the $grizY$ bands. The reconstruction codes of the photo-z require reference spectroscopic samples, which come from previous spectroscopic surveys, or fitting a spectral template library to the observations to infer the correct redshift.

3.1.7 The Large-Scale Structure in DES

The LSS is a 3D spatial distribution, but, as we already explained, DES is a photometric survey, and so its precision in the measurement of the redshift of galaxies is limited. To optimize the DES sensitivity to cosmological parameters, it is useful to measure the ACF of galaxies in photometric redshift slices, whose size is adapted to the DES precision in the redshift measurements. This method gives us the evolution of angular LSS with time, allowing the exploitation of the full sensitivity of DES to the contents of the Universe. The LSS is one of the four key probes in DES, together with supernovae, the number counts of clusters of galaxies and weak gravitational lensing.

Since LSS is one of the core measurements in DES, the so-called LSS working group has been active from the very beginning of the project, developing new analysis methods and studying the best ways to maximize the sensitivity to cosmological parameters. This enabled high precision measurements of galaxy clustering in five tomographic bins in the redshift range from 0.15 to $\sim$1, using the galaxies with best photometric redshifts, selected with the redMaGiC algorithm [73]. Such measurements provided one of the three different two-point correlation functions used to infer state-of-the-art cosmological constrains.

Several observational and theoretical systematic effects need to be accounted for in order to provide reliable LSS measurements. In particular, one has to disentangle true LSS from fluctuations induced by the changing weather and observing conditions

under which measurements are made. Also, in order to try to understand the effects of non-linear growth and galaxy bias, we run large simulations that are specially designed to mimic the DES, including all the effects that are present in the real measurements.

3.1.8 Theory and Combined Probes

Here we describe the mathematical convention and notation used for the modelling within the DES analyses. The so-called 3×2pt combines galaxy clustering, cosmic shear, and galaxy-galaxy lensing, and is one of the key cosmological probes of photometric LSS surveys. Even though the 3×2pt is not the cosmological probe we are interested in in this thesis, it is worth describing it since we will combine it with the Y3 BAO likelihood in Chapter 4 in order to infer cosmological parameters.

3.1.8.1 The 3×2pt Correlation Function

In DES we have two different catalogs: the position catalog ("lens galaxies") and the shape catalog ("source galaxies"). The lens or foreground galaxies are those who act as lenses that distort the shapes of the source or background galaxies. For both the DES Y3 and Y6 analyses, we use REDMAGIC [73] and MAGLIM [74] as the two fiducial lens samples, and we analyze the 3×2pt constraints for both of them independently. As for the source galaxies, we use the redshift distribution of the METACAL shear catalog [75, 76], which is split into 4 redshift bins.

From now on, we denote the projected density contrast of the lens galaxies in redshift bin i as $\tilde{\delta}^i_G$, the convergence field of the source galaxies at tomographic bin j as κ^j, the redshift distribution of the lens/source galaxy sample as $n^i_{G/\kappa}(z)$, and the angular number density of galaxies in this redshift bin as

$$\bar{n}^i_{G/\kappa} = \int dz\ n^i_{G/\kappa}(z). \tag{3.2}$$

The radial weight function for clustering in terms of the comoving radial distance χ is

$$q^i_G(k,\chi) = b^i(k, z(\chi)) \frac{n^i_G(z(\chi))}{\bar{n}^i_G} \frac{dz}{d\chi}, \tag{3.3}$$

where $b^i(k, z(\chi))$ is the galaxy bias of the lens sample in the redshift bin i. On the other hand, the lensing efficiency is given by

$$q^i_\kappa(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \frac{\chi}{a(\chi)} \int_\chi^{\chi_H} d\chi' \frac{n^i_\kappa(z(\chi'))}{\bar{n}^i_\kappa} \frac{dz}{d\chi'} \frac{\chi' - \chi}{\chi'}, \tag{3.4}$$

where χ_H is the comoving Hubble horizon, $\chi_H(z) = (1+z)d_H(z)$. By calculating the auto-correlation and the cross-correlation of both catalogs, we generate three two-point correlation functions:

1. A galaxy-galaxy auto-correlation, which compares the position of each galaxy with the position of every other galaxy. This is known as galaxy clustering, and we already studied it in Chapter 2. By measuring the positions of galaxies and correlating them with the positions of all other galaxies we can derive a value for the angular correlation function, $w(\theta)$, for different angular scales. We use the Landy-Szalay estimator to compute the $w(\theta)$, which we already defined in eq. (2.95).

   Under the Limber approximation (see section A.3 of Appendix A for further details), the angular power spectrum for galaxy clustering is given by

$$C^{ij}_{GG}(\ell) = \int d\chi \frac{1}{\chi^2} q^i_G \left(\frac{\ell+1/2}{\chi}, \chi\right) q^j_G \left(\frac{\ell+1/2}{\chi}, \chi\right) P_m \left(\frac{\ell+1/2}{\chi}, z(\chi)\right). \tag{3.5}$$

   The angular two-point correlation function for clustering is, then, given by eq. (2.86),

$$w^{ij}(\theta) = \sum_\ell \left(\frac{2\ell+1}{4\pi}\right) \mathcal{L}_\ell(\cos\theta) C^{ij}_{GG}(\ell). \tag{3.6}$$

   If the galaxy bias of the lens sample does not depend on the scale k, then

$$C^{ij}_{GG}(\ell) = \int d\chi \frac{1}{\chi^2} q^i_G(\chi) q^j_G(\chi) P_m \left(\frac{\ell+1/2}{\chi}, z(\chi)\right). \tag{3.7}$$

2. A galaxy-shear cross-correlation, which compares the position of each galaxy with the shape of every other galaxy. This is commonly known as galaxy-galaxy lensing. The cosmic lens distorts the shapes of background source galaxies, i.e., they appear tangentially sheared (along a line connecting lens and source galaxies). The amount of shear is quantified by a function $\gamma_t(\theta)$ that is sensitive to the mass associated with the foreground lens galaxies. We find larger distortions in the shapes of galaxies behind regions containing large amounts of matter.

   Under the Limber approximation, the angular power spectrum for galaxy-galaxy lensing is given by

$$C^{ij}_{G\kappa}(\ell) = \int d\chi \frac{1}{\chi^2} q^i_G \left(\frac{\ell+1/2}{\chi}, \chi\right) q^j_\kappa(\chi) P_m \left(\frac{\ell+1/2}{\chi}, z(\chi)\right). \tag{3.8}$$

   The angular two-point correlation function for galaxy-galaxy lensing is, then,

$$\gamma^{ij}_t(\theta) = \int d\ell \frac{\ell}{2\pi} J_2(\ell\theta) C^{ij}_{G\kappa}(\ell). \tag{3.9}$$

3. A shear-shear auto-correlation, which compares the shape of each galaxy with the shape of every other galaxy. This is commonly known as cosmic shear. We

expect that we are more likely to observe a distorted galaxy closer to another distorted galaxy, since their light path has been distorted by the same lens, i.e., the shape of neighbouring galaxies are more correlated than pairs of galaxies at greater separations. The functions $\xi_+(\theta)$ and $\xi_-(\theta)$ measure the shape information as a function of the angular separation.

Under the Limber approximation, the angular power spectrum for cosmic shear is given by

$$C^{ij}_{\kappa\kappa}(\ell) = \int d\chi \frac{1}{\chi^2} q^i_\kappa(\chi) q^j_\kappa(\chi) P_m \left(\frac{\ell + 1/2}{\chi}, z(\chi) \right). \tag{3.10}$$

The angular two-point correlation function for cosmic shear is, then,

$$\xi^{ij}_{+/-}(\theta) = \int d\ell \frac{\ell}{2\pi} J_{0/4}(\ell\theta) C^{ij}_{\kappa\kappa}(l). \tag{3.11}$$

We can combine these three two-point correlation functions together to maximize the use of all the information, this is known as 3×2pt. This allows DES to make much more accurate predictions of cosmological parameters and increases our understanding of systematics.

3.1.8.2 Models and Parameters

The DES data can be used to test different cosmological models, such as:

- ΛCDM model. In this first model we have a total of 6 parameters to estimate, namely Ω_m, Ω_b, H_0, A_s, n_s and τ.
- oΛCDM model. This second model is an extension to ΛCDM which includes the density parameter of curvature, Ω_k, as a free parameter instead of setting it to zero, i.e., we consider the possibility that the Universe is non-flat.
- wCDM model. This third model is an extension to ΛCDM which includes the dark energy equation of state parameter, w, as a free parameter instead of setting it to -1, i.e., we consider the possibility that dark energy is not a cosmological constant.

Along with these cosmological parameters, there are another 20 "nuisance"[1] parameters. These nuisance parameters account for uncertainties in our measurements of photometric redshifts, shear calibration, linear galaxy bias and the contribution of intrinsic alignment to the shear spectra.

[1] A nuisance parameter is a parameter which is required in order to model the process that generates the data, but is otherwise of little interest, i.e., it is not physically relevant.

3.1.8.3 Covariance

The likelihood that our data is indicative of a cosmological model depends on the observables that we have measured from our probes and is also influenced by two additional factors:

1. The statistical uncertainties. These statistical uncertainties decrease with the number of galaxies observed by DES.

2. The correlations in the 3×2pt datavector. The covariance matrix is a very important ingredient for the likelihood analysis, since it determines the error bars of the measurements and the confidence with which we reject or confirm a cosmological model. Galaxy clustering, galaxy-galaxy lensing and cosmic shear trace the same underlying matter density field, which means they are interdependent. Therefore, the data points cannot be treated as independent measurements, and it is necessary to take into account the correlations.

   It must be noted that other sources of cosmological information, such as type Ia supernovae or CMB measurements, can be treated independently, which means that the calculations of the correlations can be avoided.

3.2 The Dark Energy Spectroscopic Instrument

The Dark Energy Spectroscopic Instrument (DESI) is a scientific research instrument that is currently conducting a spectroscopic astronomical survey of distant galaxies. DESI observations will allow scientists to probe diverse aspects of cosmology, from dark energy to alternative theories to general relativity, and even to study neutrino masses. The data from DESI is being used to create three-dimensional maps of the distribution of matter covering an unprecedented volume of the Universe with huge amount of detail. This will provide insight into the nature of dark energy, and will help us understanding whether cosmic acceleration is due to a cosmic-scale modification of general relativity or due to a cosmological constant.

Organizationally, the Dark Energy Spectroscopic Instrument is an international collaboration that brings together more than 750 researchers from more than 70 institutions including Australia, Canada, China, Colombia, France, Germany, Korea, Mexico, Spain, Switzerland, the United Kingdom and the United States. The collaboration is led by the Lawrence Berkeley National Laboratory, which is managed by the University of California for the U.S. Department of Energy's Office of Science. DESI is located at Kitt Peak National Observatory near Tucson, Arizona. Kitt Peak is part of the NSF's National Optical-Infrared Astronomy Research Laboratory (NSF's OIR Lab).

As an instrument, the Dark Energy Spectroscopic Instrument is a Stage-IV ground-based dark energy experiment that studies the BAO and the growth of structure

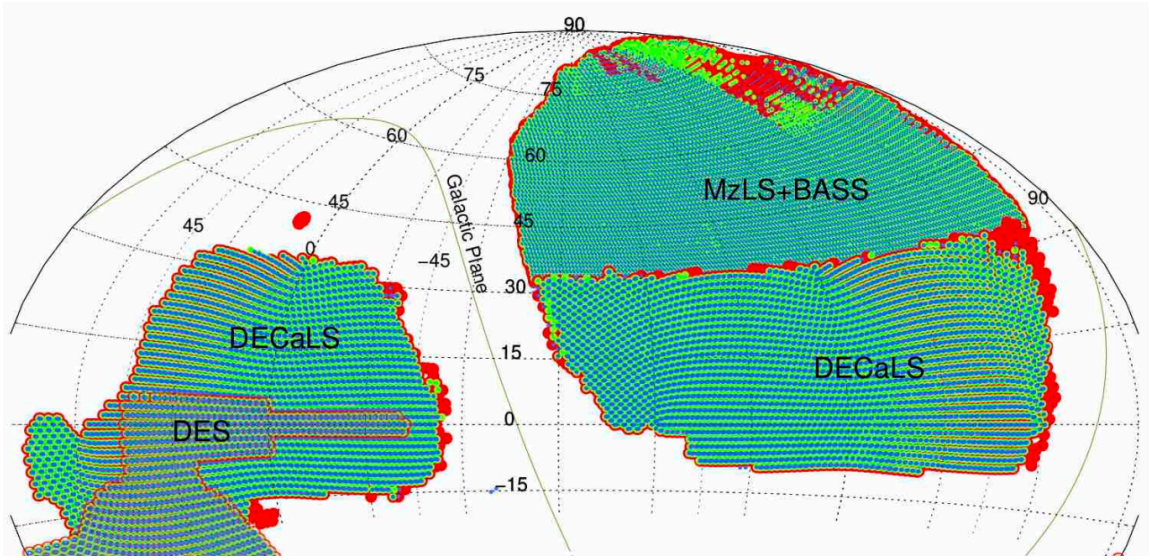

Figure 3.5: Footprint of the DESI survey, compared to that of the Dark Energy Survey, the Dark Energy Camera Legacy Survey (DECaLS), the Mosaic z-band Legacy Survey (MzLS) and the Beijing-Arizona Sky Survey (BASS). Image taken from the DESI key figures database.

through redshift-space distortions with a wide-area galaxy and quasar redshift survey. In Fig. 3.5 we show the DESI footprint, which covers, approximately, 16,000 deg^2 of the sky.

3.2.1 Scientific Goals

The primary cosmological mission of DESI is to study the nature of dark energy. Its three main science goals are:

- Dark energy. The Universe is undergoing an accelerated expansion phase, which has been confirmed by independent cosmological probes, as we already discussed in Chapter 1. DESI will use BAO to measure the expansion history of the Universe up to $z \sim 3.5$, which will allow us to constraint the equation of state of dark energy.

- Modified gravity. Modified gravity models can mimic the expansion of dark energy models, but one can distinguish between them measuring the growth of structure. DESI will study the anisotropy in the galaxy distribution in order to provide measurement of the growth rate of structures.

- Massive neutrinos. Massive neutrinos leave an imprint in the distribution of matter in the Universe, affecting both its expansion and growth. In combination with Planck, DESI will be able to accurately measure the sum of the neutrino

masses within 0.017 eV, and it might also be able to distinguish between the different mass orderings.

The main two questions that DESI aims to answer are how the dark energy density evolves with time and how it affects the clustering of matter. To do this, DESI will use its maps to measure two cosmological effects: baryon acoustic oscillations and redshift-space distortions, which are the two main probes of DESI. These same maps will also provide other opportunities to study cosmology and the physics of galaxies, quasars, and intergalactic gas.

3.2.2 The Instrument

The Dark Energy Spectroscopic Instrument is a fiber-fed spectrograph that is designed to measure the spectra of millions of galaxies and quasars over a large fraction of the sky. The instrument consists of several major components, including a corrector, a focal plane, a fiber system and a set of spectrographs. In this section we briefly go through each of them.

3.2.2.1 The Corrector

The corrector is designed to provide a larger field of view to the focal plate, allowing 5,000 fibers to be arranged over 8 square degrees. The design includes four large lenses composed of fused silica. The two smaller lenses are aspheric on one surface; all other silica surfaces are spherical. An atmospheric dispersion compensator is included in the design, in order to meet blur requirements at off-zenith angles. This compensator includes two large wedged borosilicate lenses with spherical surfaces. In Fig. 3.6 we show a model of the corrector of DESI.

3.2.2.2 The Focal Plane

The focal plane in DESI is designed to sit high on top of the Mayall Telescope and to carry 5,000 robotic positioners, each holding a fiber-optic cable. Each of these fiber-toting robots is automatically positioned to fix on a preset sequence of individual galaxies and quasars so that the fibers can collect their light. The movements of these positioners must be carefully choreographed to avoid bumping into one another. The roundish focal plane, which measures nearly a meter in diameter, consists of 10 pie-shaped wedges that are fit snugly together. Each wedge holds 500 robotic positioners. The focal plane also contains sensors and light sources called field fiducials that help ensure the positioners are in proper alignment. In Fig. 3.7 we show the focal plane of DESI.

3.2.2.3 The Fiber System

The fiber system is designed to transmit the light from 5,000 science targets imaged at the telescope's prime focus to DESI's spectrographs. Each optical fiber, with a

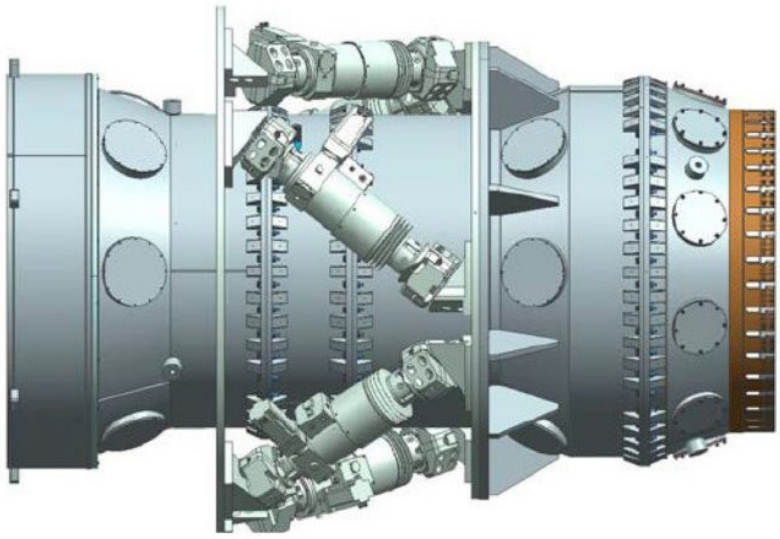

Figure 3.6: Model of the corrector of DESI (https://www.desi.lbl.gov/corrector/).

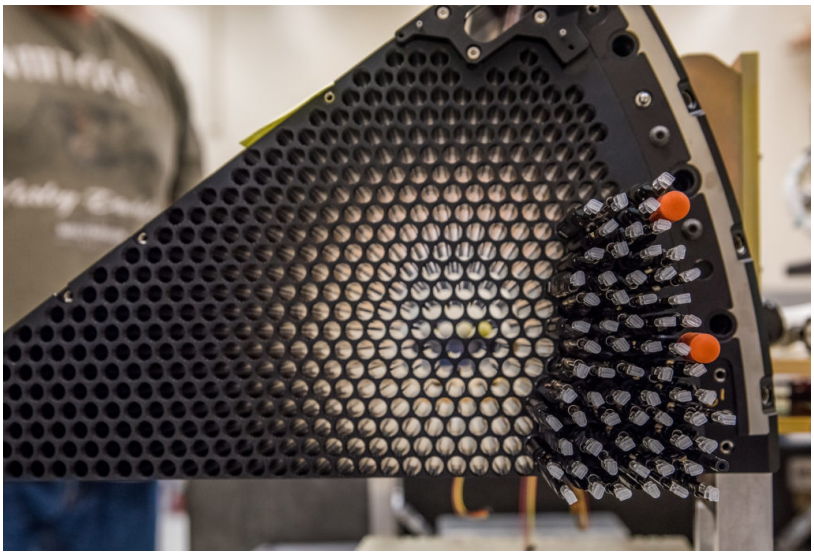

Figure 3.7: Image of 1 of the 10 wedges that form the focal plane of DESI (https://www.desi.lbl.gov/focal-plane-system/).

107-micron-diameter core, is terminated in a ferrule that is then mounted in a robotic fiber positioner. The fibers are collected behind the focal plane into 10 groups of 500. Each group spans 47.5 meters from the focal plane to the spectrograph room via guides, supports, strain-relieving spool boxes, and ruggedized cables. Each group of 500 fibers is then precisely arranged and terminated into a linear arc that provides the optical entrance-slit illumination to an associated spectrograph.

3.2.2.4 The Spectrograph

Each of DESI's 10 identical spectrographs accept one fiber-optic cable with 500 fibers. The light from the fibers enters the spectrograph through a slit that has a shutter in front of it. If the shutter is open, the light from the fibers is split using dichroics into three cameras that are each sensitive to different wavelength ranges and have different resolutions, which is given by

$$r = \frac{\lambda}{\Delta\lambda}. \tag{3.12}$$

Before the light lands on the detectors, it is spread using a volume phase holographic grating. The detector arrays have 4,096×4,096 15μm pixels and are each housed in their own cryostat so that can be kept very cold to maintain stability. The throughput of each arm of the spectrograph has been measured to demonstrate 70 percent efficiency across the whole redshift range. If the shutter in front of the slit is closed, there are several bright LEDs that are turned on so that the fibers are "backlit" and can be seen with the fiber-view camera. The Guiding, Focusing and Alignment system (GFA) is responsible for ensuring that the fibers of the instrument are correctly aligned with the targets in the sky and represents a significant technological achievement, enabling the instrument to perform highly precise and accurate measurements of the positions and spectra of galaxies and quasars over a large area of the sky.

3.2.3 The 4-Meter Mayall Telescope

DESI was mounted on the 4-meter Mayall Telescope at Kitt Peak National Observatory in 2018, and its construction and integration was the result of collaborations between several institutions. Kitt Peak National Observatory (KNPO) is located 56 miles southwest of Tucson, Arizona, in the Schuk Toak District on the Tohono O'odham Nation. In Fig. 3.8 we show an image of the KPNO. The Mayall Telescope is a reflector telescope with a 4-meter primary mirror that sits on an equatorial mount. It is the largest of the 22 optical telescopes located on Kitt Peak. It was commissioned in 1973 and it is nearly identical to the 4-meter Blanco Telescope at the Cerro Tololo Inter-American Observatory in Chile, also operated by NSF's OIR Lab and commissioned in 1976.

An extensive rebuilding process began at the Mayall Telescope in February 2018 to prepare the installation of the Dark Energy Spectroscopic Instrument (DESI). In

Figure 3.8: Image of the KPNO, with the 4-meter Mayall Telescope on the top of the mountain (https://www.flickr.com/photos/34450190@N08/8230626835).

this transition, the telescope became a sole-purpose-use telescope for deep and wide spectroscopic surveys with DESI. The Mayall was chosen to host DESI because of its unique combination of adaptable optical and mechanical designs, and the availability of a large block of dedicated survey time needed to carry out the DESI project.

3.2.4 Observations

DESI observations are split into bright and dark-time. Bright-time observations are those conducted when the moon is near full, since the targeted objects are bright enough to be observed under such conditions. This includes bright galaxies and Milky Way stars. On the other hand, the dark-time observations target three different types of galaxies: luminous red galaxies, emission line galaxies and quasars. Therefore, the DESI survey is based on four primary classes of galaxies:

1. Bright galaxies (BGS, which stands for bright galaxy sample). The redshift range of this type of galaxies extends to $\sim$0.4. These objects are bright enough that can be observed even when the moon is near full. While these are the easiest targets, they are still of high value because they probe the Universe most recently, when the accelerating expansion is strongest.

2. Luminous red galaxies (LRG). With them, the survey will reach redshift 1. These LRGs are composed mainly by old stars, and are the most massive galaxies. They can be seen at substantial distances and their red color makes them easy to select in imaging surveys.

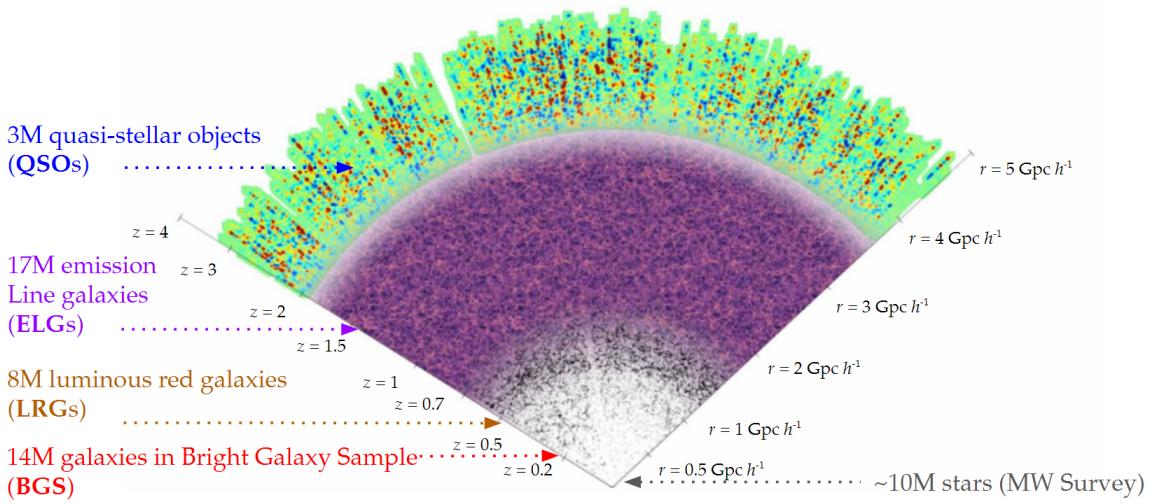

Figure 3.9: Scheme with the different dark matter tracers that DESI uses and where they are located as a function of redshift, from low to high redshift. Image taken from the DESI database, and modified to include the number of galaxies expected to be observed for each tracer.

3. Emission line galaxies (ELG). This type of galaxies extend to redshift ∼1.6. These are fainter and more distant galaxies, but their vigorous star formation and hot young stars create strong emission in distinct wavelengths that DESI can detect.

4. Quasars (quasi-stellar objects or QSO). Quasars outshine the stars of the galaxy and achieve luminosities that allow DESI to detect them out to redshift 3.5 and beyond. These are galaxies in which the central supermassive black hole is accreting large amounts of gas that is glowing as it reaches relativistic speeds. An important extra application of QSOs is that their spectra is altered by the light traveling through the intergalactic gas between them and us. The neutral hydrogen in this gas produces absorption at a key wavelength in the ultraviolet (121.6 nm), and that absorption pattern gets shifted into visible wavelengths due to the very high redshift of the gas. This produces a so-called Lyman-alpha forest and means that each quasar spectrum not only reveals its location in the map, but also a map of the intergalactic hydrogen along the LOS.

In Fig. 3.9 we show a scheme with the different dark matter tracers that DESI uses and where they are located as a function of redshift, from low to high redshift. From low to high redshift, DESI aims to measure the spectra of 14M bright galaxies, 8M LRG, 17M ELG and 3M QSO. In the upcoming section we will have a look at the status of the observations of DESI's main survey.

As we already mentioned, DESI aims to target a wide variety of galaxies to study the large-scale structure of the Universe and the nature of dark energy. To achieve

this goal, DESI uses a number of selection criteria to identify galaxies in different parts of the sky. One key selection criterion is the apparent brightness of the galaxies, which determines how easily they can be observed with the DESI spectrograph. DESI is designed to observe galaxies that are relatively bright and have well-measured redshifts, which helps to ensure accurate measurements of their distances and velocities. In addition to brightness, DESI also uses several other selection criteria to target galaxies of different types and at different distances. For example, DESI uses a technique called "color-color selection" to identify galaxies with specific colors that are indicative of their redshifts. This technique helps to select galaxies at different distances, which can be used to study the evolution of galaxy properties over cosmic time. Finally, DESI also targets galaxies in different regions of the sky, including in dense clusters, in large-scale filaments, and in underdense regions. This allows researchers to study the large-scale distribution of matter in the Universe and the relationship between galaxies and their dark matter halos.

3.2.4.1 Status of the Survey

DESI began its five-year survey in May 2021, and it is currently collecting data. DESI finished the data taking for the Y1 in September 2022, and its analysis is already ongoing. In Fig. 3.10 we show the evolution of the number of spectra for the different tracers as a function of time, from May 2021 to April 2023. The plot on the left shows the number of spectra collected in the bright-time survey, whereas the plot on the right is for the dark-time one. As of April 2023, DESI has collected spectra for about 9M BGS targets, 3.6M Milky Way targets, 3.8M LRG, 6.1M ELG and 2.2M QSO targets. The shaded regions showed on these plots represent periods in which observations were put on hold, which is the reason why the number of spectra does not increase. The first shaded region for both bright and dark-time observations represents the summer shutdown that DESI undergone between June-September 2021, whereas the second one is due to the Contreras fire, which started less than 15 km away from the KPNO and spread to thousands of acres, keeping DESI observations on hold between June-September 2022. Even though DESI had to put observations on hold for several months because of the Contreras fire, the main survey is still ahead of schedule. Taking a look at the total number of spectra that DESI aims to take for the different tracers, see Fig. 3.9, we find that DESI has already completed about a quarter of its main survey. Regarding the fraction of sky observed, as of April 2023, DESI has completed about 60% of its tiles for the bright-time survey, and about 40% for the dark-time one. In particular, about 12,000 $\deg^2$ have already been covered by at least one tile for both bright and dark-time surveys.

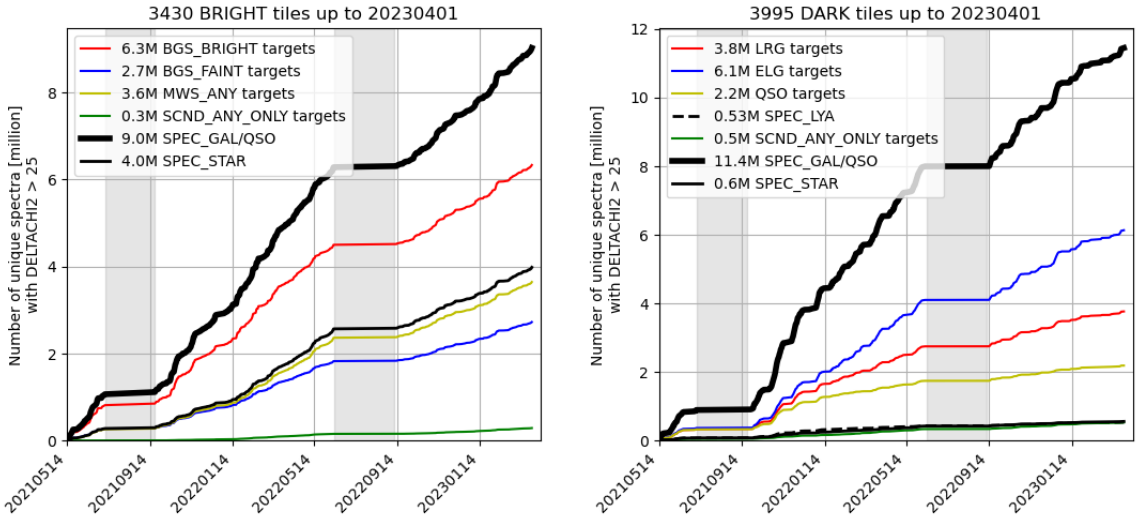

Figure 3.10: Number of DESI spectra as a function of time as of April 2023. **Left panel:** bright-time survey. **Right panel:** dark-time survey, which includes LRG, ELG and QSO. Credits to LBNL/A. Raichoor.

Chapter 4

Measurement of the BAO Scale in DES Y3

4.1 Introduction

The Dark Energy Survey has several methods to try to unveil the physical nature of dark energy, as we already mentioned in Chapter 3. One of these methods is the precise study of the spatial distribution of galaxies and, in particular, the BAO standard ruler. Since DES is a photometric survey, its precision in the measurement of redshifts is limited. However, a precise determination of the evolution of the angular diameter distance with redshift is possible through the measurement of the angular correlation function, $w(\theta)$, or the angular power spectrum, C_ℓ, in redshift shells adapted to the precision of the measurement. Here we focus on the measurement of the BAO signal in real or configuration space, i.e., using the angular correlation function.

One of the main difficulties in detecting the BAO peak in photometric surveys is the smearing of the signal produced by redshift uncertainties. In this context, it is necessary to select a galaxy population that presents a prominent spectral feature that can be captured with broadband filters and used to measure the redshift. Generally, the best practice is to select old, well-evolved galaxies with a significant Balmer break ([77, 78, 79]). This feature makes galaxies look very red, and they usually drive the target selection in galaxy surveys. The BAO signal observed in the DES data can be used in combination with other cosmological probes, such as CMB measurements, supernovae data, and weak gravitational lensing, in order to constrain cosmological parameters and shed light on the nature of dark energy, the geometry of the Universe, and other fundamental aspects of cosmology. The analysis of the BAO signal in the DES data is done through sophisticated statistical techniques and careful control of systematic uncertainties to obtain robust and accurate cosmological constraints.

4.1.1 Blind Analysis

In the context of cosmological surveys such as DES, a "blind analysis" refers to a scientific analysis approach where the researchers analyzing the data are not aware of certain details or parameters related to the survey or the specific measurements being analyzed. This is done to minimize potential biases or preconceived notions that could affect the results, such as the so-called "confirmation bias". In the case of the BAO measurements in DES, both the angular correlation function (in data) and the BAO distance measurements are blinded. We can, however, compute differences between the best-fit results obtained with different settings, but not the absolute values, which allows us to run a large battery of pre-unblinding robustness tests. We require our data to pass these consistency tests before unblinding, which makes our BAO measurement very robust. Once the analysis is completed, the blinding can be removed, and the final results can be interpreted and presented in a transparent and unbiased manner. It is worth mentioning that, in order to calibrate the clustering amplitude of the simulated galaxy catalogs (from now on, mocks) needed to validate the BAO-fitting pipeline, we do measure the angular correlation function of our BAO sample below 1 deg during the pre-unblinding period. Note that, since these scales do not contain any BAO information, these measurements do not interfere with the blinding scheme [50].

4.2 The Y3 BAO Sample

As we already discussed in Chapter 3, in DES we have two different lens galaxy samples, namely redMaGiC [73] and MagLim [74]. These are the samples that are used, together with the source sample, in order to obtain cosmological constraints via the 3×2pt probe. However, we also have a specific sample to measure the BAO signal. In the Y3 analysis, the BAO sample was created by selecting red galaxies from the DES Y3 Gold Catalog [68], following the same color selection as in the Y1 BAO analysis [80]. This selection is designed to balance the sample density with the photo-z precision above redshifts greater than 0.5 (the estimation of the photo-z is explained in 4.2.3). The Y3 BAO sample encompasses about 7 million galaxies over 4,100 deg^2 with photometric redshifts between 0.6 and 1.1. All of these galaxies were color and magnitude selected to balance trade-offs in BAO measurement between the number density and the photo-z accuracy. The Y3 BAO sample is supported by 1,952 simulated catalogs (produced using the COmoving Lagrangian Acceleration, or COLA, method, described in section 4.3) that match its main properties, in order to measure the BAO scale at an effective redshift of 0.835.

4.2.1 Angular Mask

The sample footprint is constructed directly from the high resolution HEALPix maps given in [68]. The main requirements imposed in the footprint are:

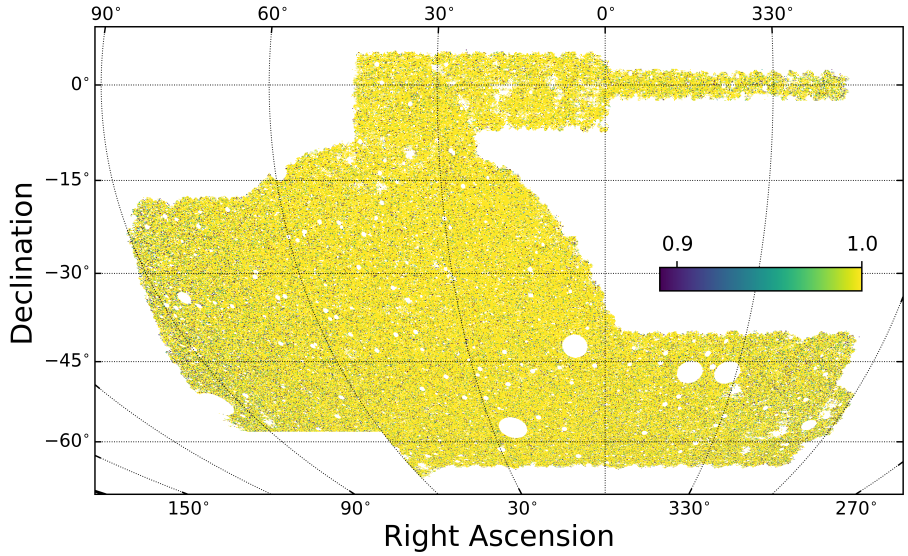

Figure 4.1: Footprint of the angular mask for the DES Y3 BAO analysis. The value plotted for each pixel represents its detection fraction. The total area of the mask, computed weighting by the detection fraction, is 4,108.47 deg^2.

1. Each pixel has to be observed at least once in each of the photometric bands, $griz$, with a coverage greater than 80% of the pixel area.

2. Pixels affected by foreground sources like extended galaxies or bright stars and regions affected by image artifacts are removed.

3. Pixels with magnitude limit of $i_{\rm lim} < 22.3$ at a signal to noise of 10 are also discarded.

The footprint of the Y3 sample is shown in Fig. 4.1. The total area of the Y3 angular mask is 4,108.47 deg^2.

4.2.2 Selection of the Sample

The DES Y3 BAO sample covers 4,108.47 deg^2, an area almost three times larger than the one of the Y1 BAO sample [80], and comprises 7,031,993 galaxies in the redshift range $0.6 < z_{\rm ph} < 1.1$, up to a magnitude limit of $i < 22.3$. In the context of a photometric galaxy survey, $z_{\rm ph}$ is the photometric redshift estimation, or photo-z. The fiducial photo-z code used in DES is the directional neighborhood fitting algorithm, or DNF, which we will describe in section 4.2.3. Changes in the data reduction between Y1 and Y3 samples increased the DES detection efficiency. This translated into a higher number density of source galaxies, even though the number of exposures per pointing in the footprint was, in average, the same. This allowed us to extend the analysis up to redshift 1.1, and also to a fainter magnitude limit with

respect to Y1. This extension of the sample meant a 10% gain in the precision of the combined BAO distance measurement.

In [81] are described the advantages of performing a multi-object, multi-epoch, multi-band fit (MOF, which stands for multi-object fitting) to the object shape to determine its morphology and flux using the NGMIX[1] software (see [82] for further details about the code). However, during the Y3 analysis a variant called SOF was introduced, which simplified the fitting process by eliminating the multi-object light-subtraction step without effectively affecting its performance, see [68]. Chromatic corrections and dereddened using SED-dependent extinction corrections were also taken into account (see [68]), while using the morphological classification based on SOF photometry in order to select secure galaxies. These choices are common to all Y3 cosmological results.

As we already mentioned, we apply the same color cuts as in Y1 to the Y3 Gold Catalog in order to select red galaxies. Despite the changes in photometry between Y1 and Y3, the color selection is still valid, as attested in [80]. The selection includes a flux-limit cut as a function of the photometric redshift, which was modified to reach $z_{\rm ph} < 1.1$. Therefore, the magnitude in the i band reaches 22.3 (for $z_{\rm ph} = 1.1$). The cuts finally applied are

$$\begin{aligned} 1.7 &< i - z + 2(r - i) \quad \text{(color selection)}, \\ 17.5 &< i < 19 + 3z_{\rm ph} \quad \text{(flux selection)}, \\ 0.6 &< z_{\rm ph} < 1.1 \quad \text{(photo-}z\text{ range)}, \end{aligned} \tag{4.1}$$

where r, i and z are the magnitudes observed in each photometric band for each galaxy, and $z_{\rm ph}$ is its photo-z. On apart from these selection cuts, we also include several quality cuts, namely

$$\begin{aligned} \text{EXTENDED\_CLASS\_MASH\_SOF} &= 3, \\ \text{FLAGS\_GOLD} &= 0, \\ \text{FLAGS\_FOOTPRINT} &= 1, \end{aligned} \tag{4.2}$$

where EXTENDED_CLASS_MASH_SOF is the star-galaxy classifier used in the Y3 Gold Catalog; FLAGS_GOLD indicates unusual characteristics of individual objects, including fitting failures and measurement anomalies; and FLAGS_FOOTPRINT ensures consistency between the object catalog and Y3 Gold footprint. See [83, 68] for further reference on these quality cuts.

4.2.3 Photometric Redshifts

We measure the BAO scale in a total of 5 tomographic redshift bins of width $\Delta z_{\rm ph} = 0.1$, from 0.6 to 1.1. In order to assign galaxies to each redshift bin, we use the photo-z estimate given by the directional neighborhood fitting algorithm (DNF) [84], which

[1] https://github.com/esheldon/ngmix

was trained using *griz* fluxes onto a large training sample. DNF is a non-parametric method that uses a training set of galaxies with known spectroscopic redshifts to establish the relationship between the observed colors and the true redshifts. It works by fitting a smooth function to the training set in a neighborhood of the data galaxy (in fluxes-color space), where the function predicts the redshift of a galaxy based on its colors. The key innovation of DNF is that it takes into account the fact that the relationship between colors and redshifts may vary in different regions of color space. It does this by defining a set of "directions" in color space, and then fitting a separate smooth function to the training set for each direction. This allows DNF to capture more complex relationships between colors and redshifts than traditional methods [84].

DNF predicts the best photo-z estimate (called $z_{\rm mean}$ in the DES catalogs), as well as the redshift of the closest neighbour ($z_{\rm mc}$) and the full PDF distribution. The photo-z estimate $z_{\rm mean}$ is computed as

$$z_{\rm mean} = \sum_i \boldsymbol{a}_i \cdot \boldsymbol{m}_i, \tag{4.3}$$

where i denotes a sum over neighbours in the spectroscopic training sample, $\boldsymbol{a}_i$ is a parameter vector and $\boldsymbol{m}_i$ are the magnitudes in the different bands. The vector $\boldsymbol{a}_i$ is obtained by fitting a linear function to the set of neighbours chosen, and then $z_{\rm mean}$ is used to assign galaxies to each redshift bin. In Table 4.1 we show the main properties of the Y3 BAO sample as a function of redshift. The $\langle z \rangle$ column is computed as the mean redshift of the redshift bin weighted with the redshift distribution of that same redshift bin,

$$\langle z \rangle = \int dz\, z n(z). \tag{4.4}$$

$N_{\rm gal}$ is the number of galaxies per redshift bin, and the variable σ_{68} is defined as the size of the region that encloses 68% of the distribution of

$$\frac{z_{\rm mean} - z_{\rm mc}}{1 + z_{\rm mc}}. \tag{4.5}$$

Finally, W_{68} is the width in redshift that encloses 68% of the integral of the $n(z)$, i.e.,

$$W_{68} = \frac{b-a}{2} \quad \text{such that} \quad \int_0^a dz\, n(z) = \int_b^\infty dz\, n(z) = 0.1585. \tag{4.6}$$

The stacking of the DNF $z_{\rm mc}$ values is a good approximation to the true $n(z)$ of the sample [84]. However, we only use it to validate the $n(z)$ estimated from a matched sample of the VIMOS Public Extragalactic Redshift Survey (VIPERS) [85], which are the ones used in the BAO distance measurement. VIPERS is a complete spectroscopic sample up to $i < 22.5$ for redshifts above 0.5 [85], and it covers 16.32 deg^2 of the DES footprint. Therefore, it is an excellent sample from which to estimate the true $n(z)$ for the 5 tomographic redshift bins. In Fig. 4.2 we plot the DNF redshift distributions

Bin	$\langle z \rangle$	$N_{\rm gal}$	σ_{68}	W_{68}
$0.6 < z < 0.7$	0.648	1,478,178	0.0246	0.0526
$0.7 < z < 0.8$	0.750	1,632,805	0.0279	0.0608
$0.8 < z < 0.9$	0.838	1,727,646	0.0298	0.0630
$0.9 < z < 1.0$	0.931	1,315,604	0.0363	0.0794
$1.0 < z < 1.1$	1.028	877,760	0.0455	0.0989

Table 4.1: Main properties of the Y3 BAO sample as a function of redshift: number of galaxies and redshift properties. The sample covers 4,108.47 deg^2.

for the BAO sample and VIPERS (as the stacking of the DNF $z_{\rm mc}$ values), and also for VIPERS $z_{\rm spec}$. The idea is that if $n(z_{\rm mc})$ for the BAO sample and VIPERS are statistically compatible, then we can use $n(z_{\rm spec})$ as the true redshift distributions of our sample. We find that, in fact, the redshift distributions $n(z_{\rm mc})$ are very similar, which allowed us to use $n(z_{\rm spec})$.

4.2.3.1 Effective Redshift

The effective redshift of the BAO sample is defined as its weighted mean redshift,

$$z_{\rm eff} = \frac{\sum_i w_{i,\rm sys} w_{\rm FKP}(z_i) z_i}{\sum_i w_{i,\rm sys} w_{\rm FKP}(z_i)} = 0.835, \tag{4.7}$$

where $w_{i,\rm sys}$ is the systematic weight of the i-th galaxy, see 4.2.4, and $w_{\rm FKP}$ is the statistical inverse-variance weight, see [83]. This definition was also used for the DES Y1 BAO analysis. Alternative definitions can lead to changes in $z_{\rm eff}$ of up to 0.035. These changes in $z_{\rm eff}$, translated into changes in α of the final measurement, are well below the statistical uncertainties reported in this thesis.

4.2.4 Observational Systematics

The number of galaxies is expected to fluctuate with the observing conditions of the survey. The method to mitigate the effect of these observing conditions consists on assigning weights to each pixel (or galaxy) in order to correct for spurious clustering. This method was developed in [86], and the weights were first applied in the DES Y1 3×2pt analysis. In Y3, the same algorithm is applied to all the clustering samples, i.e., to redMaGiC and MagLim. Details about it and results for other Y3 samples are given in [72].

In the original Y3 Gold Catalog, more than 100 survey property maps are available. First, the list of available maps is reduced to a set of 26 (almost-statistically) independent maps, including, for example, depth, airmass, stellar density and E(B-V) extinction. Then, it is studied how much the galaxy density varies with respect to a given survey property. When a significant relationship is found, a weight is assigned

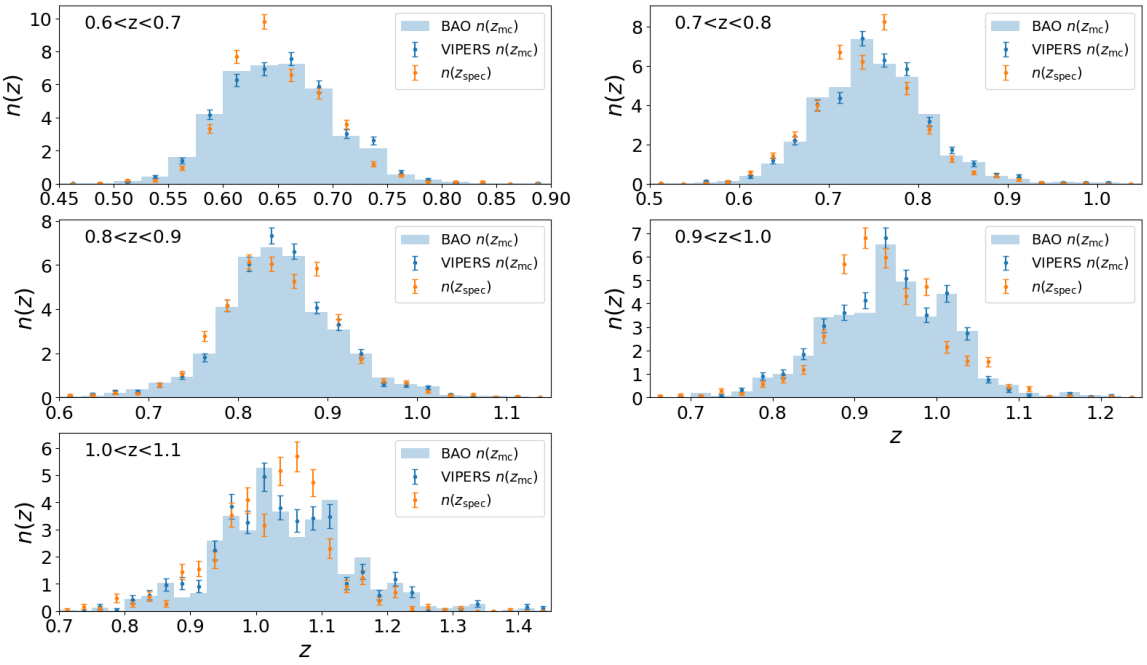

Figure 4.2: Redshift distributions of the Y3 BAO sample (blue histograms) and the VIPERS BAO sample (blue error-bars), obtained as histograms of $z_{\rm mc}$ in redshift bins selected with $z_{\rm mean}$. We also show the redshift distributions obtained as histograms of $z_{\rm spec}$ for the case of VIPERS (orange error-bars).

to each galaxy as a function of its position (seeing, airmass, ...) to correct for this effect. This process is run iteratively until all density-observing conditions relations are below a given threshold. In the case of the BAO sample, a threshold of 2σ was chosen. In Y3 it was found that the observational systematic correction was larger than the statistical error. This influence was not found for Y1 data. The main source of this change with respect to Y1 can be understood by the increase in the survey area, which implied larger variations of the observing conditions and smaller relative errors in the measurement of galaxy clustering [83].

Details about the results and weight validation tests can be found in [72]. By comparing to a set of 1,000 lognormal simulated galaxy catalogs, it was found that the amplitude of the correction was naturally arising from the observing pattern, and also that the error introduced by the corrections was small compared with the statistical errors in the angular correlation function.

4.3 Simulations: the COLA Mocks

In this section we describe the halo catalogs from dark matter simulations and the subsequent galaxy mock catalogs used for both the Y3 and the Y6 BAO analyses. We start by briefly describing the ICE-COLA fast simulations, which are based on the comoving Lagrangian acceleration (COLA) method, and then the fiducial pure N-body simulation, the so-called MICE "Grand Challenge" simulation. Both sets of simulations share the same cosmology, mass resolution and halos, defined with a friends of friends algorithm [87].

4.3.1 ICE-COLA Fast Simulations

One of our objectives is to build a huge number of mocks in order to test the BAO-fitting pipeline, and also to perform pre-unblinding and robustness tests before running on the actual data. In order to do so, we use a set of 488 fast N-body simulations generated with the ICE-COLA code [88]. The COLA method solves the evolution of the matter density field using a second-order Lagrangian perturbation theory (2LPT) combined with a particle-mesh solver to integrate the particle orbits at small scales, for which 2LPT deviates from the full N-body solution, see [89]. The ICE-COLA code uses the original COLA method but extended in order to produce on-the-fly lightcone halo catalogs together with weak lensing maps [90]. The ICE-COLA simulations use $2{,}048^3$ particles in a box of size 1,536 Mpc/h, in order to match the mass resolution of the MICE "Grand Challenge" simulation. The halo mass limit of the ICE-COLA mocks is $1.46 \times 10^{12} M_\odot$ (50 particles[2]).

[2] The threshold of 50 particles is not set deliberately, it is rather a resulting minimum number of particles of the halos of the mocks, see [87] for further reference.

4.3.2 The MICE "Grand Challenge" Simulation

In order to validate the range of scales in which we can trust our ICE-COLA fast simulations, we use the MICE "Grand Challenge" simulation as the benchmark N-body run for our BAO analyses (hereafter, we refer to this simulation as MICE). The MICE simulation is a large-scale cosmological simulation project that aims to model the formation and evolution of cosmic structures, such as galaxies and galaxy clusters, in the Universe. It is unique in its scale and scope, aiming to simulate a large volume of the Universe at high resolution. It covers a large cosmological volume, typically on the order of billions of light-years across, and includes billions of individual particles to represent the dark matter and other components of the Universe. The simulation incorporates a wide range of physical processes, including gravity, hydrodynamics, radiative cooling, and star formation, among others, to model the complex physics of cosmic structure formation. Specifically, MICE is an all-sky lightcone N-body simulation which evolves $4{,}096^3$ dark matter particles in a comoving volume of $\sim$29 Gpc^3/h^3, see [91, 27, 92]. The cosmology assumed for this simulation corresponds to the best fit of the Wilkinson Microwave Anisotropy Probe (WMAP) five-year data, see [93]. This is consistent with a flat ΛCDM model with $\Omega_b = 0.044$, $\Omega_c = 0.206$, $h = 0.7$, $\sigma_8 = 0.8$ and $n_s = 0.95$, and hereafter we refer to this particular set of parameters as "Mice" cosmology. Both simulations, i.e., the ICE-COLA and the MICE simulation, were found to be consistent with an accuracy of $\sim$5%, see [87].

4.3.3 Galaxy Lightcone Catalogs

The galaxy mock catalogs are constructed from the corresponding ICE-COLA halo catalogs described in the previous sections. A hybrid halo occupation distribution (HOD) - halo abundance matching model modeling strategy is adopted to populate the halos, as described in [94]. As we already mentioned earlier, three pre-unblinding values of the angular clustering on scales lower than one degree were provided to calibrate the clustering amplitude of the mocks. Once the mocks were created and the data passed through the rigorous process of unblinding we could compare the clustering, in both configuration and harmonic spaces, of the mocks with the final post-unblinding measurements of the data. These pre-unblinding clustering measurements below 1 deg do not violate our blinding protocol, since there is no BAO signal at such scales.

Box replications are needed in order to have lightcones reaching redshifts higher than 0.6 and also covering the DES footprint (0.6 is the limiting redshift if we set the lightcone origin at the center of a box of size 1,536 Mpc/h). Four boxes on each Cartesian direction are needed in order to create a full-sky lightcone up to redshift 1.4, since $d_M(z = 1.4) \sim 2{\times}1{,}536$ Mpc/h. Therefore, a total of 64 boxes are needed per lightcone (8 boxes up to $z \sim 0.6$, and another set of 56 boxes to reach $z \sim 1.4$). The implications of these replications are studied in [87], where the creation of the galaxy mock catalogs is also explained in detail. Since it is not one of the main purposes of

this thesis, it is not addressed here.

4.4 Measuring the BAO Signal

Our main objective here is to extract the BAO distance measurement from the clustering signal of the BAO sample, i.e., from the two-point angular correlation function or $w(\theta)$. We do it by fitting the data $w(\theta)$ to a given model, which gives us the position of the BAO peak in the correlation function. When performing the fits we do take into account correlations between angular bins and also cross-correlations between different redshift bins via the covariance matrix, i.e., we fit the $w(\theta)$ for all the redshift bins simultaneously. The χ^2 to be minimized is, then, given by

$$\chi^2(\mathbf{\Theta}) = \sum_{\mathrm{zbin}_1,\mathrm{zbin}_2} \sum_{i,j} \Delta w^{\mathrm{zbin}_1}(\theta_i, \mathbf{\Theta})(\mathrm{cov}^{-1})_{i,j}^{\mathrm{zbin}_1,\mathrm{zbin}_2} \Delta w^{\mathrm{zbin}_2}(\theta_j, \mathbf{\Theta}), \tag{4.8}$$

where

$$\Delta w^{\mathrm{zbin}}(\theta, \mathbf{\Theta}) = w_{\mathrm{data}}^{\mathrm{zbin}}(\theta) - w_{\mathrm{model}}^{\mathrm{zbin}}(\theta, \mathbf{\Theta}). \tag{4.9}$$

The parameters $\mathbf{\Theta}$ are the free parameters of our model. The covariance matrix that appears in eq. (4.8) is the covariance matrix of the $w(\theta)$, and can be calculated either analytically, using mocks or estimated with the data itself. As for the model $w_{\mathrm{model}}^{\mathrm{zbin}}(\theta, \mathbf{\Theta})$, we use a template-based method in which the position of the BAO peak is given in terms of the theoretical $w(\theta)$ for a fixed reference cosmology. Template-based methods have been widely used in the literature with great success [95, 51, 38, 52], since they allow us to fit a theoretical model of the correlation function that includes the expected BAO signal, as well as other effects such as galaxy bias and redshift-space distortions. Experiments including SDSS and DES [95, 51] have shown these methods can provide precise measurements of the BAO scale, which can then be used to constrain the properties of dark energy and the expansion history of the Universe.

4.4.1 The BAO-Fitting Pipeline

One of the main objectives of this thesis has been the development of a pipeline to perform BAO distance measurements in configuration space for the DES analysis, and in this section we describe how it works. As we already mentioned, we extract the BAO distance measurement from the clustering signal, i.e., from $w(\theta)$, using a template-based method. This approach has been widely used in the literature, for both spectroscopic and photometric datasets. As we explained in Chapter 2, the main difference between these two is that for photometric datasets we are able to extract only the angular diameter distance, i.e., the information perpendicular to the LOS, whereas for spectroscopic datasets we have access to both. The template-based method we use consists on a parametrization of the BAO feature in terms of the shift

α in the position of the BAO peak of the data $w(\theta)$ with respect to a theoretical $w(\theta)$, or template, calculated for a given reference cosmology. Since the change in the shape of the $w(\theta)$ for similar reference cosmologies, i.e., similar sets of cosmological parameters, is small, the method shifts and stretches the theoretical $w(\theta)$ in order to look for the best fit to the data. We refer to this method as the α fit, and we explain it in detail in this section. The BAO-shift parameter α is given by

$$\alpha(z_{\rm eff}) = \frac{r_d^{\rm ref}/d_A^{\rm ref}(z_{\rm eff})}{r_d/d_A(z_{\rm eff})} = \frac{r_d^{\rm ref}/d_M^{\rm ref}(z_{\rm eff})}{r_d/d_M(z_{\rm eff})}, \tag{4.10}$$

where r_d is the sound horizon scale at the photon decoupling epoch, given by eq. (2.99). We can use either $d_A(z)$ or $d_M(z)$ to quote the results, keeping in mind that $d_M(z) = (1+z)d_A(z)$; however, the distance measurements are given in terms of $d_M(z)$ throughout this chapter[3]. $r_d^{\rm ref}$ and $d_M^{\rm ref}(z_{\rm eff})$ are known quantities, since they can be calculated simply from the cosmological parameters of the reference cosmology. Therefore, the actual measurement of the α fit is the angular diameter distance divided by the scale of the sound horizon evaluated at the drag epoch,

$$\frac{d_M(z_{\rm eff})}{r_d} = \frac{d_M^{\rm ref}(z_{\rm eff})}{r_d^{\rm ref}}\alpha(z_{\rm eff}). \tag{4.11}$$

$z_{\rm eff}$ is the effective redshift of the galaxy sample, which is defined by eq. (4.7). This measurement of the angular diameter distance divided by the sound horizon scale can be used to obtain constraints on different cosmological parameters, as we will see in section 4.6, particularly when combined with other cosmological probes, since it allows to break degeneracies between parameters.

The model for the α fit is, explicitly, given by

$$w_{\rm model}^{\rm zbin}(\theta, \mathbf{\Theta}) = B^{\rm zbin} w_{\rm ref}^{\rm zbin}(\alpha\theta) + \sum_i A_i^{\rm zbin}\theta^{-i}, \tag{4.12}$$

where $w_{\rm ref}^{\rm zbin}(\theta)$ is the theoretical angular correlation function computed for some fixed reference cosmology at a certain redshift bin, and $\mathbf{\Theta} = \{\alpha, B^{\rm zbin}, A_i^{\rm zbin}\}$. The parameters $B^{\rm zbin}$ multiply the template in order to fit for the clustering amplitude of the data. The parameters $A_i^{\rm zbin}$ are the so-called broadband-term parameters, and they account for small residual differences in the shape between the data and the template. By default, the sum over i in the previous expression takes the values $i = \{0, 1, 2\}$, i.e., we have a constant additive factor, a term that goes as θ^{-1} and a term that goes as θ^{-2}. However, we also test the effect of including less broadband-term parameters or including an extra one, proportional to θ, on the BAO distance measurement. Therefore, for the default modeling case we have a total of

$$1 + 4 \times (\text{number of redshift bins}) \tag{4.13}$$

[3] We must keep in mind that the quantity that we measure here is not $d_M(z)$, but rather $d_M(z)/r_d$.

free parameters for the α fit. As we mentioned before, the BAO scale is given in terms of the shift α with respect to the reference cosmology, eq. (4.10), i.e., α is the physically-relevant parameter of our model. The other remaining parameters are treated as nuisance parameters, i.e., we marginalize over them[4], see section 4.4.2.

Taking a look at eq. (4.12), we note that there is a single α parameter regardless of how many redshift bins are fitted, and it is evaluated at the effective redshift of the sample. The reason for this is that, as we already explained, we are fitting the $w(\theta)$ for all the redshift bins simultaneously (using a covariance matrix that includes correlations between angular bins and also cross-correlations between redshift bins), which means that our BAO-fitting results are evaluated at the effective redshift of the sample. However, in order to test the BAO-fitting pipeline, we also run the BAO fits removing certain redshift bins. In these situations, the effective redshift must be recomputed not taking into account the galaxies from the removed redshift bins. We must also note that $d_M(z_{\rm eff})/r_d$, which is given by eq. (4.11), does not depend on the reference cosmology. Different reference cosmologies lead to different values for $d_M^{\rm ref}(z_{\rm eff})/r_d^{\rm ref}$, but the BAO fit also gives different values for α in such a way that the product of both quantities remains constant, i.e., independent of the reference cosmology. In the subsequent sections we will test our BAO-fitting pipeline using two different reference cosmologies in order to check that this holds.

The method we use to perform the BAO fits is a maximum likelihood estimator (MLE), similarly to the one explained in [96]. The χ^2 for the BAO fit is given by eq. (4.8), and the model is given by eq. (4.12). If the data is Gaussianly distributed, which, as we will see, is a very good approximation, then we can relate the likelihood to the χ^2 as

$$\mathcal{L}(\mathbf{\Theta}) \propto \exp\left(-\chi^2(\mathbf{\Theta})/2\right). \tag{4.14}$$

MLE is a point estimator for the parameters $\mathbf{\Theta}$ that seeks the best fit by maximizing this likelihood function. MLE itself does not require the likelihood to be Gaussian, and in that case χ^2 minimization can still be used to find the best fit. However, its connection to the probability distribution is, then, not straightforward [96].

As we already mentioned earlier, the only parameter we are interested in is the BAO-shift parameter α. The remaining free parameters are just nuisance parameters, i.e., they are not physically relevant. By taking a look at the model given by eq. (4.12), we find that the broadband-term parameters, $A_i^{\rm zbin}$, appear linearly, which means they can be fit analytically using the least-square fit method, see [97] for further reference. This allows us to obtain their best-fit values, $A_{i,\rm bf}^{\rm zbin}$, which means we end up with a χ^2 that no longer depends on these parameters,

$$\chi^2(\alpha, B^{\rm zbin}, A_i^{\rm zbin}) \to \chi^2(\alpha, B^{\rm zbin}, A_{i,\rm bf}^{\rm zbin}). \tag{4.15}$$

The next step is to numerically search for the best-fit parameters $B_{\rm bf}^{\rm zbin}$ with the prior $B^{\rm zbin} > 0$ such that our χ^2 is minimized with respect to $B^{\rm zbin}$. We add this prior

[4] For the Y3 analysis, we have a total of 20 nuisance parameters.

to avoid nonphysical solutions with $B^{\rm zbin} <= 0$, since $B^{\rm zbin}$ is the parameter that multiplies the template in eq. (4.12) and, therefore, it has to be positive. We are finally left with a one-parameter χ^2 function

$$\chi^2(\alpha, B^{\rm zbin}, A^{\rm zbin}_{i,\rm bf}) \rightarrow \chi^2(\alpha, B^{\rm zbin}_{\rm bf}, A^{\rm zbin}_{i,\rm bf}). \tag{4.16}$$

The χ^2 is then sampled as a function of α. For each value of α we have to minimize the $B^{\rm zbin}$ parameters with the condition $B^{\rm zbin} > 0$ while analytically computing $A^{\rm zbin}_i$ as explained earlier. We consider we have a detection if $\alpha \pm \sigma_\alpha$ lies within the interval $[0.8, 1.2]$. If it lies outside this range, we consider it as a non-detection. The reason for this is that this interval is about 10 times larger than the expected error (which is about 2.5%, see [50]), and since previous measurements ensure a deviation no larger than 2-3%, a deviation larger than 20% would indicate there is no BAO peak to fit for. The 1-σ error bar is derived from the deviation from the minimum χ^2, $\chi^2_{\rm min}$, by $\Delta\chi^2 = 1$ [98]. This assumes that the likelihood of α is Gaussianly distributed. Let's expand the logarithm of the likelihood about $\alpha = \alpha_0$ where the maximum of the likelihood is attained as

$$\log \mathcal{L}(\alpha) = \log \mathcal{L}_0 + \frac{1}{2}\frac{\partial^2 \log \mathcal{L}}{\partial \alpha^2}\Big|_{\alpha_0} (\alpha - \alpha_0)^2 + \mathcal{O}[(\alpha - \alpha_0)^3]. \tag{4.17}$$

When the so-called Cramér-Rao bound is saturated [99], the variance of α is given by

$$\sigma^2_\alpha = \left(-\frac{\partial^2 \log \mathcal{L}}{\partial \alpha^2}\Big|_{\alpha_0} \right)^{-1}, \tag{4.18}$$

which implies that the curvature of $\log \mathcal{L}$ encodes the error bar on α. Therefore, at 1-σ from α_0,

$$\log \mathcal{L}(\alpha_0 \pm \sigma_\alpha) = \log \mathcal{L}_0 - \frac{1}{2}. \tag{4.19}$$

Since $\mathcal{L} \propto \exp(-\chi^2/2)$, this is equivalent to the $\Delta\chi^2 = 1$ rule for the 1-σ error bar in MLE [96].

4.4.1.1 The BAO Template

Computing the template $w(\theta)$ that appears in eq. (4.8) is a very important part of the pipeline. We build the BAO template starting from a linear power spectrum $P^{\rm lin}_m(k)$ from CAMB[5] (Code for Anisotropies in the Microwave Background). CAMB has been exhaustively verified and provides a power spectrum precise enough in order to perform our BAO analysis [100]. At BAO scales the main modification due to non-linear evolution is the broadening of the BAO feature due to large-scale flows. We model this by introducing a Gaussian damping of the BAO wiggles, after isolating this

[5] https://camb.info/

component from the broadband shape. The galactic power spectrum as a function of k, μ and z is given by eq. (2.117), but modified to account for the non-linear evolution, see [50], as

$$P_G(k,\mu,z) = \left[b + f(z)\mu^2\right]^2 \times \left\{\left[P_m^{\rm lin}(k) - P_m^{\rm nw}(k)\right]\exp\left(-k^2\Sigma^2_{\rm tot}(\mu,z)\right) + P_m^{\rm nw}(k)\right\}, \tag{4.20}$$

where $P_m^{\rm nw}$ describes the smooth shape of the matter power spectrum (the no-wiggle matter power spectrum) and all the BAO information is in $P_m^{\rm lin}(k) - P_m^{\rm nw}(k)$. Therefore, the baryonic power spectrum can be expressed as

$$P_b(k,\mu,z) = \left[b + f(z)\mu^2\right]^2 \left[P_m^{\rm lin}(k) - P_m^{\rm nw}(k)\right]\exp\left(-k^2\Sigma^2_{\rm tot}(\mu,z)\right). \tag{4.21}$$

As we already explained in Chapter 2, the factor $(b + f\mu^2)^2$ accounts for linear-theory redshift-space distortions, where μ is the cosine of the angle of the wave vector $\boldsymbol{k}$ with respect to the LOS, b is the linear galaxy bias and f is the logarithmic derivative of the growth factor with respect to the scale factor, evaluated at the effective redshift of the sample. As for the no-wiggle power spectrum, there are several ways to define it. We start by defining the ratio of P and a smooth approximation to it, for which we employed the Eisenstein-Hu (EH) no-wiggle fitting formulae [101],

$$P_m^{\rm nw}(k) = P_{\rm EH}(k) \times \left[\mathcal{F} * R\right](k), \tag{4.22}$$

where $R(k) = P_m(k)/P_{\rm EH}(k)$, and the convolution with the filter $\mathcal{F}$ is given by

$$\left[\mathcal{F} * R\right](k) = \frac{1}{\sqrt{2\pi}\lambda}\int d\log q\, R(q)\exp\left(-(\log(k/q))^2/2\lambda^2\right). \tag{4.23}$$

This methodology is fully explained in Appendix A of [102]. We use $\lambda = 0.25$, for which we recover $P_m^{\rm nw} \to P_m^{\rm lin}$ at low and large k values, i.e., sufficiently away from the BAO feature scale. For the re-summation of infrared modes we take as a reference the implementation in [103, 104] and write

$$\Sigma^2_{\rm tot}(\mu,z) = \mu^2\Sigma^2_{\parallel}(z) + (1-\mu^2)\Sigma^2_{\perp}(z) + f(z)\mu^2(\mu^2-1)\delta\Sigma^2(z). \tag{4.24}$$

The parallel and perpendicular components of $\Sigma(z)$ are given by

$$\Sigma_{\parallel}(z) = \left[1 + f(z)\right]\Sigma(z), \quad \Sigma_{\perp}(z) = \Sigma(z), \tag{4.25}$$

and the redshift evolution is determined by $D(z)$ as

$$\begin{cases} \Sigma(z) = D(z)\Sigma_0, \\ \delta\Sigma(z) = D(z)\delta\Sigma_0. \end{cases} \tag{4.26}$$

	Ω_b	Ω_c	h	σ_8	n_s
Mice cosmology	0.044	0.206	0.7	0.8	0.95
Planck cosmology	0.048	0.262	0.676	0.8	0.97

Table 4.2: Cosmological parameters of the two different templates used for the BAO fits. The first one, Mice, is the same cosmology as the one of our fiducial set of simulations, the COLA mocks. The second one, Planck, is consistent with [14].

The $(1+f)$ factor in the parallel component is due to the Kaiser formula. At $z=0$, we find

$$\begin{cases} \Sigma_0^2 = \dfrac{1}{6\pi^2}\displaystyle\int_0^{k_s} dq P_m^{\rm nw}(q)\left[1 - j_0(q\ell_{\rm BAO}) + 2j_2(q\ell_{\rm BAO})\right], \\ \delta\Sigma_0^2 = \dfrac{1}{2\pi^2}\displaystyle\int_0^{k_s} dq P_m^{\rm nw}(q) j_2(q\ell_{\rm BAO}), \end{cases} \tag{4.27}$$

where j_n are the spherical Bessel functions of order n, while $\ell_{\rm BAO}$ is the wavelength of the BAO. We choose $k_s = 0.2\ h/\text{Mpc}$ and $\ell_{\rm BAO} = 110\ \text{Mpc}/h$ as reference values, but our results do not depend on these choices. For the computation of the no-wiggle power spectrum, we use the software CosmoPrimo, which is publicly available at its code repository[6]. CosmoPrimo has several different engines to compute the no-wiggle power spectrum, but we use the one based on the method developed in [105]. We use two different sets of cosmological parameters, or two different reference cosmologies, in order to generate the BAO template:

- Mice cosmology, which is consistent with WMAP five-year cosmological results, see [93]. More importantly, it is the same cosmology that we used for our fiducial set of simulations for the BAO analysis, the COLA mocks. Both the ICE-COLA fast simulations and the MICE "Grand Challenge" lightcone N-body simulation use these cosmological parameters, as we already discussed in section 4.3.

- Planck cosmology, which is consistent with Planck-2018 [14]. This is the fiducial template used when analysing the data, and the motivation is that we want to test the consistency of BAO with the high-redshift results coming from Planck.

The cosmological parameters for both reference cosmologies are displayed in Table 4.2. We obtain $\Sigma_0 = 5.80\ \text{Mpc}/h$, $\delta\Sigma_0 = 3.18\ \text{Mpc}/h$ for Mice cosmology and $\Sigma_0 = 5.30\ \text{Mpc}/h$, $\delta\Sigma_0 = 2.81\ \text{Mpc}/h$ for Planck cosmology.

For the angular correlation function computation we set the redshift dependence of all the previous functions to the mean redshift of the redshift bin, which is computed using the redshift distributions of the galaxies in that bin,

$$\langle z \rangle = \int_0^\infty dz\ z n(z). \tag{4.28}$$

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

We now define $P_G(k,\mu)$ as

$$P_G(k,\mu) \equiv P_G(k,\mu,\langle z\rangle)$$
$$= \left[b + f(\langle z\rangle)\mu^2\right]^2 \left\{\left[P_m^{\rm lin}(k) - P_m^{\rm nw}(k)\right] \exp\left(-k^2\Sigma_{\rm tot}^2(\mu,\langle z\rangle)\right) + P_m^{\rm nw}(k)\right\}. \quad (4.29)$$

From the previously computed galactic power spectrum, $P_G(k,\mu)$, we perform a multipole expansion in Legendre polynomials in μ, analogous to the multipole expansion in eq. (2.114),

$$P_{G,\ell}(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu P_G(k,\mu)\mathcal{L}_\ell(\mu). \quad (4.30)$$

Following eqs. (2.120) and (2.121), the galactic correlation function is, then, given by

$$\xi_G(s,\mu) = \sum_\ell \xi_{G,\ell}(s)\mathcal{L}_\ell(\mu), \quad (4.31)$$

where

$$\xi_{G,\ell}(s) = i^\ell \int_0^\infty dk j_\ell(ks)\frac{k^2}{2\pi^2}P_{G,\ell}(k). \quad (4.32)$$

Similarly to what we did in Chapter 2, we can write the variables s and μ in terms of the redshifts and the angular separations of any two galaxies (z_1, z_2 and θ) as

$$\xi_G(s,\mu) = \xi_G\left(s(z_1,z_2,\theta),\mu(z_1,z_2,\theta)\right)$$
$$= \sum_\ell \xi_{G,\ell}\left(\sqrt{r_1^2+r_2^2-2r_1r_2\cos\theta}\right)\mathcal{L}_\ell\left(\frac{r(z_1)-r(z_2)}{\sqrt{r_1^2+r_2^2-2r_1r_2\cos\theta}}\right). \quad (4.33)$$

Finally, the angular correlation function is obtained projecting $\xi(z_1,z_2,\theta)$ in redshift using the redshift distribution of the corresponding bin,

$$w(\theta) = \int dz_1 D(z_1)\phi(z_1)\int dz_2 D(z_2)\phi(z_2)\xi_G(z_1,z_2,\theta), \quad (4.34)$$

where the growth factor, $D(z)$, accounts for the redshift evolution of the matter power spectrum. We compute the template $w(\theta)$ between 0.001 deg and 179.5 deg in 1000 steps with logarithmic spacing. In Fig. 4.3 we show the template $w(\theta)$ computed for the Y3 analysis, using both Mice and Planck cosmologies. The amplitude of the $w(\theta)$ for both reference cosmologies is similar between 1.5 deg and 5 deg because we fitted the linear galaxy bias beforehand using the mean $w(\theta)$ of the Y3 COLA mocks. In order to fit for the linear galaxy bias, we split the computation of the $w(\theta)$ in three parts. Taking a look at eq. (4.29) we notice that the dependence on the linear galaxy bias can be written as

$$P_G(k,\mu) =$$
$$\left(b^2 + 2bf\mu^2 + f^2\mu^4\right)\left\{\left[P_m^{\rm lin}(k) - P_m^{\rm nw}(k)\right] \exp\left(-k^2\Sigma_{\rm tot}^2(\mu,\langle z\rangle)\right) + P_m^{\rm nw}(k)\right\}. \quad (4.35)$$

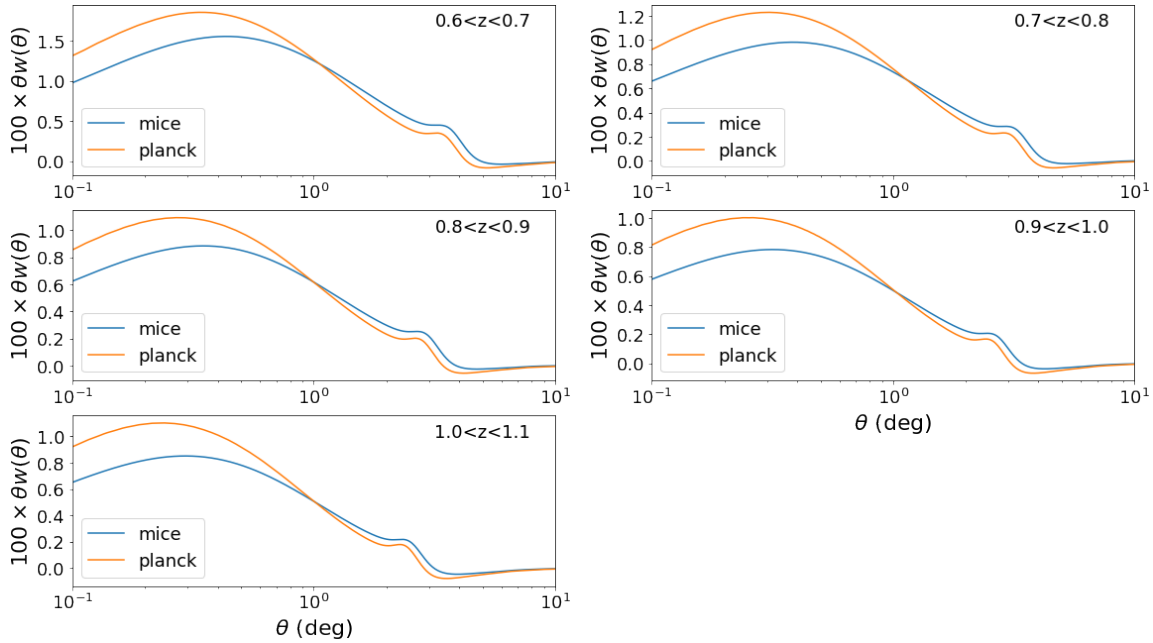

Figure 4.3: Template $w(\theta)$ used for the Y3 analysis, computed using both Mice (blue) and Planck (orange) cosmologies. The BAO peak can be observed on angular scales in the range $\theta \in (2, 4)$ deg, depending on the redshift bin.

We have a term that goes as b^2, another one that goes as b and a constant one, i.e.,

$$P_G(k, \mu) = b^2 A(k, \mu) + bB(k, \mu) + C(k, \mu). \tag{4.36}$$

Therefore, it is particularly useful to compute and store these $A(k, \mu)$, $B(k, \mu)$ and $C(k, \mu)$ instead of $P_G(k, \mu)$, since this way we do not need to assume any value for the linear galaxy bias. Furthermore, since the angular correlation function $w(\theta)$ is linear in $P_G(k, \mu)$, we can propagate the splitting into these three terms, yielding

$$w(\theta) = b^2 A(\theta) + bB(\theta) + C(\theta). \tag{4.37}$$

This allows us to easily fit for the linear galaxy bias, which is useful for both the mocks and the data itself. The galaxy bias is necessary to compute important quantities such as the analytical covariance matrices that we explained in section 4.4.1.2, which need to be computed for different sets of cosmological parameters.

Even though it is not part of this thesis, the BAO analysis in DES is also performed in Fourier space, and not just in configuration space. Therefore, it is necessary to also compute a C_ℓ theoretical template, and not just the $w(\theta)$ one. A natural way to compute this template is to transform the template $w(\theta)$ into its corresponding angular power spectrum, which is actually particularly useful because it allows us to use analogous templates for both analyses. The angular power spectrum, i.e., the Fourier transform of the angular correlation function, can be obtained from the

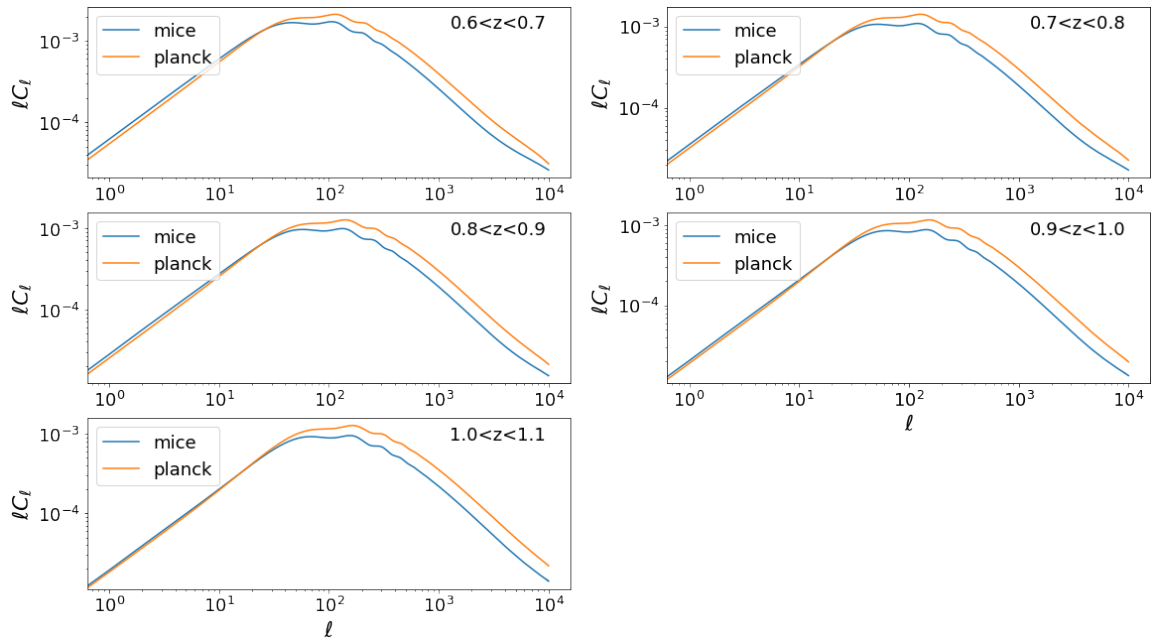

Figure 4.4: Template C_ℓ used for the Y3 analysis, computed from the $w(\theta)$ using eq. (2.88) for both Mice (blue) and Planck (orange) cosmologies. The BAO wiggles can be observed on scales with $\ell \in (50, 500)$.

angular correlation function via eq. (2.88). In Fig. 4.4 we show the template C_ℓ computed from the $w(\theta)$ using eq. (2.88). We compute up to $\ell = 10^4$, which is a much larger ℓ value than the maximum value defined by the scale cuts for the Fourier space analysis (typically, $\ell_{\rm max} \sim 500$). As a sanity check, we compute back the angular correlation function using eq. (2.86),

$$w(\theta) = \sum_{\ell \geq 0} \left(\frac{2\ell + 1}{4\pi} \right) \mathcal{L}_\ell(\cos\theta) C_\ell. \tag{4.38}$$

In Fig. 4.5 we show the ratio of the original $w(\theta)$ and the one recovered from the C_ℓ between 0 and 10 deg. We find that this ratio is very close to 1 throughout the whole angular range, which is well within the scale cuts used for the BAO analysis, as we will see in section 4.6.

4.4.1.2 Covariance Matrix

Here we briefly describe how the covariance matrix that appears in eq. (4.8) is computed. We rely on analytical estimates for our fiducial covariance computation, which were validated against covariances estimated from mocks [50]. We use the CosmoLike code to compute these analytical covariance matrices, see [106, 107, 108]. The CosmoLike code is a package for calculating cosmological likelihoods, which is used to analyze and interpret cosmological data, particularly from large-scale

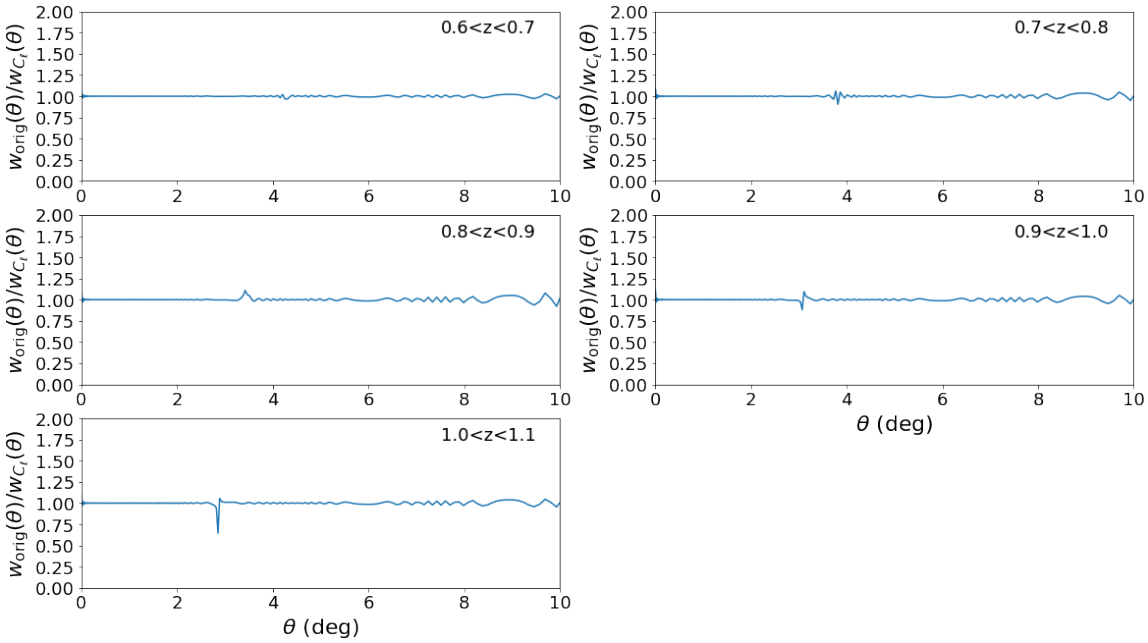

Figure 4.5: Ratio of the original template $w(\theta)$ and the one recovered from the C_ℓ. The largest oscillations happen on scales where the denominator intersects with zero. The original template was computed using Planck cosmology.

structure surveys. It includes routines for computing various quantities related to cosmological observables, such as power spectra, correlation functions, and covariance matrices. We include redshift-space distortions through the C_ℓ of eq. (2.92), and also correct the shot-noise contribution to the covariance (the term $\propto 1/n_g$) by taking into account the effect of the survey geometry on the number of galaxies in each angular bin of width $\Delta\theta$. As we already explained in section 2.4.3 of Chapter 2, the covariance matrix of our estimator of the angular correlation function, $\mathrm{cov}(\hat{w}(\theta), \hat{w}(\theta'))$, is related to the covariance of our estimator of the angular power spectrum, $\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'})$, by eq. (2.96),

$$\mathrm{cov}(\hat{w}(\theta), \hat{w}(\theta')) = \sum_{\ell\ell'} \frac{(2\ell+1)(2\ell'+1)}{(4\pi)^2} \mathcal{L}_\ell(\cos\theta)\mathcal{L}_{\ell'}(\cos\theta')\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'}). \tag{4.39}$$

The covariance matrix $\mathrm{cov}(\hat{C}_\ell, \hat{C}_{\ell'})$ can then be split into a Gaussian term and a non-Gaussian term, and the Gaussian one is given by eq. (2.92). Unless stated otherwise, we only include the Gaussian term in every covariance matrix computed throughout this chapter, since we tested that the non-Gaussian contribution does not impact our results, but makes the calculation slower.

4.4.1.3 Default Settings

The BAO-fitting pipeline we developed allows us to perform a wide variety of different fits, since there are several different settings that can be modified. However, the default settings are:

- θ ranges from (0.5−5) deg, with a step of $\Delta\theta = 0.2$ deg. In the mocks we also test the impact of fitting the $w(\theta)$ in narrower θ bins, namely $\Delta\theta = 0.05,\ 0.1,\ 0.15$ deg.

- 3 broadband-term parameters. By default, the sum over i in eq. (4.12) takes the values $i = 0, 1, 2$. This means that we have one term that goes as a constant, another one that goes as θ^{-1} and another one that goes as θ^{-2}. However, we also test the impact of using less broadband-term parameters, and also the impact of using an extra one that goes as θ.

- Mice template for the mocks, Planck template for the data. The reason for this is that the underlying cosmology of the COLA mocks is Mice. However, we also test the impact of using Planck as the reference cosmology for the mocks.

- CosmoLike covariance matrix, calculated using Mice cosmology for the mocks and Planck cosmology for the data. By default, we use the CosmoLike analytical covariance matrix to run the fits. However, we also test the impact of using the covariance matrix computed from the mocks instead.

- All the redshift bins are fitted simultaneously. However, we also perform the BAO fits removing one redshift bin at a time, and also removing all of them expect for one.

4.4.2 Cosmological Parameter Inference

Running the BAO-fitting pipeline on the data allows us to compute the BAO likelihood, which can in turn be used to infer cosmological parameters. Inferring cosmological information by fitting the measurements to different models is one of the most important steps in any cosmological analysis. Let's suppose that we have measured some data $\boldsymbol{D}$, e.g., $\boldsymbol{D} = \{d_M(z)/r_d, d_H(z)/r_d\}$ for our BAO probe or $\boldsymbol{D} \equiv \{w^i(\theta), \gamma_t^{ij}(\theta), \xi_\pm^{ij}(\theta)\}$ for the 3×2pt one, and that we have a model M to describe the data, e.g., the ΛCDM model. Our model depends on a set of physically-relevant parameters, $\boldsymbol{\Theta}$, but also on a set of nuisance parameters, $\boldsymbol{\alpha}$. In probabilistic inference, the posterior PDF, or PDF for the parameters $\{\boldsymbol{\Theta}, \boldsymbol{\alpha}\}$ given the data $\boldsymbol{D}$, is calculated via the Bayes' Theorem, which states that

$$\mathcal{P}(\boldsymbol{\Theta}, \boldsymbol{\alpha}|\boldsymbol{D}) = \frac{1}{p(\boldsymbol{D})}\mathcal{L}(\boldsymbol{D}|\boldsymbol{\Theta}, \boldsymbol{\alpha})p(\boldsymbol{\Theta}, \boldsymbol{\alpha}), \tag{4.40}$$

where $\mathcal{L}(\boldsymbol{D}|\boldsymbol{\Theta}, \boldsymbol{\alpha})$ is the likelihood function and $p(\boldsymbol{\Theta}, \boldsymbol{\alpha})$ is the prior distribution. The likelihood is usually assumed to be Gaussian, and under this assumption it has the form

$$\log \mathcal{L}(\boldsymbol{D}|\boldsymbol{\Theta}, \boldsymbol{\alpha}) = -\frac{1}{2} \left[\boldsymbol{D} - \boldsymbol{T}_M(\boldsymbol{\Theta}, \boldsymbol{\alpha})\right]^T \mathrm{cov}^{-1} \left[\boldsymbol{D} - \boldsymbol{T}_M(\boldsymbol{\Theta}, \boldsymbol{\alpha})\right], \tag{4.41}$$

where $\boldsymbol{T}_M(\boldsymbol{\Theta}, \boldsymbol{\alpha})$ is the prediction of our model M. Both the likelihood function and the prior distribution can be computed for any value of $\{\boldsymbol{\Theta}, \boldsymbol{\alpha}\}$, even though it can be hard and time-consuming. The normalization factor $p(\boldsymbol{D})$ is known as the evidence, and it is generally very expensive to compute. However, it does not depend on the parameters $\{\boldsymbol{\Theta}, \boldsymbol{\alpha}\}$, which yields to

$$\frac{\mathcal{P}(\boldsymbol{\Theta}_A, \boldsymbol{\alpha}_A|\boldsymbol{D})}{\mathcal{P}(\boldsymbol{\Theta}_B, \boldsymbol{\alpha}_B|\boldsymbol{D})} = \frac{\mathcal{L}(\boldsymbol{D}|\boldsymbol{\Theta}_A, \boldsymbol{\alpha}_A) p(\boldsymbol{\Theta}_A, \boldsymbol{\alpha}_A)}{\mathcal{L}(\boldsymbol{D}|\boldsymbol{\Theta}_B, \boldsymbol{\alpha}_B) p(\boldsymbol{\Theta}_B, \boldsymbol{\alpha}_B)}. \tag{4.42}$$

Therefore, it is possible to sample from $\mathcal{P}(\boldsymbol{\Theta}, \boldsymbol{\alpha}|\boldsymbol{D})$ without computing the evidence. In practice, what we compute when running a Markov chain Monte Carlo (hereafter, MCMC) are ratios of the PDF at pairs of points, but not the precise value at any individual point. This normalization-insensitive property of MCMC is one of its main advantages.

The priors $p(\boldsymbol{\Theta}, \boldsymbol{\alpha})$ play an important role in Bayesian inference. It is necessary to write them down because we are interested in running MCMCs, and all that MCMCs do is to draw samples from a probability distribution, and we want that to be the probability distribution for our parameters. Therefore, we cannot draw parameter samples directly from our likelihood function, since this likelihood function represents a probability distribution over datasets. In other words, we can draw representative datasets from the likelihood function but cannot draw parameter samples.

4.4.2.1 Markov Chain Monte Carlo Methods

Probabilistic data analysis involve computing and using either the posterior PDF for the parameters of the model or the likelihood function. It is sometimes sufficient to find the maximum of one of these, but it is often necessary to understand the posterior PDF in detail, i.e., to explore the parameter space around its maximum. MCMC methods are designed to sample from the posterior PDF efficiently, especially in parameter spaces with large numbers of dimensions. The process to sample this N-dimensional parameter space is known as chain, and has the following steps:

1. Draw values for the N parameters $\{\boldsymbol{\Theta}, \boldsymbol{\alpha}\}$ and store them.
2. Compute the theoretical prediction for these parameters, $\boldsymbol{T}_M(\boldsymbol{\Theta}, \boldsymbol{\alpha})$.
3. Compute the posterior value for each parameter given the observed data and the priors.

These steps are repeated until the posterior distribution has been populated with the desired density.

4.4.2.2 Marginalization over Parameters

One of the most important parts of Bayesian data analysis is the possibility of marginalizing over nuisance parameters. Marginalization is the process of integrating over all possible values of a given parameter, which propagates the effects of uncertainty about its value into the final result. Mathematically, the marginalized probability function of the parameters $\boldsymbol{\Theta}$ given the data $\boldsymbol{D}$ is given by

$$\mathcal{P}(\boldsymbol{\Theta}|\boldsymbol{D}) = \int d\boldsymbol{\alpha}\mathcal{P}(\boldsymbol{\Theta}, \boldsymbol{\alpha}|\boldsymbol{D}). \tag{4.43}$$

Once the N samples have been produced by MCMC, we can obtain the marginalized constraints on $\boldsymbol{\Theta}$ as the histogram of the samples projected into the parameter subspace spanned by $\boldsymbol{\Theta}$. In particular, we can approximate the expected value of a function $f(\boldsymbol{\Theta})$ as

$$\mathrm{E}_{\mathcal{P}(\boldsymbol{\Theta}|\boldsymbol{D})}\left[f(\boldsymbol{\Theta})\right] = \int d\boldsymbol{\Theta} f(\boldsymbol{\Theta})\mathcal{P}(\boldsymbol{\Theta}|\boldsymbol{D}) \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\Theta}_i). \tag{4.44}$$

In general, generating the N samples $\{\boldsymbol{\Theta}_i\}_{i=1}^{N}$ is a non-trivial process. The sampling is said to be good when any expectation value of interest is accurately computed via the approximation given in eq. (4.44). The word "accurately" here is related to theorems about limiting behavior: the general idea is that the sampling approximation becomes exact as N goes to infinity.

4.5 Testing the BAO-Fitting Pipeline

The main objective here is to test the BAO-fitting pipeline we developed in section 4.4, which is an important part of this thesis. In order to do so, we use the Y3 COLA mocks, which are the fiducial sets of simulations used for the DES BAO analysis, both in the Y3 and the Y6. As we already mentioned, we use two different reference cosmologies: Mice and Planck, whose cosmological parameters can be found listed in Table 4.2. The results using both reference cosmologies must be consistent in several different ways:

- since the underlying cosmology of the COLA mocks is Mice, we expect to obtain $\alpha = 1$, within errors, when using Mice as the reference cosmology. The reason for this is that α is the shift in the BAO peak of our data with respect to the reference cosmology, and there is no shift is both of them share exactly the same underlying cosmology, see eq. (4.10).

- the BAO distance measurement must be consistent among them. This means that we should obtain the same $d_M(z_{\mathrm{eff}})/r_d$ and with the same precision, even though the best-fit α and its error should not be the same (however, the relative

error, σ_α/α, should also be consistent). This is an important test, since when running on the data we do not know the exact underlying cosmology, and therefore we do not want our result to depend on the reference cosmology chosen to generate our template.

- the accuracy of the BAO measurement must decrease as we remove any number of redshift bins from the fits, i.e., removing information from any bin should increase the error in the measurement. The decrease in the accuracy must also be consistent between both templates.

On apart from these tests, we also run several more varying the default settings for the BAO fits, see section 4.4.1.3 for a complete list of the default settings. This allows us to further validate our pipeline, since we expect it to give consistent results when varying these settings.

In order to perform the BAO fits, we need to first measure the clustering signal for the 1,952 COLA mocks, i.e., measure the angular correlation function, or $w(\theta)$, for each mock (and for each redshift bin). We use the Landy-Szalay estimator, which we already introduced in eq. (2.95), implemented in the software CUTE[7]. The angular correlation functions for the 5 redshift bins, together with the mean $w(\theta)$, are shown in Fig. 4.6. We also include the three pre-unblinding clustering measurements at low angles used as input to the COLA code. As we discussed in section 4.1.1, these measurements at such low scales do not compromise our blinding protocol, since they contain no BAO information. The $w(\theta)$ covariance from the mocks can be computed as

$$\begin{aligned} &\mathrm{cov}^{\mathrm{zbin}_1,\mathrm{zbin}_2}(\theta_i,\theta_j) \\ &= \frac{1}{N-1}\sum_{n=1}^{N}\left[w_n^{\mathrm{zbin}_1}(\theta_i)-\overline{w}^{\mathrm{zbin}_1}(\theta_i)\right]\left[w_n^{\mathrm{zbin}_2}(\theta_j)-\overline{w}^{\mathrm{zbin}_2}(\theta_j)\right], \end{aligned} \tag{4.45}$$

where $N = 1,952$. The index zbin accounts for the 5 redshift bins, while the index i accounts for the angular binning of $w(\theta)$. The average here indicates mean over the 1,952 realizations. For visualization purposes, it is particularly useful to define the correlation matrix, or normalized covariance matrix, as

$$\mathrm{corr}^{\mathrm{zbin}_1,\mathrm{zbin}_2}(\theta_i,\theta_j) = \frac{\mathrm{cov}^{\mathrm{zbin}_1,\mathrm{zbin}_2}(\theta_i,\theta_j)}{\sqrt{\mathrm{cov}^{\mathrm{zbin}_1,\mathrm{zbin}_1}(\theta_i,\theta_i)}\sqrt{\mathrm{cov}^{\mathrm{zbin}_2,\mathrm{zbin}_2}(\theta_j,\theta_j)}}. \tag{4.46}$$

By definition, the correlation matrix takes values in the interval $[-1, 1]$. The correlation matrix of the $w(\theta)$ of the COLA mocks is shown in Fig. 4.7 (upper-triangle region). However, as we mentioned earlier, we do not use the COLA covariance as our default covariance when running our BAO-fitting pipeline. We rather rely of

[7] https://github.com/damonge/CUTE

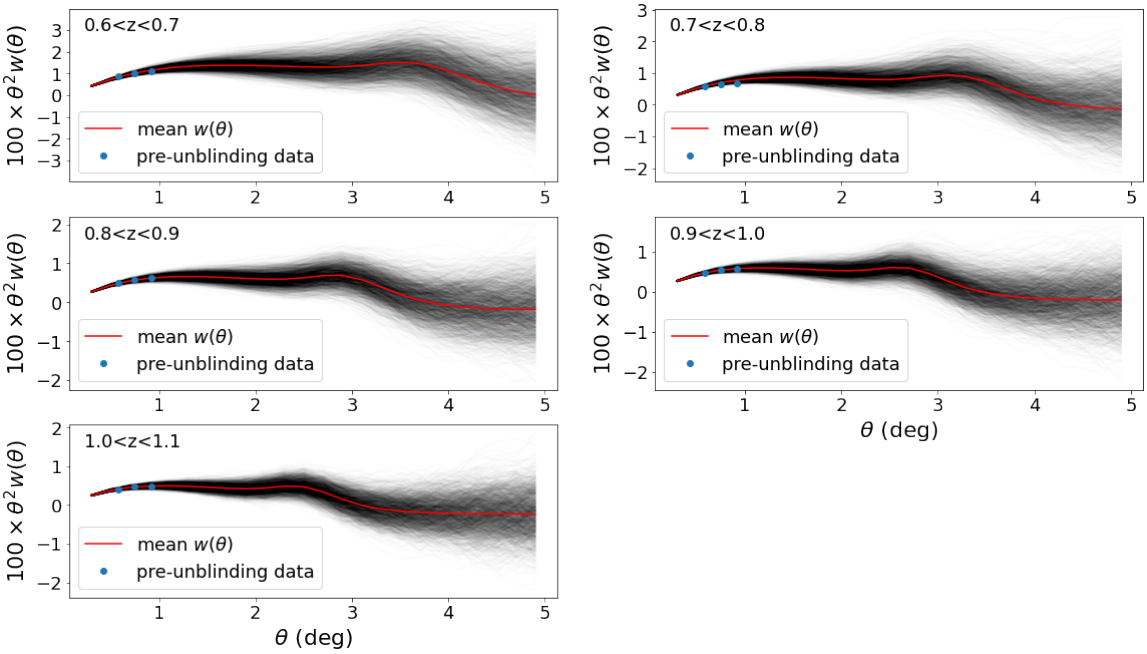

Figure 4.6: Angular correlation functions of the 1,952 COLA mocks (black curves) and mean $w(\theta)$ (red curve) for the 5 tomographic redshift bins. We also show the three pre-unblinding clustering measurements at low angles used as input to the COLA code. As we already discussed, these measurements at such low scales do not compromise our blinding protocol, since they contain no BAO information.

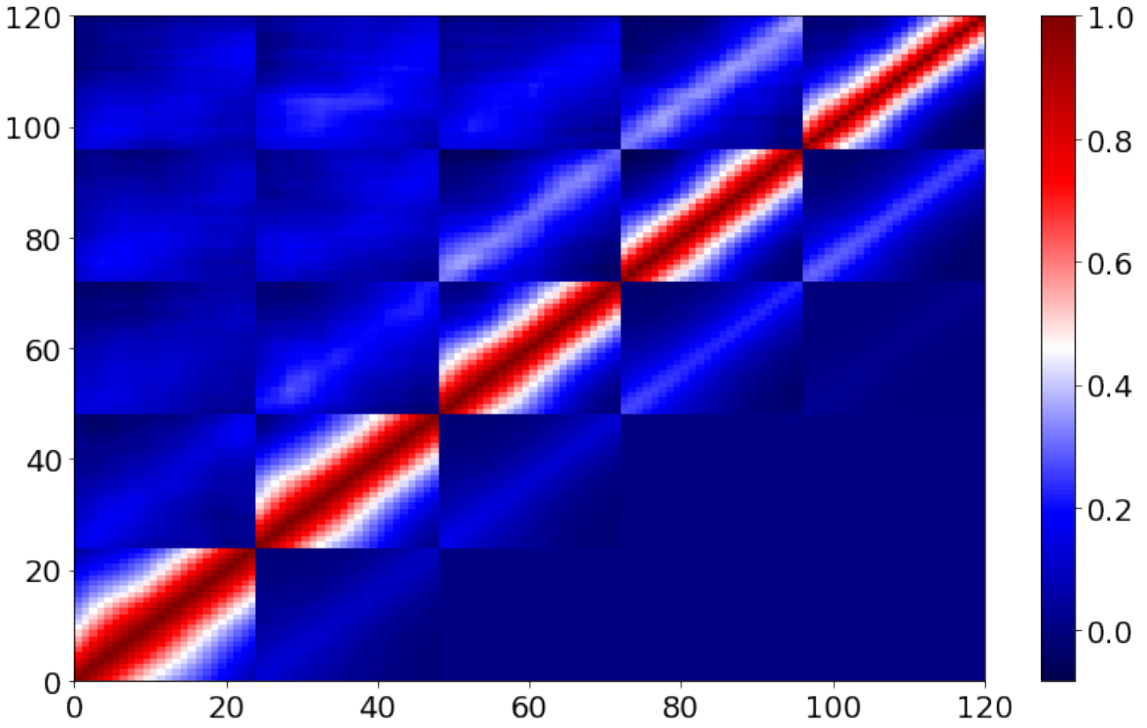

Figure 4.7: Normalized covariance matrix of the Y3 COLA mocks (upper-triangle region) and CosmoLike (lower-triangle region). In the case of CosmoLike, we show the one computed for Planck cosmology. We find that, as expected, their structure is very similar in the regions close to the diagonal, but they differ as we move to cross-correlations between more distant redshift bins. In the case of CosmoLike, the cross-correlations between non-contiguous redshift bins are negligible, whereas in the case of COLA they are not (partly due to the replications we described in section 4.3).

analytical covariances, computed using the CosmoLike code, see section 4.4.1.2 for further information. The correlation matrix for CosmoLike is also shown in Fig. 4.7 (lower-triangle region). We find that, in the case of CosmoLike, the cross-correlations between non-contiguous redshift bins are negligible, whereas in the case of COLA, they are not (this is partly due to the replications present in the COLA mocks, which were briefly described in section 4.3). These higher cross-correlation features of the COLA covariance is one of the reasons why we use the CosmoLike covariance matrix to perform the BAO fits on the Y3 COLA mocks and also on the Y3 sample. However, we also tested the effect of using the COLA covariance instead, and found that the effect is negligible in both the measurement and its error, and also the χ^2 value [50].

Now that we have measured the clustering signal for the 1,952 COLA mocks and also computed the covariance matrix, we are in good shape to test our BAO-fitting

Bins	$\langle\alpha\rangle$	$\sigma_{\rm std}$	σ_{68}	$\langle\sigma_\alpha\rangle$	f. enc.$\langle\alpha\rangle$	$\langle d_{\rm norm}\rangle$	$\sigma_{d_{\rm norm}}$	$\langle\chi^2\rangle$/dof
All	1.005	0.024	0.024	0.022	63.5%	-0.023	1.096	93.2/89
2345	1.005	0.027	0.025	0.024	65.2%	-0.022	1.114	74.3/71
1345	1.005	0.027	0.025	0.024	65.0%	-0.020	1.114	75.0/71
1245	1.005	0.027	0.026	0.024	64.3%	-0.014	1.110	74.5/71
1235	1.005	0.027	0.025	0.024	65.3%	-0.027	1.093	74.2/71
1234	1.004	0.026	0.025	0.024	64.8%	-0.023	1.091	73.7/71
1	1.002	0.051	0.051	0.049	66.5%	-0.026	1.097	17.7/17
2	1.002	0.049	0.047	0.046	66.6%	-0.021	1.124	17.0/17
3	1.002	0.046	0.043	0.041	66.4%	-0.016	1.143	17.5/17
4	1.005	0.045	0.042	0.040	66.9%	-0.029	1.155	17.8/17
5	1.008	0.049	0.046	0.044	66.5%	-0.039	1.152	18.3/17

Table 4.3: BAO fit results for the 1,952 COLA mocks using the Mice template and the CosmoLike analytical covariance matrix: mean and standard deviation of α, $\langle\alpha\rangle$ and $\sigma_{\rm std}$; width of the distribution, σ_{68}; mean error and fraction of mocks enclosed within it, $\langle\sigma_\alpha\rangle$ and f. enc.$\langle\alpha\rangle$; mean and the standard deviation of the pull statistics, $\langle d_{\rm norm}\rangle$ and $\sigma_{d_{\rm norm}}$; and mean χ^2 divided by the degrees of freedom. The results are shown for the cases of fitting the 5 redshift bins at a time, fitting removing one of them and fitting each of them individually.

pipeline. What our BAO-fitting pipeline does is to minimize the χ^2 from eq. (4.8) with the model given by eq. (4.12), following the methodology developed in section 4.4. As we mentioned earlier, we use two different reference cosmologies for our template $w(\theta)$, and then check the compatibility of the results and the robustness of the measurements.

4.5.1 Mice Template

When doing the BAO fits using the Mice template, we expect a distribution of α centered around 1, since the template has the same cosmology as the mocks. Therefore, the BAO peak in the $w(\theta)$ of the mocks must be in the same position as that of the template, which means that there is no shift between them, $\alpha = 1$. It is worth mentioning at this point that we actually expect a $\sim$0.4% deviation of our best-fit α with respect to 1 because of the non-linear evolution of the BAO feature due to the mode-coupling, which generates additional oscillations that are out of phase with those in the linear power spectrum [109].

In Table 4.3 we show the results obtained for the 1,952 COLA mocks using the Mice template in the case of fitting the 5 redshift bins at a time, fitting removing one of them and fitting each of them individually. From the results displayed in this table we conclude that $\alpha = 1$ within the error for all the cases. We find that $\langle\alpha\rangle$ is 1.005 for the case of fitting all the redshift bins simultaneously. This small bias with respect

to 1 is due to the non-linear evolution of the BAO feature we mentioned earlier (due to the mode-coupling), from which we expect a shift of about 0.4% with respect to 1. The standard deviation and the width of the distribution of α (the $\langle\sigma_{\rm std}\rangle$ and σ_{68} columns, respectively) have very similar values for all the cases, which is expected since both of them represent similar statistics. The relative error of the measurement for the case of fitting all the redshift bins is $\sigma_{\rm std}/\langle\alpha\rangle = 2.4\%$, where we used the standard deviation of the α instead of the mean error for its computation since it is a more reliable estimation of the error. The mean error is slightly smaller than $\sigma_{\rm std}$ and σ_{68} for all the cases, but still very similar, which means that our assumption that the likelihoods are Gaussian is very suitable. The fraction of mocks enclosed within $\langle\alpha\rangle \pm \langle\sigma_\alpha\rangle$ is, approximately, 68% for all the cases, which also reinforces our assumption that our likelihoods are Gaussian. The mean of the pull statistics is close to 0 and its standard deviation is close to 1, as expected, since we would expect it to be a Gaussian distribution with zero mean and unit variance. We also find that $\sigma_{\rm std}$, σ_{68} and $\langle\sigma_\alpha\rangle$ increase as we reduce the number of redshift bins included in the fit, i.e., as we reduce the degrees of freedom. This is also expected, since the less redshift bins we include, i.e., the less information we include, the larger the error of our measurement has to be. Finally, we find that the $\langle\chi^2\rangle$ are very similar to the degrees of freedom for all the different cases. The corresponding p-values are, approximately, within 0.35 and 0.45, indicating the goodness-of-fit for all of them.

4.5.2 Planck Template

When doing the BAO fits using the Planck template, we expect a distribution of α that is not centered around 1, since the template has a different cosmology to the one of the mocks. We will later compute the expected α for the case of using Planck as the reference cosmology, and compare its value with the BAO fit results (the actual value we expect, without taking into account the non-linear evolution of the BAO, is 0.960, see eq. (4.50)).

In Table 4.4 we display the BAO fit results for the 1,952 COLA mocks using the Planck template in the case of fitting the 5 redshift bins at a time, fitting removing one of them and fitting each of them individually. They can be directly compared to those obtained using Mice as the reference cosmology, which are displayed in Table 4.3. The conclusions here are similar to the ones we came to in the case of using the Mice template. However, in the case of using the Planck template we find that the mean α is quite different from one, as we expected. The ratio between $\alpha_{\rm Planck}$ and $\alpha_{\rm Mice}$, using eq. (4.10), is given by

$$\begin{aligned}\frac{\alpha^{\rm Planck}(z_{\rm eff})}{\alpha^{\rm Mice}(z_{\rm eff})} &= \left[\frac{r_d^{\rm Planck}/d_M^{\rm Planck}(z_{\rm eff})}{r_d/d_M(z_{\rm eff})}\right]\left[\frac{r_d^{\rm Mice}/d_M^{\rm Mice}(z_{\rm eff})}{r_d/d_M(z_{\rm eff})}\right]^{-1} \\ &= \frac{r_d^{\rm Planck}/d_M^{\rm Planck}(z_{\rm eff})}{r_d^{\rm Mice}/d_M^{\rm Mice}(z_{\rm eff})}.\end{aligned} \tag{4.47}$$

Bins	$\langle\alpha\rangle$	$\sigma_{\rm std}$	σ_{68}	$\langle\sigma_\alpha\rangle$	f. enc.$\langle\alpha\rangle$	$\langle d_{\rm norm}\rangle$	$\sigma_{d_{\rm norm}}$	$\langle\chi^2\rangle$/dof
All	0.966	0.023	0.022	0.021	64.1%	-0.023	1.089	93.6/89
2345	0.968	0.026	0.024	0.023	65.4%	-0.026	1.118	74.6/71
1345	0.967	0.026	0.025	0.023	65.1%	-0.022	1.118	75.3/71
1245	0.966	0.026	0.025	0.023	64.9%	-0.014	1.104	74.9/71
1235	0.966	0.026	0.024	0.023	66.0%	-0.030	1.086	74.5/71
1234	0.964	0.026	0.024	0.023	65.6%	-0.026	1.081	74.1/71
1	0.964	0.050	0.048	0.048	68.1%	-0.066	1.080	17.9/17
2	0.965	0.047	0.045	0.045	68.5%	-0.044	1.091	17.2/17
3	0.966	0.044	0.041	0.040	66.1%	-0.041	1.139	17.5/17
4	0.970	0.044	0.040	0.039	67.1%	-0.045	1.159	17.9/17
5	0.976	0.048	0.045	0.042	65.6%	-0.065	1.187	18.3/17

Table 4.4: BAO fit results for the 1,952 COLA mocks using the Planck template and the CosmoLike analytical covariance matrix: mean and standard deviation of α, $\langle\alpha\rangle$ and $\sigma_{\rm std}$; width of the distribution, σ_{68}; mean error and fraction of mocks enclosed within it, $\langle\sigma_\alpha\rangle$ and f. enc.$\langle\alpha\rangle$; mean and the standard deviation of the pull statistics, $\langle d_{\rm norm}\rangle$ and $\sigma_{d_{\rm norm}}$; and mean χ^2 divided by the degrees of freedom. The results are shown for the cases of fitting the 5 redshift bins at a time, fitting removing one of them and fitting each of them individually.

Using the cosmological parameters displayed in Table 4.2, we can compute the theoretical prediction for the sound scale horizons and the comoving angular diameter distances, and we find

$$r_d^{\rm Mice} = 153.43 \text{ Mpc}, \quad r_d^{\rm Planck} = 147.66 \text{ Mpc} \tag{4.48}$$

and

$$d_M^{\rm Mice}(z_{\rm eff}) = 2961.81 \text{ Mpc}, \quad d_M^{\rm Planck}(z_{\rm eff}) = 2969.09 \text{ Mpc}. \tag{4.49}$$

Therefore,

$$\frac{\alpha^{\rm Planck}(z_{\rm eff})}{\alpha^{\rm Mice}(z_{\rm eff})} = \frac{r_d^{\rm Planck}/d_M^{\rm Planck}(z_{\rm eff})}{r_d^{\rm Mice}/d_M^{\rm Mice}(z_{\rm eff})} = 0.960. \tag{4.50}$$

On the other hand, the value for this ratio obtained from the BAO-fit results, i.e., using Tables 4.3 and 4.4, is

$$\frac{\langle\alpha^{\rm Planck}\rangle}{\langle\alpha^{\rm Mice}\rangle} = \frac{0.966}{1.005} = 0.961. \tag{4.51}$$

The theoretical prediction, eq. (4.50), and the value obtained from the fits, eq. (4.51), are, approximately, the same. This implies that the results using the Mice template and those using the Planck one encode the same BAO distance measurement,

$$\frac{d_M(z_{\rm eff})}{r_d}, \tag{4.52}$$

which is the physical quantity that we measure. It is also useful to compute $d_M(z_{\rm eff})/r_d$ for Planck cosmology, since Planck is the reference cosmology we use when running on the data,

$$\frac{d_M^{\rm Planck}(z_{\rm eff})}{r_d^{\rm Planck}} = 20.11. \tag{4.53}$$

Continuing our comparison of the results using Mice and Planck templates, we find that the value for $\sigma_{\rm std}$ is smaller in the case of using the Planck template, which is expected since $\langle\alpha\rangle$ is also smaller. However, the relative error, i.e., the ratio of both quantities, $\langle\sigma_\alpha\rangle/\langle\alpha\rangle$, is similar to the one we obtained in the case of the Mice template: we obtain a ∼2.4% for both templates in the case of fitting all the redshift bins. This relative error is also consistent in the case of removing any number of redshift bins from the fits. We obtain a relative error in the BAO measurement of ∼2.7% in the case of removing one redshift bin (any of them); ∼5.1% for the case of fitting the first redshift bin alone; ∼4.9% for the second one; ∼4.6% for the third one; ∼4.5% for the fourth one; and ∼4.8% for the last one. Therefore, the precision of the BAO measurement does not depend on the reference cosmology used to compute the template. Finally, as in the case of the Mice template, we also find that the $\langle\chi^2\rangle$ are very similar to the degrees of freedom, with the corresponding p-values ranging within 0.35 and 0.45 for all the cases.

4.5.3 Variations of the Default Settings

Here we test the impact of varying the default settings of the analysis in the BAO fit results, which we listed in 4.4.1.3. We vary, among other things, the number of broadband-term parameters or the covariance used to perform the fits. In Table 4.5 we show the BAO fit results for the 1,952 COLA mocks with different configurations and variations of the analysis. In the first 9 rows we display the results for the cases of using different numbers of broadband-term parameters ($i = 0$, $i = 0, 1$, $i = 0, 1, 2$, $i = -1, 0, 1, 2$), being $i = 0, 1, 2$ the default case (bold row); using different bin sizes in θ for the angular correlation functions ($\Delta\theta = 0.05, 0.1, 0.15$ deg), being $\Delta\theta = 0.2$ deg the default value; and changing the minimum θ, being $\theta_{\rm min} = 0.5$ deg its default. The default case, i.e., the bold one, is the same as the one displayed in the first row of Table 4.3. We find that changing these settings does not affect the position of the BAO peak, and does not affect the precision as well, i.e., the columns α, $\sigma_{\rm std}$, σ_{68} and $\langle\sigma_\alpha\rangle$ share the same values. The row named "COLA cov." shows the results obtained using the COLA covariance instead of the CosmoLike one. The "Planck cov." entry is for the case of using a CosmoLike covariance matrix computed for the Planck cosmology (and also fitting with Planck template). Finally, the last 4 rows of the table show the results for the cases of using different numbers of broadband-term parameters but for the Planck template (the "Pla. $i = 0, 1, 2$" entry is the same case as the one displayed in the first row of Table 4.4). All the $\langle\chi^2\rangle$ values are consistent with the degrees of freedom, with their p-values ranging, approximately, within 0.15

case	$\langle\alpha\rangle$	$\sigma_{\rm std}$	σ_{68}	$\langle\sigma_\alpha\rangle$	f. enc.$\langle\alpha\rangle$	$\langle d_{\rm norm}\rangle$	$\sigma_{d_{\rm norm}}$	$\langle\chi^2\rangle$/dof
$i=0$	1.002	0.022	0.021	0.021	66.7%	0.012	1.051	101.6/99
$i=0,1$	1.004	0.024	0.023	0.022	64.7%	-0.025	1.077	97.6/94
$\boldsymbol{i=0,1,2}$	**1.005**	**0.024**	**0.024**	**0.022**	**63.5%**	**-0.023**	**1.096**	**93.2/89**
$i=-1,0,1,2$	1.005	0.024	0.023	0.022	63.4%	-0.026	1.102	88.4/84
$\Delta\theta=0.05$ deg	1.005	0.024	0.023	0.022	64.4%	-0.018	1.080	398.6/429
$\Delta\theta=0.1$ deg	1.005	0.024	0.023	0.022	64.6%	-0.018	1.086	198.2/204
$\Delta\theta=0.15$ deg	1.005	0.024	0.023	0.022	63.8%	-0.018	1.090	130.9/129
$\theta_{\rm min}=1$ deg	1.005	0.024	0.024	0.022	64.1%	-0.028	1.098	82.9/79
COLA cov.	1.004	0.024	0.023	0.022	65.7%	-0.023	1.042	91.4/89
Planck cov.	0.966	0.023	0.023	0.025	73.5%	-0.020	0.910	72.1/89
Pla. $i=0$	0.948	0.022	0.022	0.021	66.4%	0.048	1.042	111.6/99
Pla. $i=0,1$	0.964	0.024	0.023	0.024	69.5%	-0.019	0.975	102.6/94
Pla. $i=0,1,2$	0.966	0.023	0.022	0.021	64.1%	-0.023	1.089	93.6/89
Pla. $i=-1,0,1,2$	0.966	0.024	0.023	0.021	64.4%	-0.029	1.099	88.8/84

Table 4.5: BAO fit results for the 1,952 COLA mocks with different configurations and variations of the analysis. In the first 9 rows we display the results for the cases of using different numbers of broadband-term parameters ($i=0$, $i=0,1$, $i=0,1,2$, $i=-1,0,1,2$), being $i=0,1,2$ the default case (bold row); using different bin sizes in θ for the angular correlation functions ($\Delta\theta=0.05, 0.1, 0.15$ deg), being $\Delta\theta=0.2$ deg the default value; and changing the minimum θ, being $\theta_{\rm min}=0.5$ deg its default. The next entry shows the results obtained using the COLA covariance instead of the CosmoLike one (all the other cases are fitted using the CosmoLike covariance). The "Planck cov." entry is for the case of using a CosmoLike covariance computed for the Planck cosmology. In the last 4 rows we display the results for the cases of using different number of broadband-term parameters but for the Planck template.

and 0.85. The worst-case fit out of all the different cases displayed in this table is for the entry labeled as "Pla. $i=0$": we find that one broadband-term parameter is not enough to accurately perform the fits using the Planck template (the $\langle\chi^2\rangle$ column is also larger than the same case using the Mice template). However, even this case has a reasonable p-value of 0.18, and therefore we conclude that the results of our pipeline are robust against variations in the default settings of the analysis.

4.5.4 Summary of the Results

From the results displayed in Tables 4.3 and 4.4, we conclude that:

- For the fits using all the redshift bins, we have obtained a mean error of the 1,952 α values of $\sim$2.4% with respect to the mean α. This result does not depend on the reference cosmology of our template: we obtain the same relative

error for both Mice and Planck templates. Therefore, the precision of the BAO measurement does not depend on the cosmology chosen to generate the template $w(\theta)$.

- As expected, using the Mice template gives $\langle\alpha\rangle \approx 1$, while using the Planck one gives $\langle\alpha\rangle \approx 0.966$. If we compute the expected α for Planck, we find that the theoretical prediction agrees very well with the result obtained from the fits. Therefore, we conclude the algorithm is robust - the results are independent of the cosmology used for the template, both the distance measurement itself and its precision.

- For the fits removing any number of redshift bins, we have found that the more redshift bins removed, the larger $\sigma_{\rm std}$, σ_{68} and $\langle\sigma_\alpha\rangle$ are, which is an expected result. This means that the error is always increasing by removing the information from one or more bins.

- The fraction of mocks enclosed within $\langle\alpha\rangle \pm \langle\sigma_\alpha\rangle$ is, approximately, 68%. This holds for all the different cases shown in Tables 4.3 and 4.4. This means that our likelihoods are, in a very good approximation, Gaussian.

- The distribution of the best-fit χ^2 for the 1,952 mocks follow, in a very good approximation, a χ^2 distribution with the corresponding degrees of freedom. Also, the mean χ^2 are very similar to the degrees of freedom, with their p-values ranging, approximately, within 0.35 and 0.45.

On the other hand, from the results displayed in Table 4.5 we find that our pipeline is very robust against variations in the default settings. The $\langle\alpha\rangle$ and $\sigma_{\rm std}$, as well as all the other quantities displayed in this table, are barely affected by changing these settings. Therefore, we conclude that our BAO-fitting pipeline has now been successfully validated.

4.6 BAO Measurement on the Y3 Data

We already described the Y3 BAO sample in section 4.2.2, in which we included information about how the sample is selected, how the angular mask looks, how the observational systematics are mitigated and how the photometric redshifts are obtained. In section 4.5 we applied out BAO-fitting pipeline to a set of 1,952 COLA mocks. We successfully performed several tests to check the robustness and accuracy of the code, and now we are in good shape to use it on the actual data. Firstly, we measure the clustering signal of our BAO sample. Then, we run the BAO-fitting pipeline on the angular correlation function of the data. Finally, we discuss about the cosmological implications of our measurement.

In the fiducial Y3 analysis we had to pass several pre-unblinding tests before unblinding, as we mentioned in section 4.1.1 (see [50] to find the complete battery of

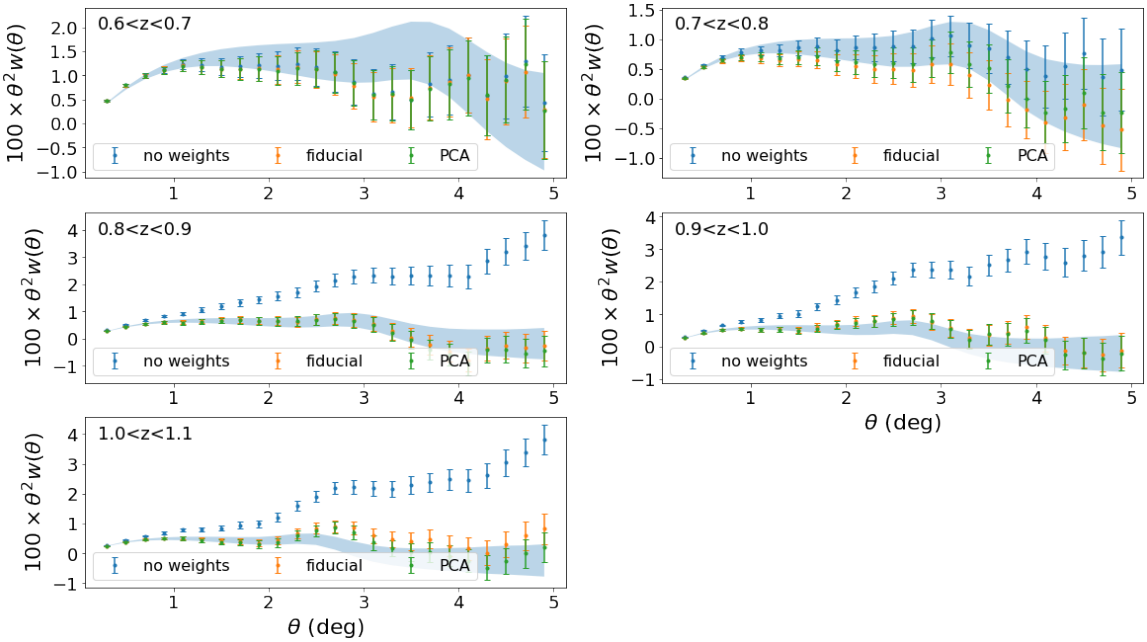

Figure 4.8: Angular correlation function of the Y3 BAO sample, computed using the Landy-Szalay estimator with the software TreeCorr [110]. The cases of not correcting the systematics (blue), correcting with the fiducial weights (orange) and correcting with the alternative PCA weights (green) are shown. The errorbars come from the CosmoLike covariance matrix. The shaded region corresponds to the 1-σ region of the 1,952 COLA mocks. The angular separation θ goes from 0.5 to 5 deg, with $\Delta\theta = 0.2$ deg (for the 5 redshift bins).

tests run). For instance, one of these pre-unblinding tests consists of checking if the BAO is detected: we consider to have a detection if the 1-σ region of our measured α lies within $[0.8, 1.2]$. In this thesis we do not go through these pre-unblinding tests, since we passed them all in the fiducial analysis, and therefore here we directly run our BAO-fitting pipeline on the data and show its results.

4.6.1 Clustering Signal of the Y3 BAO Sample

The first thing we need to do is to measure the angular correlation functions of the BAO sample for the 5 redshift bins. We do it similarly to what we did it for the COLA mocks, i.e., using the Landy-Szalay estimator [30], but this time with the TreeCorr Python package [110]. The main difference with respect to the computation in the mocks is that here we must also take into account the weights that come from the correction of the observational systematics. In practice, what we do is to weight each galaxy in our sample with its associated systematic weight [72]. The angular correlation functions of the Y3 BAO sample are shown in Fig. 4.8. We plot the cases of not correcting the systematics, correcting with the fiducial weights and correcting

with the alternative PCA weights (the two different sets of weights are explained in [72]). We also include the 1-σ region computed from the 1,952 COLA mocks. The angular range in which the $w(\theta)$ are plotted is the one used for the BAO fits, and goes from 0.5 to 5 deg for the 5 redshift bins, with $\Delta\theta = 0.2$ deg. The BAO feature moves towards lower angular scales (lower θ) as the redshift increases, reflecting its fixed comoving scale.

4.6.1.1 Projected Clustering Results

An alternative to the angular correlation function is to obtain an estimate of the 3-dimensional correlation function. In photometric surveys such as DES, it is not possible to directly estimate 3-dimensional correlation function since there is no measurement of the spectroscopic redshift of the galaxies, and hence it is not possible to calculate the distance between these objects in redshift space. However, if we approximate the redshift of each galaxy by its photometric redshift, $z_{\rm mean}$, and then we assume some cosmological parameters, we can obtain the distance to each galaxy. Here we assume a flat ΛCDM model with the same cosmological parameters as the Planck template, see Table 4.2, to transform photometric redshifts and angular positions to comoving distances. In Fig. 4.9 we show the 3-dimensional correlation function calculated in this way and projected along the LOS. It can be directly computed integrating the 3-dimensional $\xi(r_\perp, r_\parallel)$, or $\xi(s, \mu)$, as

$$w_\perp(r_\perp) = \int dr_\parallel \xi(r_\perp, r_\parallel). \tag{4.54}$$

The projected correlation function of the Y3 sample, plotted in Fig. 4.9, shows a higher clustering amplitude in the case in which the observational systematics are not corrected, as happened in the angular correlation function for most redshift bins, see Fig. 4.8, which is the expected behavior of a contaminated clustering signal. We remark that we only use the projected clustering for illustration purposes since it is the single correlation with highest signal-to-noise ratio, but we do not use it to derive distance measurements in this thesis.

4.6.2 BAO Fit Results

For the fits on the data, we use the Planck template and Planck CosmoLike covariance matrix, i.e., the fiducial cosmology we use to quote our primary results is Planck. The sound horizon scale and the comoving angular diameter distance at our effective redshift for Planck cosmology can be found in eqs. (4.48) and (4.49), respectively, from which we already obtained $d_M^{\rm Planck}(z_{\rm eff})/r_d^{\rm Planck} = 20.11$. In Table 4.6 we display the α parameter and its error, the χ^2/dof for the best-fit parameters and the corresponding p-value. We show all these results for the cases of fitting all the redshift bins simultaneously, removing one of them at a time and fitting each of them individually. We find that for the case of fitting the first redshift bin alone, there is

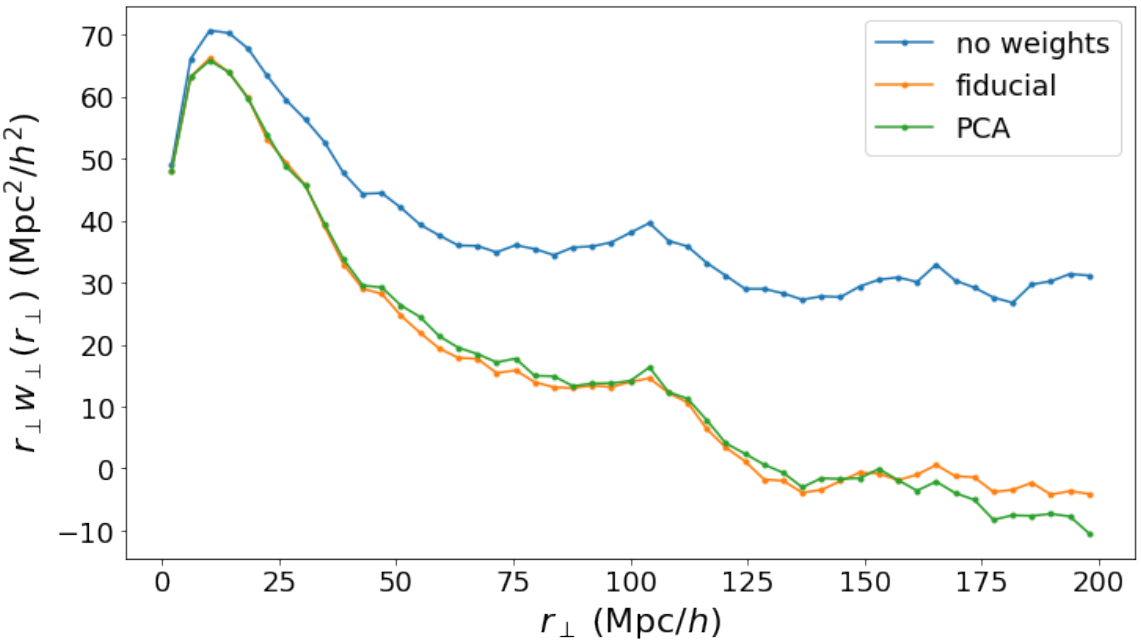

Figure 4.9: Projected two-point correlation function of the Y3 sample as a function of comoving separation perpendicular to the LOS, $r_\perp$, as obtained with NBODYKIT [111]. The positions of the galaxies were transformed into distances using their photometric redshifts and assuming Planck cosmology [14]. The cases of not correcting the systematics (blue), correcting with the fiducial weights (orange) and correcting with the alternative PCA weights (green) are shown.

Bins	α	σ_α	χ^2/dof	p-value
All	0.942	0.024	95.0/89	0.313
2345	0.953	0.025	68.2/71	0.572
1345	0.930	0.026	84.1/71	0.136
1245	0.938	0.027	75.3/71	0.342
1235	0.934	0.028	69.2/71	0.540
1234	0.967	0.027	81.0/71	0.194
1	-	-	-	-
2	1.000	0.050	10.2/17	0.895
3	0.978	0.047	19.7/17	0.288
4	0.978	0.040	23.2/17	0.143
5	0.903	0.036	10.8/17	0.865
All (no weights)	0.933	0.025	101.8/89	0.166
All (PCA weights)	0.930	0.022	104.4/89	0.126

Table 4.6: Best-fit α parameter and its error, together with the best-fit χ^2 of the fit divided by the degrees of freedom, and the corresponding p-value. The results are shown for the cases of fitting the 5 redshift bins at a time, fitting removing one of them and fitting each of them individually (for the fiducial systematic weights). We also display the results obtained when fitting all the redshift bins but without correcting the systematics and correcting them with the alternative PCA weights. The results for the case of fitting only the first redshift bin are not displayed in the table because the 1-σ region lies outside the interval $[0.8, 1.2]$, and therefore we consider it as a non-detection.

no BAO detection, i.e., the 1σ region of α does not lie within $[0.8, 1.2]$. This is not completely unexpected, since the probability for a non-detection in one redshift bin is 30%, as computed from the COLA mocks. The goodness-of-fit for all the cases are good, in the sense that the χ^2 is always similar to the number of degrees of freedom. Our primary results for the measurement of the BAO from the angular correlation function is, then,

$$\alpha(z_{\rm eff}) = 0.942 \pm 0.024. \tag{4.55}$$

In Fig. 4.10 we show the final BAO likelihood in terms of the α parameter, following the procedure explained in section 4.4. In Fig. 4.11 we show the BAO fit results for the individual tomographic bins. Also, the DES Y3 $w(\theta)$ and the best-fit model are shown in Fig. 4.12, together with the template $w(\theta)$ used to do the fits.

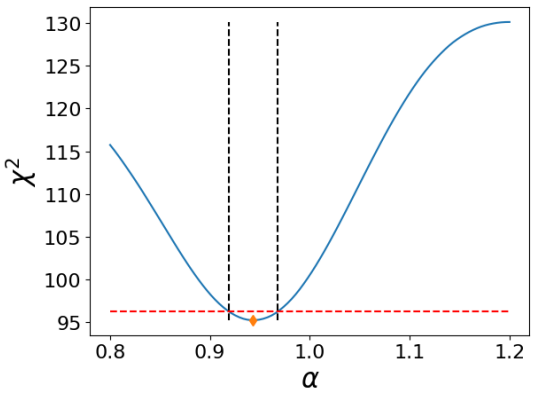

Figure 4.10: BAO likelihood for the DES Y3 BAO sample. The horizontal orange dashed line corresponds to $\Delta\chi^2 = 1$, whereas the vertical black dashed ones enclose the 1-σ region for α.

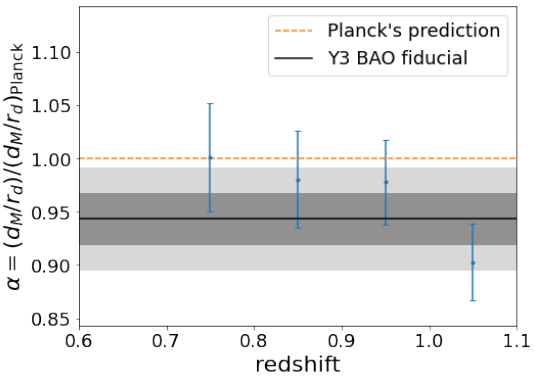

Figure 4.11: BAO fits for the individual tomographic bins. The shaded region corresponds to our Y3 measurement ($\alpha = 0.942 \pm 0.024$). We show the individual results for the fiducial analysis.

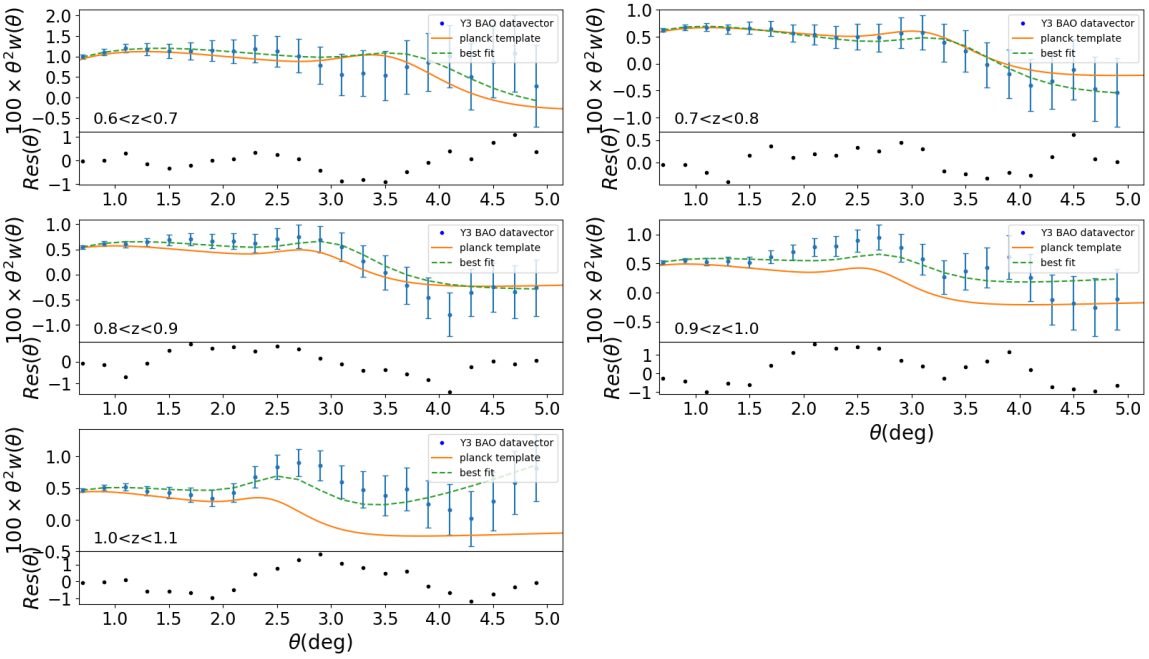

Figure 4.12: Angular correlation function of the Y3 BAO sample (blue), together with the BAO template for the Planck cosmology (orange) and the best-fit model (green).

4.6.2.1 Impact of Standard Weights for Systematics

We already know that the weights for systematics have little or no impact on the recovered BAO, see [83]. However, as a robustness test of the measurement of the BAO peak, in Table 4.6 we added the result of the BAO fit for the $w(\theta)$ computed without applying the weights, i.e., without correcting the systematics. The insensitivity to the weights demonstrates the robustness of the BAO fit results against the observational systematics.

4.6.2.2 Impact of PCA Weights for Systematics

A second method for determining the decontamination weights was applied in the DES Y3 analysis. This method consisted on doing a principal components analysis of the survey properties maps used as templates for the observational systematics, see [72]. A second set of weights (PCA50), came out from this analysis. The new weights had more impact over the clustering amplitude of the second and fifth redshift bins, but they essentially leave the BAO fit results unchanged.

As another robustness test, we performed the BAO fit for the $w(\theta)$ calculated with the PCA weights and added the result in Table 4.6. The PCA weights have their biggest impact over the clustering amplitude of the second and fifth redshift bins, but they essentially leave the BAO fit results unchanged.

4.6.3 Cosmological Implications of the BAO Measurement

Here we aim to discuss the cosmological implications of our BAO measurement. Using the Planck cosmology and its measurement of the sound horizon at the drag epoch, we can convert our best-fit α, displayed in Table 4.6, into d_M/r_d measurements at our effective redshift via eq. (4.10),

$$\frac{d_M(z_{\rm eff})}{r_d} = \frac{d_M^{\rm ref}(z_{\rm eff})}{r_d^{\rm ref}}\alpha(z_{\rm eff}) = 20.11 \times (0.942 \pm 0.024) = 18.94 \pm 0.48. \tag{4.56}$$

This result represents a fractional error of 2.6%, which is the most precise determination of the BAO scale ever done in a photometric galaxy redshift survey, and the most precise ever at this redshift as well. It is consistent with what we expected from the mocks, i.e., a relative error of ∼2.4% (see section 4.5).

In Fig. 4.13 we display the angular diameter distance measurement from the DES Y3 BAO sample normalized by the sound horizon scale compared to other $d_M(z)/r_d$ in the literature, and the prediction from Planck assuming a flat ΛCDM cosmological model (with fixed neutrino mass). We include measurements from the combined BOSS LOWz + CMASS galaxy samples (at $0.2 < z < 0.5$ and $0.4 < z < 0.6$) [112], from the eBOSS luminous red galaxies (LRG, $0.6 < z < 1.0$) [51, 38] and the eBOSS emission line galaxies (ELG, $0.6 < z < 1.1$) [113, 52], as well as from the eBOSS quasars ($0.8 < z < 2.2$) [53, 54] and Lyman-α auto-correlations ($z > 2.1$) [47]. These

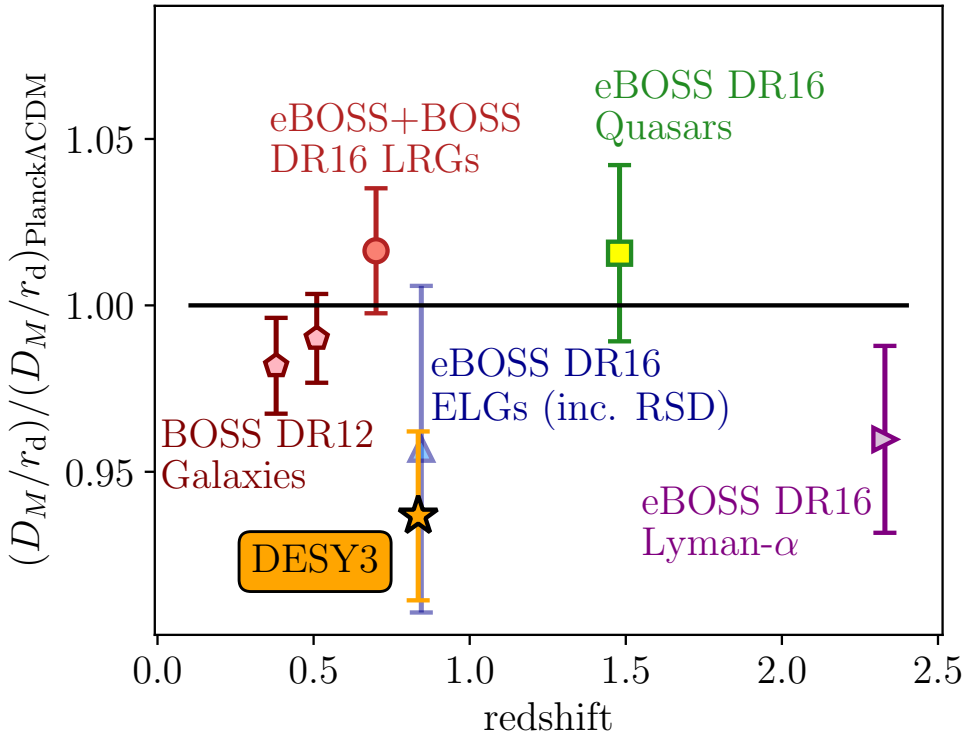

Figure 4.13: Ratio between the angular diameter distances measured using the BAO feature at different redshifts for several galaxy surveys and the prediction from the cosmological parameters determined by Planck. Plot taken from [50].

measurements and their cosmological implications are compiled and discussed in [35], and represent the most updated BAO distance ladder. Our BAO measurement is about 2.5σ away from Planck's prediction. This deviation from Planck is mainly due to the fifth redshift bin, which is the one that drives the final result to smaller α values, see Table 4.6. However, it passed all the pre-unblinding and post-unblinding consistency tests in the fiducial analysis, which essentially means this behaviour is statistically compatible with simulations, and therefore there is no reason to exclude it from the measurements, see [50].

The likelihood of our data given α can be written as

$$\mathcal{L}(\boldsymbol{D}_{\mathbf{DESY3}}|\alpha) \propto \exp\left(-\chi^2/2\right), \tag{4.57}$$

where the χ^2 profile is shown in Fig. 4.10. Since we are interested in obtaining constraints in some set of cosmological parameters $\boldsymbol{\Theta}$, we express α as a function of them. Explicitly,

$$\alpha = \alpha(\boldsymbol{\Theta}) = \frac{d_M(z_{\mathrm{eff}}, \boldsymbol{\Theta})}{r_d(\boldsymbol{\Theta})} \left[\frac{d_M^{\mathrm{ref}}(z_{\mathrm{eff}})}{r_d^{\mathrm{ref}}}\right]^{-1}. \tag{4.58}$$

We combine our likelihood with the likelihoods for the SDSS "main galaxy sample" (MGS), which was selected and analyzed from the first two generations of SDSS, SDSS-I and SDSS-II; the BOSS DR12 galaxies; and all the different eBOSS measurements, which can be found on the public SDSS SVN repository[8]. Therefore, our combined likelihood can be expressed as

$$\mathcal{L}_T(\boldsymbol{D}|\boldsymbol{\Theta}) = \mathcal{L}(\boldsymbol{D}_{\mathbf{DESY3}}|\boldsymbol{\Theta})\mathcal{L}(\boldsymbol{D}_{\mathbf{MGS}}|\boldsymbol{\Theta})\mathcal{L}(\boldsymbol{D}_{\mathbf{BOSS}}|\boldsymbol{\Theta})\mathcal{L}(\boldsymbol{D}_{\mathbf{eBOSS}}|\boldsymbol{\Theta}). \tag{4.59}$$

The posterior distribution of the cosmological parameters given the data is, then,

$$\mathcal{P}(\boldsymbol{\Theta}|\boldsymbol{D}) \propto \mathcal{L}_T(\boldsymbol{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}), \tag{4.60}$$

where $p(\boldsymbol{\Theta})$ are the priors for the cosmological parameters. Unless specified a prior on any cosmological parameter, by default we take

$$\begin{cases} 0.4 \leq h \leq 0.9, \\ 0.12 \leq \Omega_m \leq 0.5, \\ 0.01 \leq \Omega_b \leq 0.12 \end{cases} \tag{4.61}$$

as the priors for the ΛCDM chains and

$$\begin{cases} 0.4 \leq h \leq 0.9, \\ 0.12 \leq \Omega_m \leq 0.5, \\ 0.01 \leq \Omega_b \leq 0.12, \\ -0.03 \leq \Omega_k \leq 0.03 \end{cases} \tag{4.62}$$

for the oΛCDM ones. In order to constrain cosmological parameters, we run MCMCs to maximize the full posterior distribution, eq. (4.60), as we explained in section 4.4.2. We make use of the PYTHON package EMCEE in order to run the chains, see [114]. As for plotting the results of the chains, we use the PYTHON package GETDIST [115].

4.6.3.1 BAO Chains in ΛCDM

Here we run chains to obtain cosmological constraints for a ΛCDM model. We sample over H_0 (or h), Ω_m and Ω_b. As derived parameters, we also include r_d and $d_M(z_{\rm eff})/r_d$. We consider three different chains, similarly to what is done in [35]:

- BAO-only chain: This chain does only include the BAO data, i.e., we do not include any extra data and/or specific prior on any cosmological parameter (on apart from the flat priors we set previously). By BAO-only we mean the combination of SDSS' MGS, BOSS, eBOSS and our Y3 BAO likelihood.

[8] https://svn.sdss.org/public/data/eboss/mcmc/trunk/likelihoods

Chain#	Type	Dataset
1	BAO only	eBOSS . . . . .
2		eBOSS+DESY3 . . . . .
3	BAO+DL	eBOSS+DL . . . . .
4		eBOSS+DESY3+DL .
5	BAO+BBN	eBOSS+BBN . . . . .
6		eBOSS+DESY3+BBN

Table 4.7: List of the different chains run in this section. We consider three types of chains: BAO only, BAO+DL and BAO+BBN, each of them run without and with the DES Y3 BAO likelihood.

- BAO + distance ladder (DL) chain. This chain includes the BAO data and also a prior on h (or H_0), which we take from [116]. In this paper, the authors perform a study using SN Ia distances calibrated from 70 long-period cepheids in the Large Magellanic Cloud, and find $h = 0.7403 \pm 0.0142$, including systematic errors. We take this measurement as a Gaussian prior on h when running this chain.

- BAO + Big Bang nucleosynthesis chain. This chain includes the BAO data and also a prior on $\omega_b = h^2\Omega_b$ that comes from BBN measurements. We use the results of recent high resolution spectroscopic measurements of seven quasar absorption systems that indicate a primordial deuterium abundance $D/H = (2.527 \pm 0.030) \times 10^{-5}$, see [117]. Using the empirically-derived reaction cross-section from [118], one can estimate ω_b.

We run each of the previous chains without and with the DES Y3 measurement, since we are interested in finding out how much information our measurement adds to the final contours. The list of chains is displayed in Table 4.7, where we associated each chain with a number from 1 to 6. The marginalized contours for all the chains listed in this table, assuming ΛCDM, are shown in Fig. 4.14, and the constraints on the different parameters are displayed in Table 4.8. The BAO-only chains, i.e., the red and green contours in Fig. 4.14, are not too sensitive to the Hubble constant: the contours on the plane $h - r_d$ are completely degenerate. Therefore, we cannot trust the values of h or r_d for chains 1 and 2 (in fact, the uncertainties in these parameters for these chains are very large). This is the reason why it is necessary to combine the BAO likelihood with additional measurements of h or ω_b[9] (in order to break this degeneracy). The chains run with the distance ladder prior, i.e., the blue and purple contours in Fig. 4.14, and those with the BBN prior, i.e., the yellow and black

[9] Adding a prior on ω_b would effectively allow us to constrain r_d and, therefore, also h.

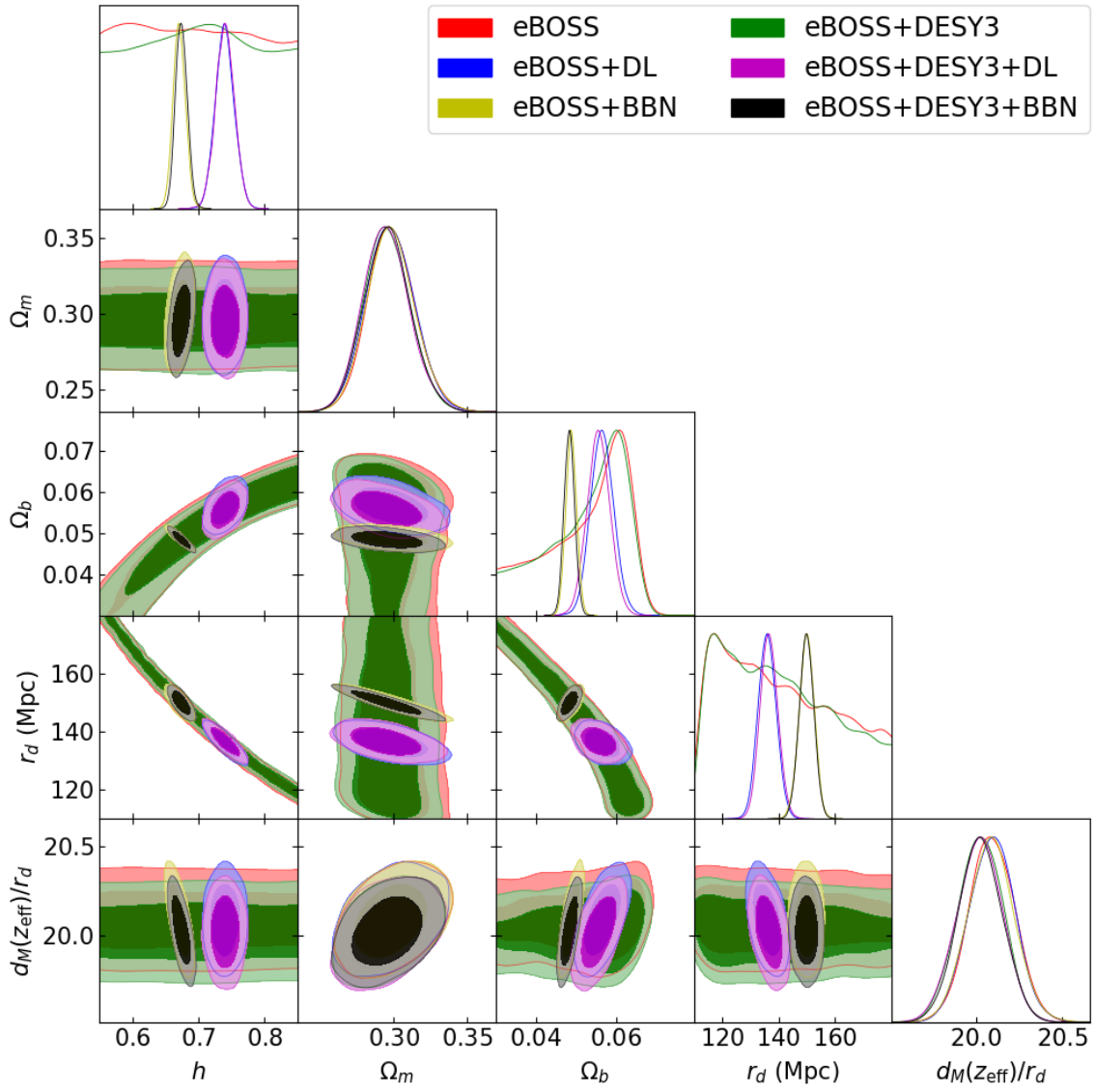

Figure 4.14: Marginalized constraints in h, Ω_m and Ω_b, and also in the derived parameters r_d and $d_M(z_{\rm eff})/r_d$, for the different chains listed in Table 4.7, assuming ΛCDM. Red and green contours: BAO-only chains (without and with our BAO likelihood, respectively). Blue and purple contours: BAO+DL chains (without and with our BAO likelihood, respectively). Yellow and black contours: BAO+BBN chains (without and with our BAO likelihood, respectively).

Chain#	h	Ω_m	Ω_b	r_d	$d_M(z_{\rm eff})/r_d$
1	$0.726^{+0.151}_{-0.149}$	$0.298^{+0.017}_{-0.016}$	$0.055^{+0.009}_{-0.023}$	$138.4^{+36.1}_{-23.8}$	$20.09^{+0.13}_{-0.13}$
2	$0.721^{+0.153}_{-0.149}$	$0.295^{+0.016}_{-0.016}$	$0.054^{+0.009}_{-0.024}$	$140.2^{+36.4}_{-24.6}$	$20.03^{+0.13}_{-0.13}$
3	$0.741^{+0.014}_{-0.014}$	$0.298^{+0.016}_{-0.015}$	$0.057^{+0.003}_{-0.003}$	$135.7^{+3.1}_{-3.0}$	$20.09^{+0.13}_{-0.13}$
4	$0.740^{+0.014}_{-0.014}$	$0.294^{+0.016}_{-0.015}$	$0.056^{+0.003}_{-0.003}$	$136.6^{+3.0}_{-3.0}$	$20.03^{+0.13}_{-0.13}$
5	$0.670^{+0.009}_{-0.009}$	$0.298^{+0.017}_{-0.016}$	$0.049^{+0.001}_{-0.001}$	$150.0^{+2.6}_{-2.8}$	$20.10^{+0.13}_{-0.13}$
6	$0.673^{+0.009}_{-0.009}$	$0.295^{+0.016}_{-0.015}$	$0.048^{+0.001}_{-0.001}$	$150.0^{+2.5}_{-2.7}$	$20.02^{+0.14}_{-0.13}$

Table 4.8: Constraints on the cosmological parameters h, Ω_m and Ω_b obtained running the different BAO chains listed in Table 4.7, assuming ΛCDM. We also display constraints on two derived parameters: r_d and $d_M(z_{\rm eff})/r_d$.

contours in that same figure, give closed contours on all the parameters. We find that local h measurements prefer values for the sound-scale horizon of

$$r_d \sim 136 \pm 3 \text{ Mpc}, \tag{4.63}$$

whereas early Universe measurements of ω_b prefer values of

$$r_d \sim 150 \pm 2.7 \text{ Mpc}, \tag{4.64}$$

independently of running eBOSS-only chains or including also our BAO measurement. These results are clearly in tension, since they are not compatible with each other, not even at the 2σ level. The contribution of our measurement to the whole BAO picture is moderate, as expected, since it only represents a point in a datavector that already had a total of 14 measurements from eBOSS results. However, since our measurement is more than 2σ away from Planck's prediction, the final constraint on $d_M(z_{\rm eff})/r_d$ moves towards lower values when including our likelihood. From the results displayed in Table 4.8, we find that

$$d_M(z_{\rm eff})/r_d \sim 20.09 \pm 0.13 \tag{4.65}$$

for all the chains without our BAO measurement, whereas

$$d_M(z_{\rm eff})/r_d \sim 20.03 \pm 0.13 \tag{4.66}$$

for all the chains that include it. Both of them are compatible with Planck's prediction, which is ∼20.11 (see eq. (4.53)), but our measurement clearly drives this parameter to lower values, as expected. It is also interesting that, adding our BAO likelihood, the data prefers a slightly larger value of r_d for the chains run with the distance ladder prior (chains 3 and 4 of Table 4.8). As for the goodness-of-fit, we find $\chi^2 \sim 11.5$ for the chains that only include eBOSS data, for a total of 11 degrees of freedom (p-value of about 0.4), whereas $\chi^2 \sim 16.3$ for the chains that also include

Ch.#	h	Ω_m	Ω_b	Ω_k	r_d	$d_M(z_{\rm eff})/r_d$
1	$0.717^{+0.157}_{-0.154}$	$0.285^{+0.023}_{-0.023}$	$0.060^{0.015}_{-0.025}$	$0.076^{+0.092}_{-0.086}$	$138.9^{+38.2}_{-25.0}$	$20.13^{+0.14}_{-0.14}$
2	$0.716^{+0.157}_{-0.152}$	$0.283^{+0.023}_{-0.022}$	$0.057^{0.016}_{-0.024}$	$0.063^{+0.094}_{-0.086}$	$140.1^{+37.7}_{-25.0}$	$20.05^{+0.14}_{-0.13}$
3	$0.739^{+0.014}_{-0.013}$	$0.286^{+0.023}_{-0.022}$	$0.064^{0.011}_{-0.010}$	$0.070^{+0.091}_{-0.085}$	$134.7^{+3.4}_{-3.3}$	$20.13^{+0.14}_{-0.14}$
4	$0.739^{+0.014}_{-0.014}$	$0.283^{+0.023}_{-0.023}$	$0.063^{0.012}_{-0.011}$	$0.066^{+0.092}_{-0.090}$	$135.5^{+3.5}_{-3.3}$	$20.06^{+0.13}_{-0.14}$
5	$0.647^{+0.033}_{-0.029}$	$0.287^{+0.022}_{-0.022}$	$0.052^{0.005}_{-0.005}$	$0.068^{+0.086}_{-0.089}$	$154.0^{+5.9}_{-6.2}$	$20.12^{+0.14}_{-0.14}$
6	$0.656^{+0.032}_{-0.032}$	$0.287^{+0.022}_{-0.023}$	$0.051^{0.005}_{-0.005}$	$0.047^{+0.093}_{-0.081}$	$153.0^{+6.4}_{-6.0}$	$20.05^{+0.13}_{-0.13}$

Table 4.9: Constraints on the cosmological parameters h, Ω_m, Ω_b and Ω_k obtained running the different BAO chains listed in Table 4.7, assuming oΛCDM. We also display constraints on two derived parameters: r_d and $d_M(z_{\rm eff})/r_d$.

our BAO measurement, for a total of 12 degrees of freedom (p-value of about 0.18). We find that adding our Y3 measurement makes the χ^2 worse slightly large compared to the degrees of freedom, but the p-value is still reasonable. The reason for this is that our measurement prefers lower values for $d_M(z_{\rm eff})/r_d$ when compared to eBOSS likelihoods, and combining them all drives the result to an intermediate value to the one each of them prefer.

4.6.3.2 BAO Chains in oΛCDM

It is also interesting to study the cosmological constraints assuming a oΛCDM cosmological model, i.e, considering Ω_k as a free parameter when running the chains. In Fig. 4.15 we plot the marginalized contours on the different cosmological parameters, and in Table 4.9 we explicitly display the constraints on them, which can be compared with those in Table 4.8. We find that adding Ω_k as a free parameter slightly decreases the value of Ω_m for all the different chains, and also increases Ω_b. It also decreases the value of h, except for those chains with the distance ladder prior (chains 3 and 4). Interestingly, for these two chains we find a slightly smaller value for r_d compared to the ΛCDM chains, but for chains 5 and 6 (the ones that include the BBN prior), r_d is larger in the case of oΛCDM. As for the goodness-of-fit, we find a $\chi^2 \sim 11.15$ for the chains that only include eBOSS data, for a total of 10 degrees of freedom (p-value of about 0.35), whereas $\chi^2 \sim 16.1$ for the chains that also include our BAO measurement, for a total of 11 degrees of freedom (p-value of about 0.14). The p-values are reasonable for all the different chains, although, as happened in the ΛCDM case, adding our BAO likelihood drives the χ^2 towards larger values compared to the degrees of freedom. We also find that the comparison between the results adding or not adding our BAO likelihood are qualitatively similar to the results obtained for ΛCDM.

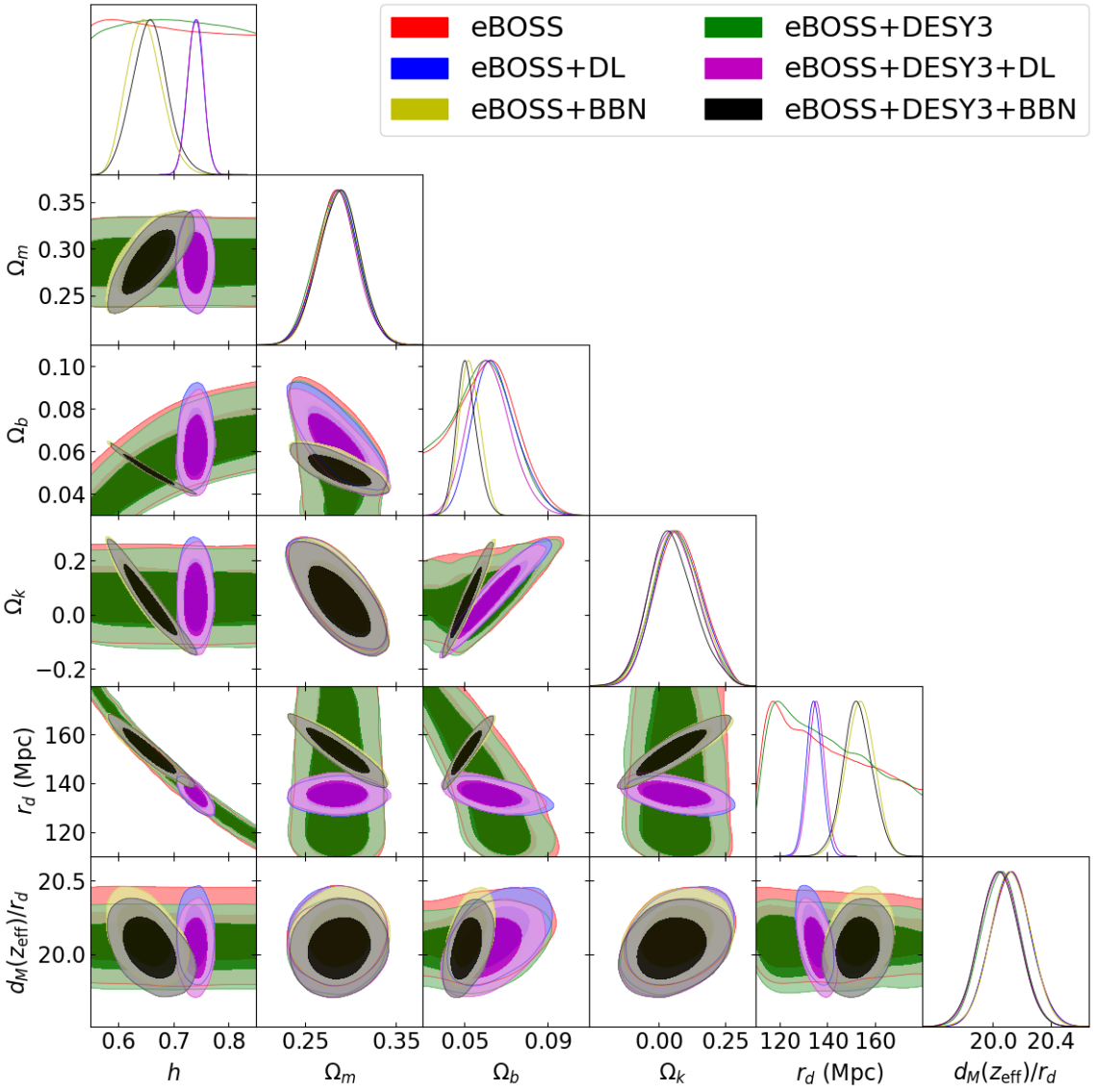

Figure 4.15: Marginalized constraints in h, Ω_m, Ω_b and Ω_k, and also in the derived parameters r_d and $d_M(z_{\rm eff})/r_d$, for the different chains listed in Table 4.7, assuming oΛCDM. Red and green contours: BAO-only chains (without and with our BAO likelihood, respectively). Blue and purple contours: BAO+DL chains (without and with our BAO likelihood, respectively). Yellow and black contours: BAO+BBN chains (without and with our BAO likelihood, respectively).

4.6.3.3 Combination of BAO with Other DES Probes

The combination of BAO with other cosmological probes, such as CMB, SN and WL, has shown to provide powerful constraints on the properties of the Universe and its evolution. These combined analyses can significantly improve the precision and accuracy of cosmological parameter measurements, and help to shed light on the nature of dark energy, dark matter, and the overall cosmic expansion history. The BAO itself is sensitive to the large-scale structure of the Universe and can provide precise measurements of the expansion rate of the Universe at different redshifts, which is influenced by the presence of dark energy and dark matter. Other cosmological probes, such as CMB, SN, and WL, provide complementary information on different aspects of the Universe, such as its early evolution, the distance-redshift relation, and the growth of cosmic structures. By combining BAO measurements with these other cosmological probes it is possible to break degeneracies between different parameters and obtain more robust and precise constraints on cosmological parameters, such as the matter density, the dark energy density, and the equation of state of dark energy. For example, the combination of BAO and CMB data can provide tight constraints on the geometry of the Universe and the total amount of matter it contains. The addition of SN data can further constrain the expansion history and the properties of dark energy, while WL data can provide insights into the growth of cosmic structures and the properties of dark matter.

Here we combine our DES Y3 BAO measurement with the result from the combination of galaxy clustering and weak lensing, i.e., the 3×2pt, which we already introduced in Chapter 3, and the DES SN [119]. Converting the 3×2pt ΛCDM MCMC from [28] into a posterior distribution on d_M/r_d at the effective redshift of our sample, we obtain

$$d_M(z_{\rm eff})/r_d = 20.15 \pm 0.70. \tag{4.67}$$

This result is larger than our Y3 BAO measurement of eq. (4.56) by $\sim 1.43\sigma$ accounting for the joint 68% confidence level, i.e.,

$$(20.15 - 18.94)/\sqrt{0.70^2 + 0.48^2} = 1.43. \tag{4.68}$$

However, it is consistent with the Planck result of $d_M(z_{\rm eff})/r_d = 20.11$. The relevance of our BAO measurement for the DES experiment itself is summarized in Fig. 4.16, which shows marginalized contours on cosmological parameters for the combination of 3×2pt [28], SN [119] and BAO, assuming the flat ΛCDM cosmological model. These chains were run using CosmoSIS [120], which is an open-source software framework for cosmological parameter estimation and data analysis. The details of the DES likelihoods, scale cuts and priors can be found in ([28, 119]), and also in the references therein.

As we already mentioned in Chapter 3, DES 3×2pt can constrain well Ω_m and σ_8. However, it cannot constraint that well the Hubble constant. The constraining power in Ω_m is further increased by the inclusion of DES SN [119], as discussed in

[28]. These two datasets help constraining r_d through the baryon density, Ω_b, and therefore the inclusion of BAO enables a much tighter constrain in h, shrinking the (Ω_m, h) and (σ_8, h) contours shown in Fig. 4.16. For DES 3×2pt+SN we obtain

$$\begin{cases} \Omega_m = 0.342^{+0.029}_{-0.025}, \\ h = 0.681^{+0.109}_{-0.082}, \\ \sigma_8 = 0.724^{+0.020}_{-0.042}, \end{cases} \tag{4.69}$$

whereas for DES 3×2pt+SN+BAO the results are

$$\begin{cases} \Omega_m = 0.317^{+0.021}_{-0.020}, \\ h = 0.709^{+0.089}_{-0.059}, \\ \sigma_8 = 0.758^{+0.033}_{-0.036}. \end{cases} \tag{4.70}$$

We clearly see that adding the BAO measurement yields to more precise constraints on these cosmological parameters. The posterior for h is more symmetrical, since adding the BAO penalizes low h values (the prior for h is flat in $0.55 - 0.91$), with a gain in constraining power of ∼20%. Also, the error in Ω_m is reduced by ∼25%, and the constraining power in σ_8 improves by ∼16% with the inclusion of our BAO measurement.

We have found that the contribution of our BAO measurement to the whole BAO picture is moderate when compared to all the other measurements over the last decade. However, it plays an important role for DES itself, since its combination with other DES probes, such as 3×2pt and SN, yields to more precise constraints in Ω_m, h and σ_8.

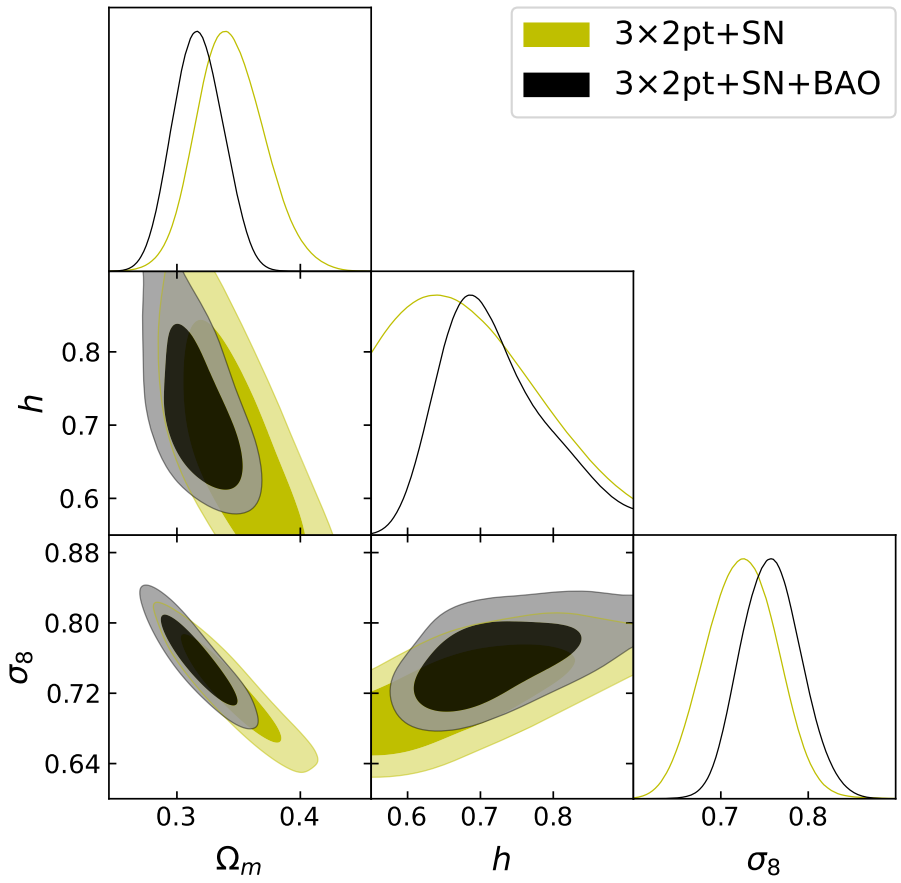

Figure 4.16: Marginalized constraints in Ω_m, h and σ_8 for the combination of DES 3×2pt, SN and BAO. These contours highlight the degeneracy breaking introduced by the BAO, enabling us to improve the constrains in Ω_m and, particularly, in h.

Chapter 5

Measurement of the BAO Scale in DES Y6

In this chapter we present the BAO analysis on the DES Y6 data. For the Y6 analysis, the selection cuts of the BAO sample are re-optimized with respect to those of Y3 in order to get the most precise distance measurement possible. The redshift range of the sample is also modified by increasing its upper limit up to 1.2. With these changes, we can estimate a decrease in the size of the error-bar associated to the BAO measurement of about 25% with respect to the Y3. However, since the Y6 analysis is still ongoing, we are not allowed to show the final measurement on the data in this manuscript (because the analysis is still on its blinded phase). Therefore, we show measurements on simulations and the expected improvement when combining 3×2pt with BAO.

5.1 The Y6 BAO Sample

The sample selection for the Y6 is done in two steps. First, we select and analyze what we refer to as the "preliminary Y6 BAO sample", which is selected with the same cuts as in the Y3, eq. (4.1). This preliminary Y6 sample is already better in terms of photo-z[1] accuracy and expected precision in the BAO measurement than the Y3 sample. Second, we perform an optimization of the sample selection in order to get the best possible forecasted precision. This optimization is carried out in terms of two parameters a and b that define the i-magnitude limit of our sample as

$$i < a + b z_{\rm ph}, \tag{5.1}$$

where $z_{\rm ph}$ is the point estimate of the photometric redshift of the source.

[1] The fiducial photo-z code for the Y6 analysis is DNF, as in Y3. DNF was already briefly described in section 4.2.3, and the only difference with respect to the Y3 analysis is that now we are also including the information in the Y band when estimating the photo-z of our galaxies.

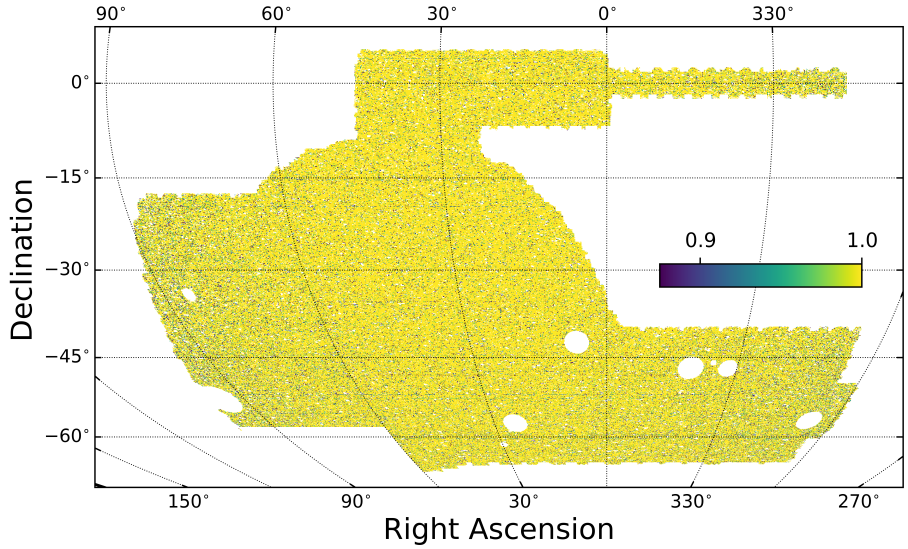

Figure 5.1: Footprint of the angular mask for the DES Y6 BAO analysis, with the $i < 22.5$ cut already applied. The value plotted for each pixel represents its detection fraction. The total area of the mask, computed weighting by the detection fraction, is 4,357.01 deg^2.

5.1.1 Angular Mask

The angular mask is constructed in a similar way to the Y3 analysis, see section 4.2.1. In Fig. 5.1 we show the footprint of the Y6 angular mask. Since for the Y6 analysis we want to optimize the sample as a function of the i-magnitude selection cut, see eq. (5.1), it is necessary to carefully account for the depth maps related to our footprint. Each pixel plotted in Fig. 5.1 reaches a different depth in the i band, which essentially means that, if we want to build a galaxy sample that goes up to a certain $i_{\rm max}$, we have to keep only those pixels that reach a depth greater than $i_{\rm max}$. This effectively limits the area of the mask as a function of the i-magnitude cut that we set, as shown in Fig 5.2. The deeper we want our sample to be, the more area of the full footprint we need to drop. The orange dashed line shown in this figure corresponds to the Y3 i-magnitude limit, which was simply $i_{\rm max} = 19 + 3z_{\rm ph}^{\rm max} = 19 + 3 \times 1.1 = 22.3$ (directly computed from the Y3 selection cuts, eq. (4.1)). The green dashed line corresponds to the i-magnitude limit chosen for the Y6, which is set to 22.5.

Since for the Y6 we have better data quality than for the Y3, i.e., less noisy magnitude estimations (because of the longer exposure time), we can afford going deeper in magnitude (and also in redshift). However, we cannot arbitrarily go to higher magnitudes: first, because we would lose a lot of area; second, because very faint galaxies might be complicated for our photo-z code to deal with; and third, because these galaxies may be an important source of systematics. The limiting magnitude for the Y6 was chosen to be 22.5 because we do not have reference spectra

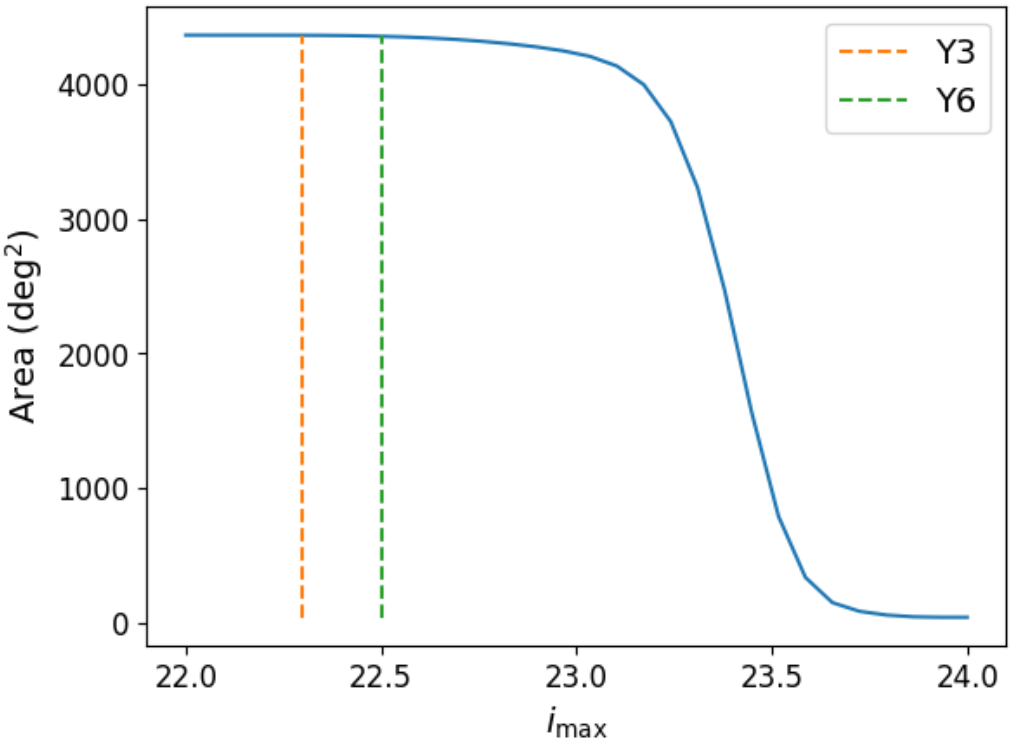

Figure 5.2: Area of the footprint of the Y6 angular mask as a function of the i-magnitude limit (blue solid line). The orange dashed line indicates the limit for the Y3 analysis ($i_{\mathrm{max}} = 22.3$), whereas the green dashed line represents the same but for the Y6 ($i_{\mathrm{max}} = 22.5$).

for fainter galaxies. Therefore, we would not be able to trust and/or validate the photo-z of such sample[2]. In section 5.1.5.1 we discuss the validation of the photo-z.

5.1.2 Preliminary Selection of the Sample

The DES Y6 BAO sample is selected directly from the DES Y6 Gold Catalog. Before going into detail about the selection and optimization of the Y6 sample, we firstly

[2] This validation is done using VIPERS, which is a complete sample above redshift 0.5, but only up to $i < 22.5$ (we already described VIPERS in section 4.2.3), see [85] for further details.

Description	EXT_MASH
Data not available	-9
High confidence stars	0
Candidate stars	1
Mostly galaxies	2
High confidence galaxies ..	3
Ultra-pure galaxy selection	4

Table 5.1: Possible values of the star-galaxy separator, EXT_MASH, of the Y6 Gold Catalog. The ultra-pure galaxy selection has a stellar contamination of $< 1\%$.

apply some quality cuts to the DES Y6 Gold Catalog, namely

$$\begin{aligned} \text{EXT\_MASH} &\geq 3, \\ \text{FLAGS\_GOLD} &= 0, \\ \text{N\_IMAGES\_G} &> 1, \\ \text{N\_IMAGES\_R} &> 1, \\ \text{N\_IMAGES\_I} &> 1, \\ \text{N\_IMAGES\_Z} &> 1, \\ \text{N\_IMAGES\_Y} &> 1, \end{aligned} \tag{5.2}$$

where EXT_MASH is the star-galaxy separator[3] for the Y6 Gold Catalog, and its possible values are summarized in Table 5.1; FLAGS_GOLD indicates unusual characteristics of individual objects, including fitting failures and measurement anomalies; and the variables N_IMAGES_GRIZY are the number of *grizY* band exposures at the object location (from the HEALPix map). These quality cuts are common to both the BAO sample and the MagLim lens sample, and are applied by default to every galaxy sample throughout this chapter. However, they will be refined in section 5.1.4.

The most natural way to start the Y6 BAO analysis is to create a preliminary Y6 BAO sample with the same selection cuts used for the Y3, i.e., those given in eq. (4.1). Selecting a sample with the same selection cuts as in the Y3 allows us to directly compare its properties with those of the Y3 sample, which is useful to check if everything is working as intended. In Table 5.2 we display the number of galaxies, the width of the photo-z distribution, the mean redshift and the width of the redshift distributions for the preliminary Y6 BAO sample and also the Y3 one (which is useful for comparison purposes). The $\langle z \rangle$, σ_{68} and W_{68} columns are defined in the exact same way as in section 4.2.3, see eqs. (4.4), (4.5) and (4.6), respectively.

[3] The star-galaxy separator is based on two main properties of the objects DES observes: on the one hand, their extensions (point-like objects are star candidates, whereas extended objects are galaxy candidates); on the other hand, their colors.

Bin	$\langle z \rangle$		$N_{\rm gal}$		σ_{68}		W_{68}	
	Y3	Y6-pre	Y3	Y6-pre	Y3	Y6-pre	Y3	Y6-pre
1	0.648	0.650	1,478,178	1,574,405	0.0246	0.0210	0.0526	0.0486
2	0.750	0.749	1,632,805	1,661,151	0.0279	0.0228	0.0608	0.0544
3	0.838	0.847	1,727,646	1,809,905	0.0298	0.0258	0.0630	0.0583
4	0.931	0.937	1,315,604	1,474,098	0.0363	0.0324	0.0794	0.0712
5	1.028	1.017	877,760	953,731	0.0455	0.0379	0.0989	0.0829

Table 5.2: Comparison of some of the main properties of the Y3 and the preliminary Y6 BAO samples. The ones for the Y3 sample were already displayed in Table 4.1, but we also include them here for comparison purposes. The area of the angular mask is 4,108.47 deg^2 for the Y3 and 4,364.72 deg^2 for the Y6 (corresponding to an i-magnitude limit of 22.3, see Fig. 5.2). The number density is very similar for both samples, and the photo-z precision is better in the case of the Y6 (both σ_{68} and W_{68} are smaller for every redshift bin).

From Table 5.2 we to find that the number density of galaxies is very similar between the Y3 sample and the preliminary Y6 one (the increase in the number of galaxies scales with the increase in the area of the mask). Explicitly,

$$\frac{N_{\rm gal}}{\rm Area}({\rm Y3}) = 1711.59 \text{ gal/deg}^2, \quad \frac{N_{\rm gal}}{\rm Area}({\rm Y6}) = 1714.02 \text{ gal/deg}^2. \tag{5.3}$$

This is what we expected, since the Y3 sample was already complete in the sense that there were no missing galaxies in its redshift range for its magnitude limits. We also find that the preliminary Y6 BAO sample has better photo-z precision: both the photo-z dispersion, σ_{68}, and the width of the redshift distributions, W_{68}, are smaller compared to the Y3 values, and this happens in all the redshift bins. The reason for this is simple: as we already mentioned, in the Y6 analysis we accumulated a longer exposure time with respect to the Y3, which means the photometry in the Y6 is better than that of the Y3, and since DNF estimates the redshift from this photometry, its results are also be more accurate. Also, for the Y6 analysis we also included the information on the Y-band filter to estimate the photo-z with DNF, i.e., we used the 5 *grizY* bands, as we mentioned earlier.

In Fig. 5.3 we compare the density of galaxies and the total number of galaxies, both of them as a function of redshift, between the Y3 and the preliminary Y6 BAO samples. The galaxy density as a function of redshift (left panel of Fig. 5.3) is very similar for the Y3 and the Y6 samples. The main difference is the peak found in the Y3 case at around redshift 0.83, which is due to stellar contamination. The reason for this is that in this redshift region, the colors of some stars are very similar to those of the BAO sample galaxies. The star-galaxy separation is better characterized in the Y6 analysis, and this explains the different shape of the distributions in this region. As for the number of galaxies (right panel of Fig. 5.3), we find a higher number of

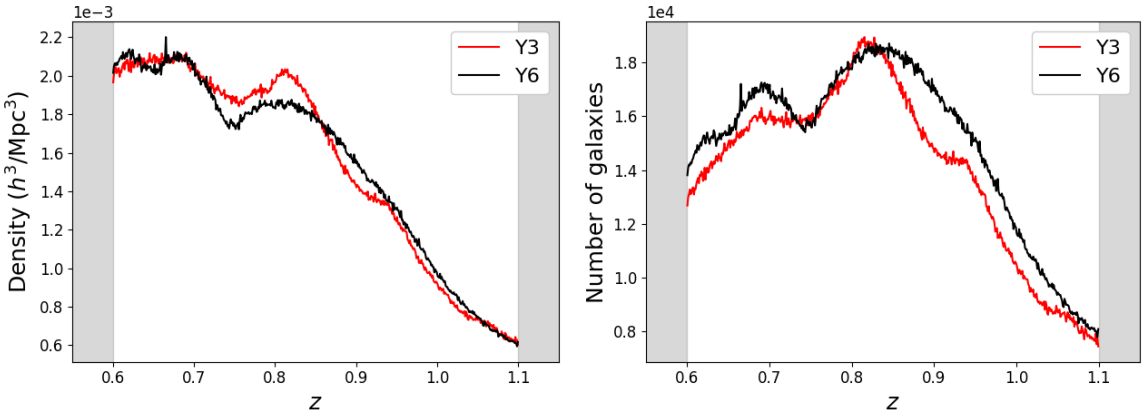

Figure 5.3: Comparison between the Y3 and the preliminary Y6 BAO samples. **Left panel:** density of galaxies as a function of redshift. **Right panel:** number of galaxies as a function of redshift.

galaxies for the Y6 case for all redshift. This is expected, because the area of the Y6 mask is larger than that of the Y3, which directly translates into a larger amount of galaxies while keeping the galaxy density at almost the same value, as we saw in eq. (5.3).

From the results displayed in Table 5.2, we already know that the preliminary Y6 sample has more galaxies (because it also has more area) and better photo-z estimates (the dispersion on the photo-z error and the width of the redshift distributions, σ_{68} and W_{68}, respectively, are smaller). These are properties for the two individual galaxy samples; however, it would also be interesting to study the properties of the common galaxies between them, i.e., the matching between both samples. Since the DES Data Management (DESDM) reprocessed the whole data to create the Y6 Gold Catalog, the unique identifier for each galaxy changes from Y3 to Y6 (the so-called COADD_OBJECT_ID). Therefore, it is necessary to match the catalogs with their coordinates in the sky, i.e., with the right ascension (ra) and the declination (dec). We consider two galaxies to be the same if

$$\theta < 10^{-4}\text{deg}, \tag{5.4}$$

where θ is given by eq. (2.73). Matching the Y3 and the preliminary Y6 samples with the criteria defined in eq. (5.4), we find that there are a total of 5,622,143 galaxies which are common between the Y3 and the preliminary Y6 samples, i.e., approximately 80% of the Y3 sample is included in the preliminary Y6 one, and ∼75% of the preliminary Y6 sample is included in the Y3 one. As expected, most of the preliminary Y6 sample is also part of the Y3 one. The reason that 25% of the sample is constituted of different galaxies is partly due to the fact that the area of the Y6 angular mask is larger, but also because of the photo-z resolution.

5.1.3 Optimization of the Selection Cuts

Using the DES Y6 data we can go deeper in magnitude (and also in redshift) compared to the Y3 analysis. Therefore, for the Y6 analysis it is interesting to study the limiting magnitude in the i band that optimizes the BAO scale measurement (in the sense of the lowest error in the BAO distance measurement, $\sigma_{\rm BAO}$ hereafter), and not just use the same selection as in the Y3. We use a forecast method based on the Fisher matrix to compute the expected $\sigma_{\rm BAO}$ for a given galaxy sample and, then, develop an algorithm to run this forecast over several different galaxy samples. This allows us to search for the optimal sample out of thousands of them.

5.1.3.1 Forecast Method

In cosmology, Fisher forecasts are a way to estimate how well a particular experiment or survey will be able to constrain the parameters that describe our Universe. This is done by computing the Fisher information matrix, which relates the covariance of the observables to the parameters of interest. The diagonal elements of the Fisher matrix correspond to the variance of the estimated parameters, while the off-diagonal elements correspond to the covariances between them.

The forecasting method is based upon the methodology developed in [121]. Following [122] and assuming the likelihood function of the band powers of the galactic power spectrum to be Gaussian, the Fisher matrix can be approximated as

$$\begin{aligned} F_{ij} &= \int_{\boldsymbol{k}_{\rm min}}^{\boldsymbol{k}_{\rm max}} \frac{d^3\boldsymbol{k}}{2(2\pi)^3} \frac{\partial \log P_G(\boldsymbol{k})}{\partial p_i} \frac{\partial \log P_G(\boldsymbol{k})}{\partial p_j} V_{\rm eff}(\boldsymbol{k}) \\ &= \int_{-1}^{1} d\mu \int_{k_{\rm min}}^{k_{\rm max}} \frac{2\pi k^2 dk}{2(2\pi)^3} \frac{\partial \log P_G(k,\mu)}{\partial p_i} \frac{\partial \log P_G(k,\mu)}{\partial p_j} V_{\rm eff}(k,\mu). \end{aligned} \tag{5.5}$$

$P_G(\boldsymbol{k})$ is the observed galactic power spectrum at $\boldsymbol{k}$, μ is the cosine of the angle of $\boldsymbol{k}$ with respect to the LOS, p_i are the cosmological parameters to be constrained, and $V_{\rm eff}$ is the effective volume of the survey, given by

$$\begin{aligned} V_{\rm eff}(k,\mu) &= \int d^3\boldsymbol{r} \left[\frac{n_{\rm gal}(\boldsymbol{r}) P_G(k,\mu)}{n_{\rm gal}(\boldsymbol{r}) P_G(k,\mu) + 1} \right]^2 \\ &= \left[\frac{(1+\beta\mu^2)^2 P(k)}{(1+\beta\mu^2)^2 P(k) + n_{\rm gal}^{-1}} \right]^2 V_{\rm survey}. \end{aligned} \tag{5.6}$$

Here, $n_{\rm gal}(\boldsymbol{r})$ is the comoving number density of galaxies (that we assumed constant in position) and β is the linear redshift-space-distortion parameter. We also defined

$$V_{\rm survey} = \frac{4\pi}{3} f_{\rm sky} \left[\chi(z_{\rm max})^3 - \chi(z_{\rm min})^3 \right]. \tag{5.7}$$

Following [121], the baryonic features in the nonlinear power spectrum can be expressed as

$$\begin{aligned} P_b(k,\mu) &= P_b^{\rm lin}(k,\mu)\exp\left(-k_\parallel\Sigma_\parallel^2 - k_\perp\Sigma_\perp^2\right) \\ &= P_b^{\rm lin}(k,\mu)\exp\left(-k^2\left[\mu^2\Sigma_\parallel^2 + (1-\mu^2)\Sigma_\perp^2\right]\right) \\ &= P_b^{\rm lin}(k,\mu)\exp\left(-k^2\Sigma_{\rm tot}^2(\mu)\right). \end{aligned} \tag{5.8}$$

The previous expression is equivalent to the one we already deduced in eq. (4.21) if we define

$$P_b^{\rm lin}(k,\mu) = \left(b + f\mu^2\right)^2\left[P_m^{\rm lin}(k) - P_m^{\rm nw}(k)\right]. \tag{5.9}$$

$P_b^{\rm lin}$ is the part of the linear power spectrum with the acoustic feature, including Silk damping effects, and

$$\Sigma_{\rm tot}^2(\mu,z) = \mu^2\Sigma_\parallel^2(z) + (1-\mu^2)\Sigma_\perp^2(z), \tag{5.10}$$

which is the same expression as eq. (4.24) but without the factor proportional to $\delta\Sigma^2(z)$. $\Sigma_\parallel$ and $\Sigma_\perp$ are the same root mean square radial displacements along and perpendicular to the LOS we already introduced in Chapter 4, and they are given by eq. (4.25). Introducing these expressions into eq. (5.5), the full Fisher matrix becomes

$$\begin{aligned} F_{ij} = \int_{-1}^{1} d\mu \int_{k_{\rm min}}^{k_{\rm max}} &\frac{2\pi k^2 dk}{2(2\pi)^3}\frac{\partial \log P_m^{\rm lin}(k,\mu)}{\partial p_i}\frac{\partial \log P_m^{\rm lin}(k,\mu)}{\partial p_j} \\ &\times V_{\rm eff}(k,\mu)\exp\left(-2k^2\Sigma_{\rm tot}^2(\mu)\right). \end{aligned} \tag{5.11}$$

This Fisher matrix can, then, be approximated based on how well we can centroid the location of the peak (1-dimensional approximation, or $\beta = 0$), i.e., the sound horizon scale s_o at the drag epoch when observed in the reference cosmology,

$$F_{\log s_o} = V_{\rm survey}\int_{k_{\rm min}}^{k_{\rm max}} \frac{4\pi k^2 dk}{2(2\pi)^3}\frac{1}{\left[P(k) + n_{\rm gal}^{-1}\right]^2}\left[\frac{\partial P_b(k)}{\partial \log s_o}\right]^2. \tag{5.12}$$

Following [121], the baryonic power spectrum is found to be given by

$$P_b(k) = \sqrt{8\pi^2}A_0 P_{0.2}\frac{\sin(ks_o)}{ks_o}\exp\left(-(k\Sigma_{\rm Silk})^{1.4}\right)\exp\left(-k^2\Sigma_{\rm tot}^2\right), \tag{5.13}$$

where $P_{0.2}$ is the galaxy power spectrum at $k = 0.2\ h/{\rm Mpc}$ at the given redshift and the factors $\Sigma_{\rm Silk}$ and $\Sigma_{\rm tot}$ give the broadening of the BAO peak with a Gaussian function due to the Silk damping effect and the Lagrangian displacement, respectively. Therefore,

$$\begin{aligned} F_{\log s_o} &= 2V_{\rm survey}A_0^2 \\ \times\int_{k_{\rm min}}^{k_{\rm max}} dk &\left[\cos(ks_o) - \frac{\sin(ks_o)}{s_o}\right]^2 \frac{k^2\exp\left(-2(k\Sigma_{\rm Silk})^{1.4}\right)\exp\left(-2k^2\Sigma_{\rm tot}^2\right)}{\left(\frac{P(k)}{P_{0.2}} + \frac{1}{n_{\rm gal}P_{0.2}}\right)^2}. \end{aligned} \tag{5.14}$$

The sound horizon scale is $s_o \sim 100$ Mpc/h and most baryonic information comes from $k > 0.05$ h/Mpc, which means ks_o is large over the wave-number of our interest. Therefore, we can approximate the leading term $\cos^2(ks_o)$ as 1/2, and $\sin^2(ks_o)/(ks_o)^2$ is small relative to $\cos^2(ks_o)$. The fractional error on the location of the baryonic peak can, then, be written as

$$\sigma_{\log s_o} = \frac{\sigma_{s_o}}{s_o} = \sqrt{F^{-1}_{\log s_o}}$$
$$= \left\{ V_{\rm survey} A_0^2 \int_{k_{\rm min}}^{k_{\rm max}} dk \frac{k^2 \exp\left(-2(k\Sigma_{\rm Silk})^{1.4}\right) \exp\left(-2k^2\Sigma^2_{\rm tot}\right)}{\left(\frac{P(k)}{P_{0.2}} + \frac{1}{n_{\rm gal} P_{0.2}}\right)^2} \right\}^{-1/2} . \tag{5.15}$$

From eq. (5.15) it follows that the distance precision depends only on the survey volume, $n_{\rm gal}$ and the redshift of the survey. However, for photometric redshift surveys such as DES it also depends on the width of the photo-z distribution, $\Sigma_z \equiv \sigma_{68}$, since photometric redshift errors result in an exponential suppression of the power spectrum,

$$P \to P \exp\left(-k^2\mu^2\Sigma_z^2\right). \tag{5.16}$$

The fractional error on the location of the peak from the observed galaxy redshift surveys, i.e., $\sigma_{\log s_o}$, is equivalent to the fractional error on the distance estimation when the physical location of the peak is well known from the CMB. We compute $\sigma_{\log s_o}$ for each redshift bin and then combine them as

$$\sigma_{\rm BAO} = \left[\sum_{\rm zbin} \frac{1}{(\sigma^{\rm zbin}_{\log s_o})^2} \right]^{-1/2} . \tag{5.17}$$

Following this methodology we find that, in order to run the forecast, we need:

- $z_{\rm ph}^{\rm min}$ and $z_{\rm ph}^{\rm max}$ (0.6 and 1.1 for the Y3 BAO sample, as we saw in Chapter 4; and 0.6 and 1.2 for the Y6 one, as we will see).
- the width of the redshift bins, $\Delta z_{\rm ph}$ (0.1 for both the Y3 and the Y6 BAO analyses).
- the area of the angular mask, $A_{\rm mask}$, which depends on the i-magnitude limit of the sample (see Fig. 5.2). It allows us to compute $f_{\rm sky}$ as

$$f_{\rm sky} = \frac{A_{\rm mask}({\rm rad}^2)}{4\pi}. \tag{5.18}$$

- the value of σ_{68}, or Σ_z, for each individual redshift bin.
- the number of galaxies $N_{\rm gal}$ in each redshift bin, from which we compute the number density,

$$n_{\rm gal} = \frac{N_{\rm gal}}{A_{\rm mask}}. \tag{5.19}$$

All the previous information is displayed in Table 5.2 for the Y3 and the preliminary Y6 BAO samples. Hence, we can compute the expected $\sigma_{\rm BAO}$ for both samples with our forecast, which yields

$$\begin{cases} \sigma_{\rm BAO}^{\rm Y3} = 0.0214, \\ \sigma_{\rm BAO}^{\rm Y6-pre} = 0.0185. \end{cases} \tag{5.20}$$

We find that the forecast for our preliminary Y6 BAO sample is already better compared to the Y3 case. However, we can still optimize the sample by slightly modifying its selection, which we will find to result in an even better forecast for $\sigma_{\rm BAO}$. The optimization of the Y6 sample will allow us to reduce by a 25% the error-bar of our BAO measurement.

5.1.3.2 Optimization Algorithm

In order to optimize the sample selection for the best BAO scale measurement, we need to include, at least, one free parameter in our sample selection. We add this freedom in the flux selection[4], leaving the slope and the intercept on the y-axis as free parameters, i.e., as in eq. (5.1),

$$\begin{aligned} &1.7 < i - z + 2(r - i) && \text{(color selection)}, \\ &17.5 < i < a + b z_{\rm ph} && \text{(flux selection)}, \\ &0.6 < z_{\rm ph} < 1.1 && \text{(photo-}z\text{ range)}. \end{aligned} \tag{5.21}$$

Setting $a = 19$ and $b = 3$ corresponds to the Y3 selection. We must keep in mind, however, that the previous selection could yield $i > 22.5$ for some values of a and b. Since we use VIPERS for the validation of the photo-z, which is complete only up to $i = 22.5$, here we add an extra selection requiring $i < 22.5$. Therefore, the sample selection for the Y6 is given by

$$\begin{aligned} &1.7 < i - z + 2(r - i) && \text{(color selection)}, \\ &17.5 < i < a + b z_{\rm ph} && \text{(flux selection)}, \\ &i < 22.5 && \text{(magnitude limit)}, \\ &0.6 < z_{\rm ph} < 1.1 && \text{(photo-}z\text{ range)}. \end{aligned} \tag{5.22}$$

We allow a and b to vary in the ranges

$$19 \leq a \leq 22, \quad 1.5 \leq b \leq 3.5, \tag{5.23}$$

with 100 linearly-spaced values in each interval (for a total of 10,000 test samples), in order to search for their optimal values to perform the BAO measurement. The optimization algorithm works as follows:

[4] The other cuts are fixed by the survey characteristics.

1. Select a pair of values for a and b.

2. Compute $i_{\rm max} = \min(a + bz_{\rm ph}^{\rm max}, 22.5)$, where $z_{\rm ph}^{\rm max} = 1.1$. By default, we set the i-magnitude limit to 22.5 (see eq. (5.22)). However, depending on the values of a and b, for some samples $a + bz_{\rm ph}^{\rm max} < 22.5$ and, therefore, we would unnecessarily lose area for them if we simply set $i_{\rm max} = 22.5$. Therefore, the correct way to compute $i_{\rm max}$ for a given sample is to calculate the minimum between $a + bz_{\rm ph}^{\rm max}$ and 22.5.

3. Remove pixels with depth in the i-magnitude band smaller than $i_{\rm max}$ from the angular mask. Compute the total area of the remaining pixels (see Fig. 5.2 for reference). This step implies using a different angular mask for each sample (as a function of the value of $i_{\rm max}$).

4. Create a galaxy sample applying the Y6 quality cuts, defined by eq. (5.2), and also the corresponding selection cuts, given by eq. (5.22), to the Y6 Gold Catalog.

5. Compute σ_{68} and the number of galaxies $N_{\rm gal}$ for each redshift bin for the galaxy sample created in step 4.

6. Compute $\sigma_{\rm BAO}$ with the Fisher forecast code using the area of the angular mask (computed in step 3.), σ_{68} and $N_{\rm gal}$ (both of them computed in step 4.), as explained in section 5.1.3.1.

We apply this algorithm to the grid in the (a, b) plane defined by eq. (5.23) and find that the minimum value for $\sigma_{\rm BAO}$ obtained is

$$\sigma_{\rm BAO}^{\rm Y6-opt} = 0.0170, \tag{5.24}$$

which is lower than the one we obtained for the preliminary Y6 BAO sample (displayed in eq. (5.20)), as expected. The corresponding optimal parameters are $a = 19.64$ and $b = 2.894$, which are well within the limits of the (a, b) plane, eq. (5.23).

We can further optimize the BAO sample by increasing the photo-z range, i.e., adding one more redshift bin from 1.1 to 1.2 (we already discussed that the photo-z estimation is more accurate in the Y6 because of the longer time exposures, which means we can increase the redshift range). We now run forecasts in samples selected as

$$\begin{aligned} &1.7 < i - z + 2(r - i) && \text{(color selection)}, \\ &17.5 < i < a + bz_{\rm ph} && \text{(flux selection)}, \\ &i < 22.5 && \text{(magnitude limit)}, \\ &0.6 < z_{\rm ph} < 1.2 && \text{(photo-}z\text{ range)}. \end{aligned} \tag{5.25}$$

We find that the minimum value for $\sigma_{\rm BAO}$ is

$$\sigma_{\rm BAO}^{\rm Y6-opt} = 0.0162, \tag{5.26}$$

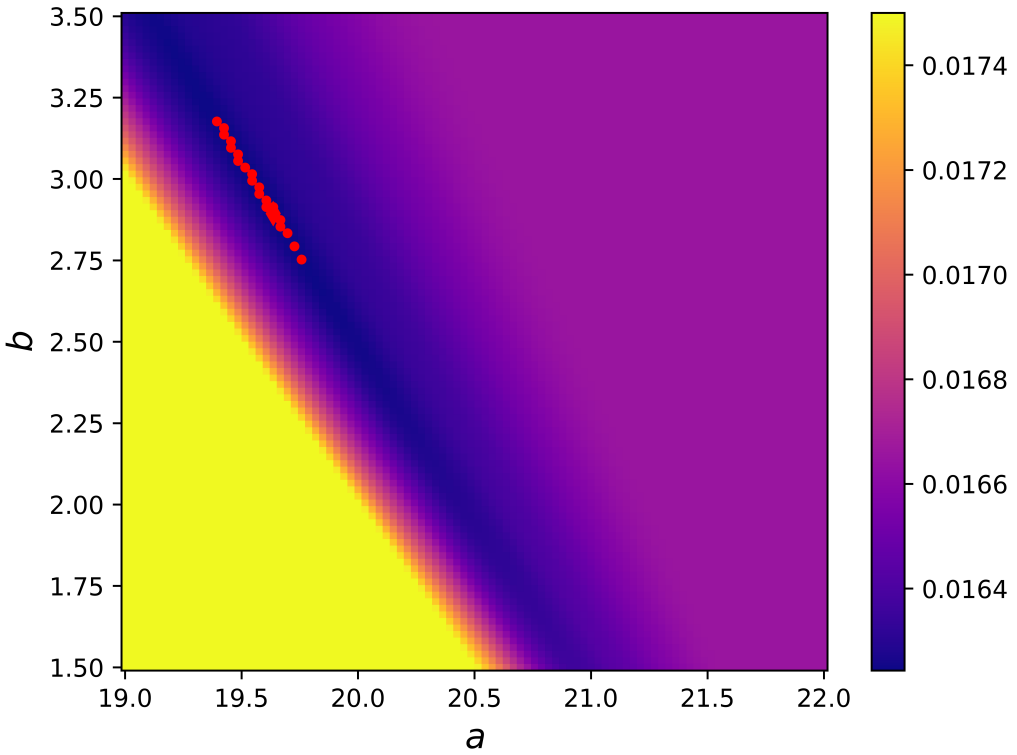

Figure 5.4: Heat-map of $\sigma_{\rm BAO}$ obtained for samples selected with different values of a and b following eq. (5.25). The red points represent the 20 samples with the lowest $\sigma_{\rm BAO}$.

which is smaller than the minimum value for the 5-bin case, i.e., $\sigma_{\rm BAO}$ decreases when adding one extra redshift bin, as expected. However, the optimal parameters found for the 6-bin case are exactly the same as the ones we found for the 5-bin one, namely

$$a_{\rm Y6-opt} = 19.64, \quad b_{\rm Y6-opt} = 2.894, \tag{5.27}$$

i.e., adding one extra redshift bin does not impact the optimal values for a and b. In Fig. 5.4 we show the $\sigma_{\rm BAO}$ heat-map obtained from the forecasts as a function of a and b. We find that there is a diagonal-like region in which $\sigma_{\rm BAO}$ reaches its minimum value, i.e., all the samples in this region have, approximately, the same $\sigma_{\rm BAO}$. We studied several different properties for all these samples, namely the width of the photo-z distribution, the total number of galaxies, the limiting magnitude and the number of photo-z outliers, but we did not find any significant difference between them: all their properties were quite similar. Therefore, we decided to choose the values of a and b corresponding to the sample with the lowest $\sigma_{\rm BAO}$, i.e., the ones

Bin	$\langle z \rangle$		$N_{\rm gal}$		σ_{68}		W_{68}	
	Y6-pre	Y6-opt	Y6-pre	Y6-opt	Y6-pre	Y6-opt	Y6-pre	Y6-opt
1	0.650	0.653	1,572,078	2,955,384	0.0211	0.0234	0.0486	0.0503
2	0.749	0.751	1,658,492	3,397,640	0.0228	0.0256	0.0544	0.0581
3	0.847	0.844	1,807,099	4,156,868	0.0258	0.0294	0.0583	0.0631
4	0.937	0.930	1,471,552	3,746,809	0.0324	0.0358	0.0719	0.0771
5	1.017	1.015	952,301	1,974,025	0.0379	0.0395	0.0829	0.0844
6	1.110	1.105	687,932	828,396	0.0406	0.0413	0.0919	0.0917

Table 5.3: Comparison of some of the main properties of the preliminary Y6 and the optimal Y6 BAO samples. The ones for the preliminary Y6 sample were already displayed in Table 5.2; however, the numbers here are slightly different because in this case we also included a sixth redshift bin from $1.1 < z < 1.2$, which means adding the $i < 22.5$ cut, which in turn reduces the area of the mask with respect to the 5-bin preliminary sample. The area of the angular mask is 4,357.01 deg^2 for both samples (corresponding to an i-magnitude limit of 22.5, see Fig. 5.2). The number density is much larger for the optimal sample, and the photo-z precision is slightly worse. However, since the optimization algorithm is a trade-off between these two quantities, we end up with a much more appropriate sample to measure the BAO feature after the optimization.

displayed in eq. (5.27). Therefore, the final selection of the Y6 sample is given by

$$\begin{aligned} &1.7 < i - z + 2(r - i) && \text{(color selection)}, \\ &17.5 < i < 19.64 + 2.894 z_{\rm ph} && \text{(flux selection)}, \\ &i < 22.5 && \text{(magnitude limit)}, \\ &0.6 < z_{\rm ph} < 1.2 && \text{(photo-}z\text{ range)}. \end{aligned} \tag{5.28}$$

In Table 5.3 we display the main properties of the optimal Y6 BAO sample, together with those of the preliminary Y6 sample for comparison purposes. The number of galaxies in each redshift bin is much larger for the case of the optimal Y6 sample, and this is due to the value of the parameter a in the photo-z-dependent selection cut of eq. (5.25). In the case of the optimal Y6 sample, a is significantly larger than that of the Y3 ($a_{\rm Y6-opt} = 19.64$ compared to $a_{\rm Y3} = 19$). This gives rise to a galaxy sample with a larger amount of galaxies (even though b decreases from Y6 to Y3, $b_{\rm Y6-opt} = 2.894$ compared to $b_{\rm Y3} = 3$, this decrease is not enough to compensate the increase in a, and we end up getting much more galaxies when applying the optimal selection). We also note that the photo-z precision is worse for the optimal sample, i.e., σ_{68} and W_{68} are lower in the case of the preliminary Y6 sample for every redshift bin. However, this is an expected result, since the optimization is actually a trade-off between the number density and the photo-z precision. Since the number of galaxies is much larger for the optimal sample, the

Case	$\sigma_{\rm BAO}$
Y3	0.0214
Y6-pre (5 redshift bins) ...	0.0185
Y6-pre (6 redshift bins) ...	0.0176
Y6-opt (5 redshift bins) ...	0.0170
Y6-opt (6 redshift bins)	**0.0162**

Table 5.4: Summary of the results of the Fisher forecast code applied to the Y3 BAO sample and different selections of the Y6 one. We include the cases of running the code on the preliminary Y6 sample (5 and 6 redshift bins) and the optimal Y6 sample (5 and 6 redshift bins). As expected, the most appropriate sample to measure the BAO feature is the optimal one with 6 redshift bins (bolded case).

photo-z precision is worse, but the forecast for the BAO measurement is significantly better. In Table 5.4 we summarize the results of the forecast applied to the different samples considered in this chapter. In the Y6, we expect the error associated to the BAO distance measurement to be reduced by about a 25% with respect to the Y3 analysis, which is an important increase in precision.

5.1.4 Improvement of the Y6 Quality Cuts

Additional color cuts were added to the selection of the Y6 BAO sample to remove objects with badly measured magnitudes in some band, which produced poorly estimated values of the photo-z. These new magnitude-based cuts ensure a signal-to-noise greater than 3 at grz bands, and are given by

$$g < 25.5, \tag{5.29}$$
$$r < 25, \tag{5.30}$$
$$z < 24. \tag{5.31}$$

The star-galaxy separator was modified from the original EXT_MASH ≥ 3, see eq. (5.2), to

$$\text{EXT\_MASH} = 4, \tag{5.32}$$

which allows us to better remove the remaining stellar contamination. The Y6 BAO mask was also slightly modified in order to remove regions with globular clusters. The new angular mask has an area of 4,273.42 deg^2, compared to the original one, which was 4,357.01 deg^2. Around 5% of the galaxies of the Y6 optimal sample were removed with the combined effect of applying the new selection cuts and the modified angular mask. In Table 5.5 we display the properties of the final Y6 BAO sample. Comparing the results displayed in this table with the ones of Table 5.3, we find that the average redshifts, number densities, σ_{68} and W_{68} were not affected.

Bin	$\langle z \rangle$	$N_{\rm gal}$	σ_{68}	W_{68}
$0.6 < z < 0.7$	0.654	2,854,542	0.0232	0.0501
$0.7 < z < 0.8$	0.752	3,266,097	0.0254	0.0577
$0.8 < z < 0.9$	0.844	3,898,672	0.0292	0.0626
$0.9 < z < 1.0$	0.929	3,404,744	0.0358	0.0770
$1.0 < z < 1.1$	1.013	1,752,169	0.0403	0.0861
$1.1 < z < 1.2$	1.107	761,332	0.0415	0.0927

Table 5.5: Main properties of the Y6 BAO sample as a function of redshift: number of galaxies and redshift properties. The sample covers 4,273.42 deg^2, with a total of 15,937,556 galaxies. The numbers are similar to those displayed in Table 5.3, since the additional quality cuts and the slightly modified mask barely changed our Y6 sample.

5.1.5 Photometric Redshifts

In Fig. 5.5 we show the redshift distributions of the optimal BAO sample, obtained with several different methods that we describe throughout this section. DNF provides two alternative estimations of the redshift distributions: $n(z_{\rm mc})$ and DNF PDF. The first one is obtained as histograms of DNF $z_{\rm mc}$ in redshift bins defined by $z_{\rm mean}$ (see section 4.2.3 for further details on these variables), whereas the second one is obtained as the stacking of individual galaxy PDFs [84]. These two are shown in Fig. 5.5 as blue histograms and green lines, respectively. We find that the distributions of DNF $z_{\rm mc}$ are quite smooth, being the last redshift bin the noisiest one (the same happened in the Y3 case, see Fig. 4.2). This is what we expected, since the last redshift bin is the one with the lowest number density, and also the one for which it is more complicated to estimate the photo-z (there are fewer galaxies in the DNF spectroscopic training sample at higher redshifts), i.e., we expect the photo-z quality to decrease at large redshifts, and also the redshift distributions to be noisier. On the other hand, we find that DNF PDF is qualitatively similar to $n(z_{\rm mc})$, but smoother.

In this section we perform different tests to validate the redshift distributions of the BAO sample. First, we validate the photo-z of our BAO sample, DNF $z_{\rm mc}$, using VIPERS. Then, we use both DNF $z_{\rm mc}$ and DNF PDF as template shapes for the redshift distributions, and shift and stretch them to match the distributions obtained using spectroscopic samples (namely VIPERS $z_{\rm spec}$ and clustering redshift using BOSS and eBOSS galaxies, which are two consistent alternative methods). Finally, we test the effect of the redshift distributions on the position of the BAO peak, and we find it is negligible.

5.1.5.1 Validation of the Photo-z Using VIPERS

For the validation of the photo-z we use VIPERS, which is a complete spectroscopic sample for redshifts above 0.5 and up to $i = 22.5$, see [85] (the same i-magnitude

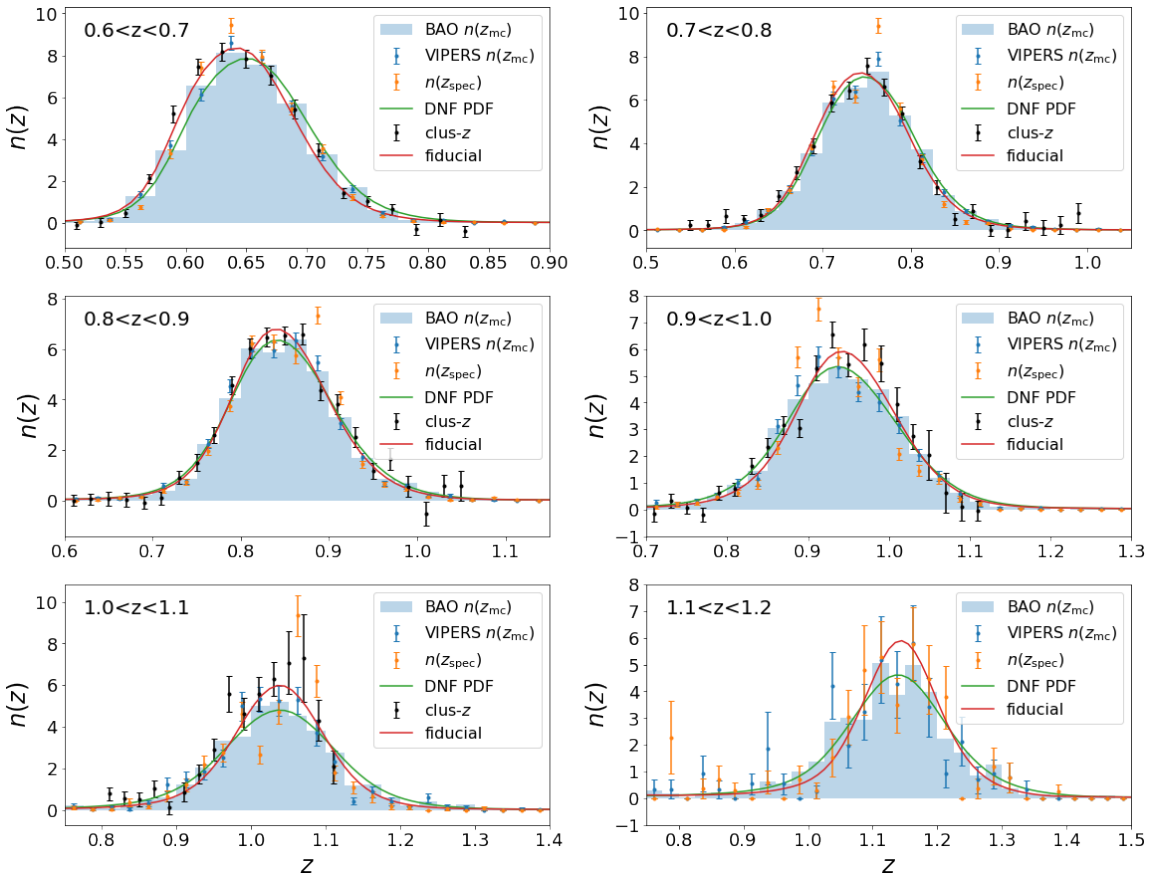

Figure 5.5: Redshift distributions of the Y6 BAO sample (computed with DNF) and VIPERS, together with those obtained with clustering redshift, and also the fiducial choice for the Y6 analysis. In the case of the BAO sample, we show both DNF $n(z_{\rm mc})$ and the stacking of DNF PDF (blue histograms and green lines, respectively). In the case of VIPERS, we include the distributions of $z_{\rm mc}$ and $z_{\rm spec}$ (blue and orange error-bars, respectively). In the case of clustering-z (black error-bars), we only show them for the first 5 redshift bins, since it was not possible to compute it for the last one because of the lack of spectroscopic galaxies in that bin. The fiducial choice (red lines) corresponds to the redshift distributions of DNF PDF but shifted and stretched with respect to clustering-z in the first 4 redshift bins, and with respect to VIPERS $n(z_{\rm spec})$ in the last 2, following the methodology described in section 5.1.5.3.

Bin	χ^2/dof	p-value
$0.6 < z < 0.7$	1.29	0.23
$0.7 < z < 0.8$	0.87	0.57
$0.8 < z < 0.9$	1.08	0.37
$0.9 < z < 1.0$	1.68	0.08
$1.0 < z < 1.1$	1.54	0.12
$1.1 < z < 1.2$	1.07	0.38

Table 5.6: Reduced χ^2 between the redshift distributions, computed as $n(z_{\rm mc})$, of the Y6 BAO sample and VIPERS.

limit of our BAO sample). The idea is to compare the redshift distributions of the BAO sample, computed with DNF $z_{\rm mc}$, with those of VIPERS (actually using the subsample of VIPERS matched to the BAO sample), also computed with DNF $z_{\rm mc}$, and check their compatibility. If they are statistically compatible, that means we can use the redshift distributions computed with VIPERS $z_{\rm spec}$ as the true redshift distributions of the BAO sample.

The first step to validate the photo-z is to select those galaxies from VIPERS that are also part of the BAO sample. Hereafter, we refer to this sample as VIPERS for simplicity, even though it is actually the matching between VIPERS and the BAO sample. After matching with the BAO sample, we end up with 11,202 VIPERS galaxies, i.e., VIPERS represents, approximately, a 0.066% of the total number of galaxies in the BAO sample. It is also worth mentioning that, in order to take into account the spectroscopic success ratio of VIPERS, which is encoded in the variable SSR of the VIPERS catalog [85], we must weight each VIPERS galaxy with 1/SSR.

As we already mentioned, the idea behind the validation of the photo-z using VIPERS is to check that $n(z_{\rm mc})$ for both BAO and VIPERS are compatible, which would allow us to use $n(z_{\rm spec})$ as the true redshift distributions of our BAO sample (these three redshift distributions shown in Fig. 5.5). To quantitatively compare the distributions of DNF $z_{\rm mc}$ of the BAO sample and VIPERS, we calculate the χ^2 between them as

$$\chi^2 = \sum_i \frac{\left[n_{\rm BAO}(z_i) - n_{\rm VIPERS}(z_i)\right]^2}{\Delta n_{\rm VIPERS}(z_i)^2}, \tag{5.33}$$

where the sum over i means summing over histogram bins, and $\Delta n(z)$ is the shot-noise contribution to the error in the redshift distributions (which is negligible for the BAO sample, because of the large number of galaxies compared to that of VIPERS). In Table 5.6 we show the reduced χ^2 obtained using eq. (5.33), and also their corresponding p-values. All the χ^2 and p-values show that both samples are compatible. However, in order to obtain these χ^2 we had to compute the redshift distributions of DNF $z_{\rm mc}$ for a certain histogram binning. In Fig. 5.6 we show the reduced χ^2 using different bin sizes but in the same redshift range. We find that the χ^2 between the two redshift distributions does not depend on the binning, i.e., we conclude they

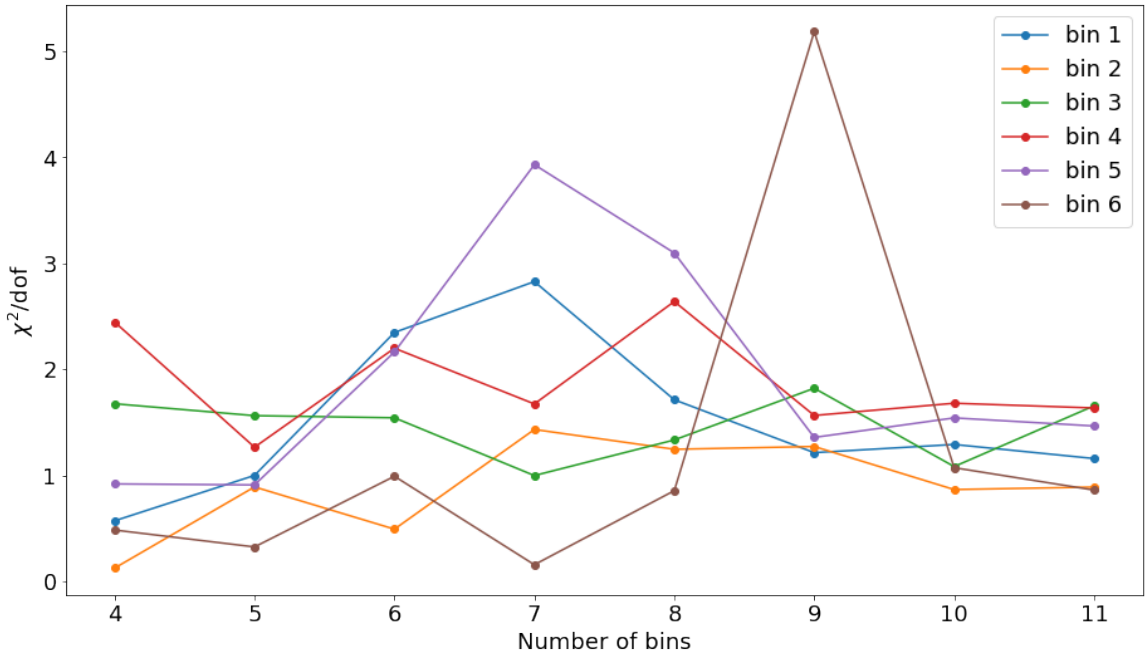

Figure 5.6: Reduced χ^2 between the redshift distributions of the BAO sample and VIPERS as a function of the number of bins in $z_{\rm mean}$. Each color represents one redshift bin, from 1 to 6.

are compatible, which allows us to use the distributions of $z_{\rm spec}$ as the true redshift distributions of our BAO sample.

Even though we just showed that the distributions of $z_{\rm mc}$ for the BAO sample and VIPERS are compatible, the distributions of $z_{\rm spec}$ are quite noisy, particularly in the last redshift bin (see Fig. 5.5). This noise is due to the small number of galaxies of VIPERS (11,202 galaxies in total), and can be complicated to deal with during the analysis, e.g., when computing the $w(\theta)$ template. Also, we find that DNF $z_{\rm mc}$ and PDF have similar properties, but both of them are significantly wider than $z_{\rm spec}$ (see the results displayed in Table 5.7). To attack the problem of the small statistics of $z_{\rm spec}$, for the fiducial analysis on the data we decided to use a modified version of the DNF redshift distributions as our default choice: shifted and stretched to match the properties of $z_{\rm spec}$. Therefore, both DNF $z_{\rm mc}$ and DNF PDF are used as templates for the shape of our final redshift distributions.

5.1.5.2 Clustering Redshift

There are alternative ways to estimate the redshift distributions of our BAO sample, such as the so-called clustering redshift technique, or clustering-z. Clustering redshifts make use of the fact that galaxies with unknown redshifts reside in the same structures as galaxies that have known redshifts. Thus, spatial cross-correlations can be used to estimate the redshift distribution of the sample with unknown redshifts. The basic

Bin	$\langle z \rangle$				W_{68}			
	$n(z_{\rm mc})$	$n(z_{\rm spec})$	PDF	clus-z	$n(z_{\rm mc})$	$n(z_{\rm spec})$	PDF	clus-z
1	0.653	0.650	0.658	0.646	0.0503	0.0449	0.0509	0.0474
2	0.751	0.746	0.754	0.744	0.0581	0.0503	0.0571	0.0576
3	0.844	0.850	0.847	0.854	0.0631	0.0561	0.0652	0.0623
4	0.930	0.931	0.934	0.939	0.0771	0.0597	0.0787	0.0649
5	1.015	1.024	1.020	1.014	0.0844	0.0674	0.0889	0.0644
6	1.105	1.110	1.111	-	0.0917	0.0792	0.0958	-

Table 5.7: Mean and width of the redshift distributions of DNF $z_{\rm mc}$, $z_{\rm spec}$, DNF PDF and clustering-z. For the case of clustering-z, there is no entry in the last redshift bin because of the lack of spectroscopic galaxies in the redshift range $1.1 < z < 1.2$, which did not allow us to compute the clustering-z redshift distribution for that bin.

method bins the sample with known redshifts in z and then spatially cross-correlates each of these bins against the unknown sample. The amplitude of the resultant correlation can then be used to estimate the amount of redshift overlap and thus the redshift distribution of the sample with unknown redshifts, i.e., the BAO sample. The clustering redshift technique was already validated and applied to the Y3 MagLim sample in order to calibrate its redshift distributions, see [123]. In that case and also in this one, spectroscopic galaxies from BOSS and its extension, eBOSS, are used to cross-correlate with our sample.

In Fig. 5.5 we show the redshift distributions for clustering redshift (black errorbars). We did not include the plot for the last redshift bin because the galaxy sample used as the spectroscopic catalog, i.e., the combination of BOSS and eBOSS, does not have enough galaxies in the redshift range $1.1 < z < 1.2$, and, therefore, it was not possible to compute the redshift distribution for that one bin. We find that clustering-z is compatible with $z_{\rm spec}$, see Table 5.7. Since we have two independent determinations of the redshift distributions that agree, i.e., VIPERS $z_{\rm spec}$ and clustering-z, we consider them as validated. However, as in the case of $z_{\rm spec}$, clustering-z is quite noisy. Therefore, we use the redshift distributions from clustering-z to shift and stretch DNF $z_{\rm mc}$ and DNF PDF, in the same way as we use $n(z_{\rm spec})$. For the fiducial analysis on the data, we use DNF PDF shifted and stretched with respect to clustering-z in the first 4 redshift bins, and with respect to $n(z_{\rm spec})$ in the last 2, as our default choice[5].

5.1.5.3 Shift and Stretch Algorithm

In this section we describe the algorithm developed to perform the shift and stretch of DNF $z_{\rm mc}$ and DNF PDF. In general, we perform the shift and stretch of a given

[5] For the analysis of the COLA mocks, we use $n(z_{\rm mc})$ as the default redshift distributions, since these are the ones they were generated with.

$n(z)$ in two different steps:

1. Shift of the original $n(z)$. The shifted redshift distribution is, simply, given by

$$n_{\rm shifted}(z, \Delta z) = n(z - \Delta z). \tag{5.34}$$

2. Stretch of the shifted $n(z)$. The shifted and stretched redshift distribution is given by

$$n_{\rm 2-param}(z, \Delta z, \sigma_z) = n_{\rm shifted}(\sigma_z(z - \langle z \rangle_{\Delta z}) + \langle z \rangle_{\Delta z}, \Delta z), \tag{5.35}$$

where

$$\langle z \rangle_{\Delta z} \equiv \int dz\, n_{\rm shifted}(z, \Delta z). \tag{5.36}$$

We, then, compute the best fit parameters Δz and σ_z by minimizing

$$\chi^2(\Delta z, \sigma_z) = \sum_i \left[\frac{n_{\rm 2-param}(z_i, \Delta z, \sigma_z) - n_{\rm ref}(z_i)}{\Delta n_{\rm ref}(z_i)} \right]^2. \tag{5.37}$$

Since the redshift distributions we shift and stretch are either DNF $z_{\rm mc}$ or DNF PDF, we neglect their contribution to the denominator of the previous expression, since their shot-noise is much smaller than that of the reference redshift distribution, which is either VIPERS or clustering-z. This methodology is similar to the one used for the DES Y3 3×2pt analysis, see [28]. These shift and stretch parameters appear because of our uncertainty in the photo-z, and this implementation is particularly useful since it allows us to fit for them when running the 3×2pt chains[6].

Once we have run the shift and stretch code for all the different combinations of redshift distributions and also for the fiducial choice, we end up with 5 sets of $n(z)$:

- $n(z_{\rm mc})$ shifted and stretched with respect to $n(z_{\rm spec})$.
- $n(z_{\rm mc})$ shifted and stretched with respect to clustering-z.
- DNF PDF shifted and stretched with respect to $n(z_{\rm spec})$.
- DNF PDF shifted and stretched with respect to clustering-z.
- Fiducial: DNF PDF shifted and stretched with respect to clustering-z in the first 4 redshift bins, and with respect to $n(z_{\rm spec})$ in the last 2. We explicitly show these redshift distributions in Fig. 5.5 (red lines), since they are the fiducial ones for the analysis on the data.

[6] The 3×2pt chains include these parameters as free parameters with Gaussian priors. Since they are nuisance parameters we simply marginalize over them, as we will see in section 4.4.2.

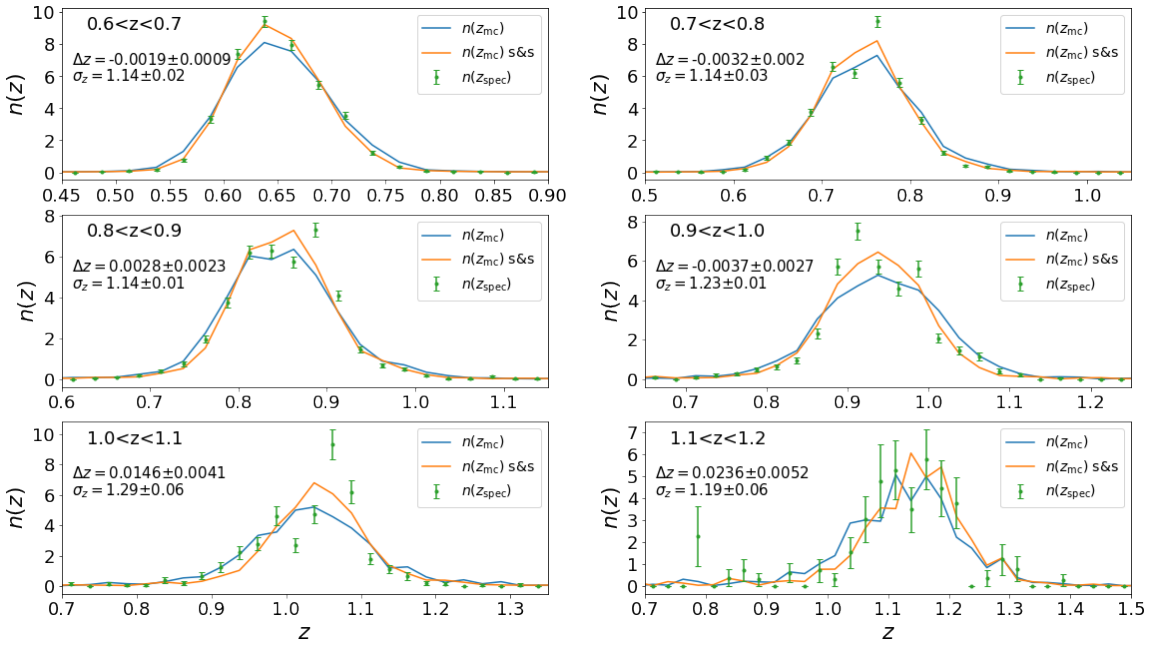

Figure 5.7: Example of running the shift and stretch code over $n(z_{\rm mc})$, using $n(z_{\rm spec})$ as the reference redshift distributions. We find that the shifted and stretched redshift distributions are more similar to $n(z_{\rm spec})$ than the original ones, as expected.

Since we have 5 different cases, here we explicitly show the results of running the shift and stretch code only for the first one, as an example. In Fig. 5.7 we show the results of running our shift and stretch code on the redshift distributions $n(z_{\rm mc})$ using as reference $n(z_{\rm spec})$. We also display the values for the best fit parameters Δz and σ_z, with their errors. We find that for the first, second and fourth redshift bins, the $n(z_{\rm mc})$ redshift distributions had to be shifted towards lower redshifts ($\Delta z < 0$), whereas for the remaining bins they had to be shifted towards larger redshifts ($\Delta z > 0$). As for the stretch parameter, we find that $\sigma_z > 1$ for all the redshift bins, as expected, i.e., the redshift distributions for all the redshift bins had to be stretched in order to fit $n(z_{\rm spec})$, which is narrower.

In Fig. 5.8 we show the average redshift for each $n(z)$ (top panel), computed using eq. (4.4), and also their width (bottom panel), computed using eq. (4.6). We find that $n(z_{\rm mc})$ and DNF PDF follow a similar trend (blue and green), as we already discussed earlier. We also find that the two redshift distributions shifted and stretched with respect to VIPERS follow the same trend as $n(z_{\rm spec})$ (orange, purple and pink), as expected. Also, when shifting and stretching with respect to clustering-z, we find that our redshift distributions also follow a similar trend to that of clustering-z, in both average redshift and width (black, brown and grey). All these different choices of redshift distributions are used in section 5.1.5.4 to test the impact of the $n(z)$ in the BAO measurement.

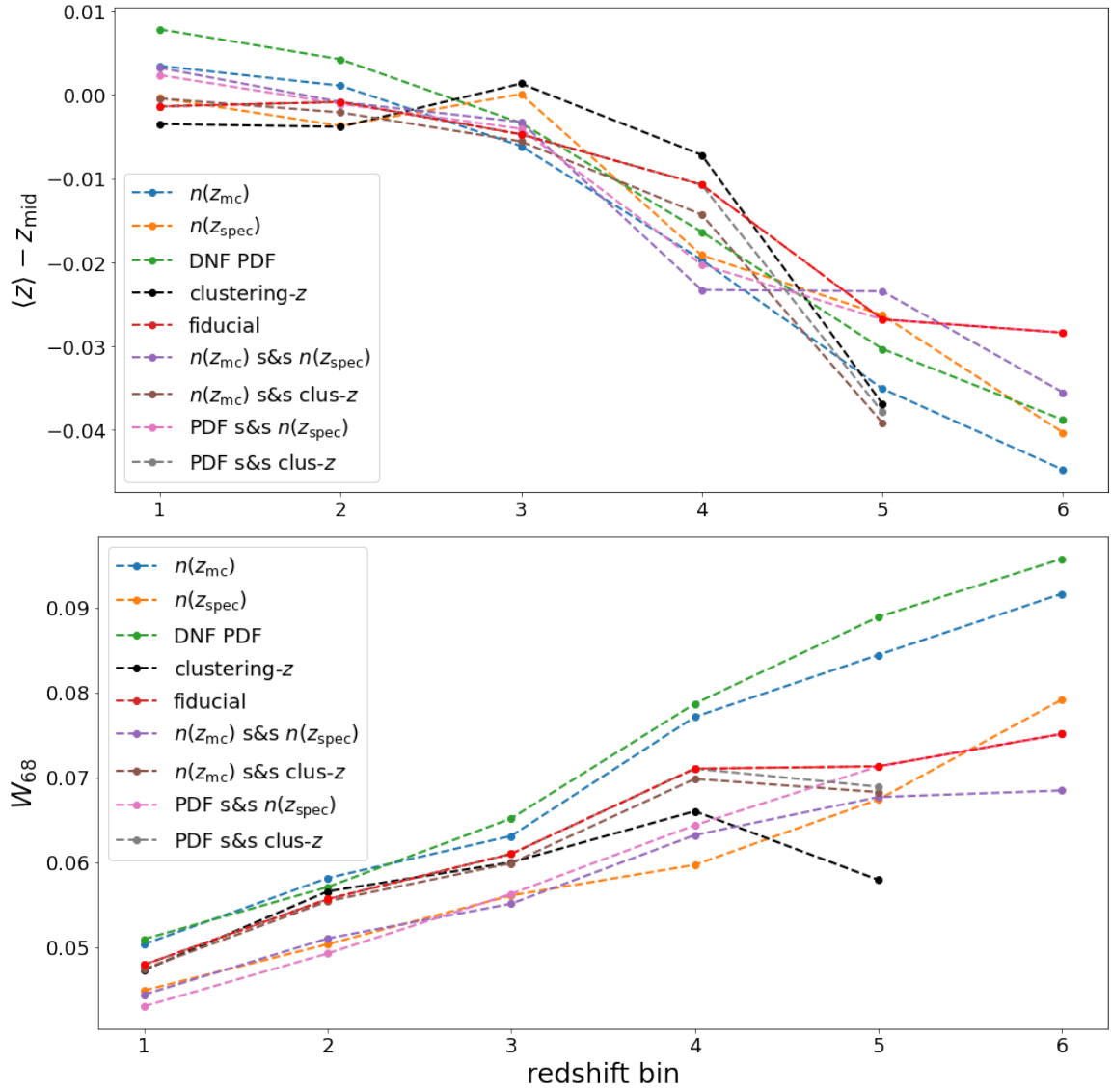

Figure 5.8: **Top panel:** Average redshift of the different redshift distributions of the Y6 analysis , computed using eq. (4.4). For visualization purposes, we subtracted the middle redshift, which is given by the average of the limits for each redshift bin (0.65, 0.75, 0.85, 0.95, 1.05 and 1.15, respectively). **Bottom panel:** Width of the different redshift distributions of the Y6 analysis, computed using eq. (4.6). Cases included in this plot: $n(z_{\rm mc})$ (blue), $n(z_{\rm spec})$ (orange), DNF PDF (green), clustering-z (black), fiducial choice (red), $n(z_{\rm mc})$ shifted and stretched with respect to $n(z_{\rm spec})$ (purple), $n(z_{\rm mc})$ shifted and stretched with respect to clustering-z (brown), DNF PDF shifted and stretched with respect to $n(z_{\rm spec})$ (pink) and DNF PDF shifted and stretched with respect to clustering-z (grey). The fiducial choice is a combination of the grey (first 4 redshift bins) and the pink cases (last 2 redshift bins).

5.1.5.4 Effect of the Redshift Distributions in the BAO Measurement

The redshift distributions are necessary to compute the template $w(\theta)$ we use for the BAO fits, see eq. (4.34). In this section we study the impact of using different redshift distributions in the determination of the position of the BAO peak. The idea is to compute a default $w(\theta)$ template with $n(z_{\rm mc})$ as its redshift distributions, and then run the BAO-fitting pipeline to fit this $w(\theta)$ with the template computed with different redshift distributions, but with the same cosmological parameters (Mice cosmology). This allows us to estimate how much α deviates from 1 as a function of the $n(z)$.

In Table 5.8 we show the results for this test, for the cases of fitting the default template, computed with $n(z_{\rm mc})$, with another one computed with $n(z_{\rm spec})$, DNF PDF, the fiducial choice and the other 4 shifted and stretched cases. The column labeled as "$n(z_{\rm mc})$" represents the case of fitting the default $w(\theta)$ template with itself, and we display it as a consistency check (it is consistent with 1 up to, at least, the third decimal digit). For all the other columns, we find that, as expected, the impact in the position of the BAO peak is negligible, i.e., all the numbers displayed in Table 5.8 are very close to 1 within the errors (the effect is always smaller than 0.25σ). We also find that the uncertainty in α is consistent between all the different columns, i.e., it is not affected by the redshift distributions assumed to generate the template used to run the fits. We conclude that our BAO measurement is robust against different estimations of the redshift distributions.

5.2 Tests on Simulations

The BAO-fitting pipeline described in section 4.4 was already tested and validated using the Y3 COLA mocks, see section 4.5, where we found its results are very robust against the selection of the reference cosmology and also against changes in the default settings for the fits. Although it was not the default pipeline used for the official DES Y3 analysis, it is the one being used for the official Y6 one. The code was developed during the last stages of the Y3 analysis and upgraded in several ways during the Y6 analysis, adding a wide variety of different options to perform the fits, as detailed in section 4.4. For instance, it includes the option to fit for the linear galaxy bias, which needs to be measured from the COLA mocks, and the option to transform the $w(\theta)$ into C_ℓ, which we also showed in section 4.4.1.1. In this section we test the effect of observational systematics using a set of 1,000 contaminated lognormal mocks by comparing the BAO fit results with those obtained from a set of 1,000 uncontaminated lognormal mocks. We also run the BAO-fitting pipeline on the Y6 COLA mocks, and compare the results with those of the Y3 analysis.

Bins	$n(z_{\rm mc})$	$n(z_{\rm spec})$	DNF PDF	fiducial	$z_{\rm mc}$ s&s $z_{\rm spec}$	$z_{\rm mc}$ s&s clus-z	PDF s&s $z_{\rm spec}$	PDF s&s clus-z
All	1.000±0.018	1.000±0.018	0.999±0.018	1.000±0.018	1.000±0.018	1.001±0.018	1.002±0.018	1.002±0.018
23456	1.000±0.020	1.000±0.020	0.999±0.020	0.999±0.020	1.000±0.020	1.001±0.020	1.001±0.020	1.001±0.020
13456	1.000±0.020	1.000±0.020	0.998±0.020	0.999±0.020	1.000±0.020	1.000±0.020	1.001±0.020	1.001±0.020
12456	1.000±0.021	1.002±0.020	0.998±0.020	1.000±0.020	1.001±0.020	1.002±0.020	1.002±0.020	1.002±0.021
12356	1.000±0.020	0.998±0.020	0.998±0.020	1.001±0.020	0.999±0.020	1.000±0.020	1.004±0.020	1.004±0.020
12346	1.000±0.020	1.002±0.020	0.999±0.020	1.001±0.020	1.002±0.020	1.002±0.020	1.001±0.020	1.001±0.020
12345	1.000±0.019	1.001±0.019	0.999±0.019	1.001±0.019	1.000±0.019	1.002±0.019	1.002±0.019	1.003±0.019
1	1.000±0.049	1.003±0.049	0.995±0.049	1.007±0.049	1.001±0.048	1.001±0.048	1.007±0.049	1.007±0.049
2	1.000±0.044	1.002±0.044	1.000±0.044	1.006±0.044	1.004±0.044	1.006±0.044	1.005±0.044	1.006±0.044
3	1.000±0.038	0.996±0.038	1.000±0.038	1.001±0.038	0.999±0.037	1.000±0.038	1.001±0.038	1.001±0.038
4	1.000±0.037	1.009±0.037	1.001±0.037	0.997±0.036	1.004±0.037	1.005±0.037	0.996±0.037	0.997±0.037
5	1.000±0.041	0.994±0.041	0.997±0.041	0.997±0.041	0.994±0.041	0.997±0.041	1.006±0.042	1.005±0.042
6	1.000±0.059	0.998±0.059	0.995±0.059	0.993±0.059	0.999±0.059	0.993±0.058	0.998±0.059	0.995±0.059

Table 5.8: Effect of the redshift calibration in the BAO measurement ($\alpha \pm \sigma_\alpha$). Each column displays the results of fitting a theoretical $w(\theta)$ computed using $n(z_{\rm mc})$ as its redshift distributions with another template with redshift distributions given by its column name. The cases of fitting all the redshift bins simultaneously, removing one at a time and fitting each of them individually are shown. Since there is no clustering-z measurement for the last redshift bin, for the cases shifted and stretched with respect to clustering-z we simply used the original redshift distribution in that bin. We find that the impact in the BAO distance measurement and its error due to assuming different redshift distributions is negligible.

5.2.1 Effect of Observational Systematics

Lognormal mocks are synthetic galaxy catalogs that are commonly used in large-scale structure surveys to test and validate analysis pipelines, such as those used to measure the BAO signal or to perform galaxy clustering analysis. These mocks are constructed by applying a lognormal transformation to the matter density field, which has the effect of smoothing out the density fluctuations on small scales while preserving the large-scale structure. The lognormal transformation is applied to the linear matter density field, which is then used to generate a synthetic galaxy catalog by assigning galaxies to the matter density peaks.

Lognormal distributions have shown to be an excellent approximation to cosmological fields ([124, 125]). They have been used as an useful tool for cosmological simulations, much less computationally expensive in comparison to the COLA mocks we described in section 4.3. We do not use them as the fiducial simulations for our analysis because they are less realistic than N-body simulations[7]. However, they can be particularly useful for specific tests, just like testing the impact of observational systematics on the BAO measurement. The lognormal-distributed density contrast $\delta(\boldsymbol{x})$ is related to a Gaussian field $G(\boldsymbol{x})$ as

$$\delta(\boldsymbol{x}) = \exp\left(-\sigma_G^2 + G(\boldsymbol{x})\right) - 1, \tag{5.38}$$

where $\sigma_G^2 \equiv \langle G^2(\boldsymbol{x})\rangle$ and $G(\boldsymbol{x})$ is given by

$$G(\boldsymbol{x}) \equiv \log\left[1+\delta(\boldsymbol{x})\right] - \langle \log\left[1+\delta(\boldsymbol{x})\right]\rangle. \tag{5.39}$$

It can be easily checked that

$$\langle G(\boldsymbol{x})\rangle = 0. \tag{5.40}$$

The pre-factor with the variance of the Gaussian field σ_G^2 ensures that the mean of $\delta(\boldsymbol{x})$ vanishes. It can be noted that $\delta(\boldsymbol{x}) \geq -1$ by definition, which is the natural constraint for the density contrast field.

In this section we use a set of 1,000 lognormal mocks, generated using the cosmological parameters for Planck (see Table 4.2), to test how robust the BAO measurement is against observational systematics. These lognormal mocks, which are originally free of systematic contamination, are contaminated following the methodology described in [72]. This means that we end up with 1,000 uncontaminated plus 1,000 contaminated lognormal mocks, and we run our BAO-fitting pipeline on both sets separately. In Fig. 5.9 we show the mean $w(\theta)$ for the 1,000 lognormal mocks, for the uncontaminated and the contaminated cases. The effect of systematics is an effective increase in the clustering amplitude at all scales, leaving the BAO peak position unchanged. In the first few redshift bins, the effect of the contamination is smaller, and it increases as the redshift increases. This is exactly the same behaviour we found in the clustering measurements of the Y3 BAO sample, see Fig. 4.8.

[7] Lognormal mocks may not accurately capture the nonlinear evolution of the density field on small scales, and the galaxy assignments may not accurately reflect the true relationship between galaxies and matter.

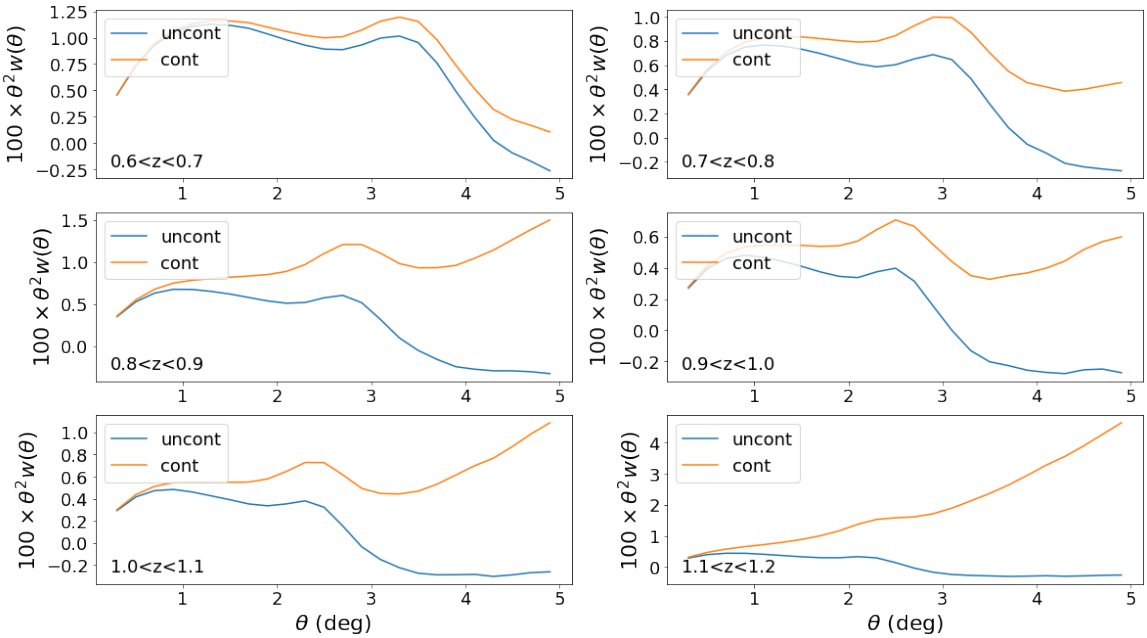

Figure 5.9: Mean angular correlation function of the 1,000 lognormal mocks, for the uncontaminated (blue lines) and contaminated (orange lines) cases. The $w(\theta)$ are shown for the 6 redshift bins. We find that the effect of the systematics is an increase in the clustering amplitude at all scales. This increase in the clustering amplitude is more important as we move towards larger redshifts, being the last redshift bin the one with the most shifted amplitude. However, the systematic contamination does not change the position of the BAO peak.

In Table 5.9 we display the results of the BAO measurement for both the 1,000 uncontaminated (top) and the 1,000 contaminated (bottom) lognormal mocks using the Planck template. We obtain the same BAO distance measurement (same α and its error, σ_α) for both the uncontaminated and the contaminated cases[8]. We also find that the mean χ^2 does not increase because of the contamination; in fact, all the columns displayed in Table 5.9 are very consistent between the uncontaminated and contaminated cases. Therefore, we conclude that the BAO measurement is very robust against observational systematics, i.e., the precision of the measurement, the position of the peak and the goodness-of-fit are not affected by them.

5.2.2 Tests on the COLA Mocks

The COLA mocks are the fiducial set of simulations used for the DES Y6 BAO analysis (analogously to the Y3 analysis, see section 4.5). These mocks are necessary to test the BAO-fitting pipeline, and also to run all the robustness tests and the pre-unblinding tests before we are allowed to unblind our results. The Y6 COLA mocks were created as explained in section 4.3: we end up having 1,952 galaxy mocks designed to mimic the DES Y6 BAO sample properties over its full photometric redshift range, $0.6 < z_{\rm ph} < 1.2$.

Similarly to what we did in section 4.5 for Y3, we need to measure the clustering signal for the 1,952 COLA mocks, i.e., the angular correlation functions. The clustering measurements were computed using the Landy-Szalay estimator implemented in the software CUTE[9]. The angular correlation functions for the 1,952 COLA mocks, together with its mean, are shown in Fig. 5.10. We also include the three pre-unblinding clustering measurements at low angles used as input to the COLA code. As we already discussed, these measurements at such low scales do not compromise our blinding protocol, since they contain no BAO information. From these angular correlation functions we can also compute the covariance matrix of the COLA mocks using eq. (4.5). The correlation matrix (i.e., the normalized covariance matrix, see eq. (4.46)) of the $w(\theta)$ of the COLA mocks is shown in Fig. 5.11 (upper-triangle region). As in the case of the Y3 analysis, the default covariance matrix we use for the BAO fits in the Y6 is not the COLA covariance we just calculated, but rather the analytical one computed with CosmoLike. The correlation matrix for CosmoLike is also shown in Fig. 5.11 (lower-triangle region). We find that the cross-correlations between non-contiguous redshift bins are negligible for the CosmoLike covariance matrix, whereas for the COLA covariance they are not (similarly to what we found in Fig. 4.7).

Now that we have computed the clustering measurements for the 1,952 Y6 COLA

[8] We note that, however, the best fit alpha is slightly more biased with respect to 1 than that of the COLA mocks (see section 4.5 for the results on the Y3 COLA mocks). This is somewhat expected, since the BAO feature is not as precisely modeled in lognormal mocks as in the case of the COLA mocks (which are N-body simulations).

[9] https://github.com/damonge/CUTE

Bins	$\langle\alpha\rangle$	$\sigma_{\rm std}$	σ_{68}	$\langle\sigma_\alpha\rangle$	f. enc.$\langle\alpha\rangle$	$\langle d_{\rm norm}\rangle$	$\sigma_{d_{\rm norm}}$	$\langle\chi^2\rangle$/dof
Uncontaminated								
All	1.010	0.023	0.022	0.020	63.2%	-0.017	1.165	110.7/107
23456	1.009	0.025	0.023	0.021	63.8%	-0.019	1.182	92.0/89
13456	1.010	0.025	0.023	0.021	63.7%	-0.031	1.176	92.0/89
12456	1.010	0.027	0.026	0.022	62.4%	-0.047	1.200	92.2/89
12356	1.009	0.025	0.024	0.021	63.1%	-0.028	1.154	92.1/89
12346	1.011	0.025	0.023	0.021	64.6%	-0.002	1.158	92.1/89
12345	1.010	0.025	0.024	0.021	63.4%	-0.012	1.155	91.9/89
1	1.013	0.055	0.056	0.051	64.3%	-0.011	1.199	17.3/17
2	1.011	0.054	0.051	0.048	64.5%	0.010	1.207	17.2/17
3	1.009	0.049	0.045	0.043	65.5%	0.015	1.156	17.4/17
4	1.006	0.048	0.044	0.040	64.6%	-0.030	1.216	17.3/17
5	1.001	0.047	0.043	0.039	65.0%	-0.044	1.230	17.2/17
6	1.006	0.053	0.053	0.046	63.2%	-0.026	1.298	17.2/17
Contaminated								
All	1.010	0.023	0.022	0.020	62.5%	-0.027	1.171	110.8/107
23456	1.009	0.025	0.024	0.021	63.2%	-0.023	1.186	92.1/89
13456	1.010	0.025	0.023	0.021	64.2%	-0.033	1.177	92.2/89
12456	1.009	0.027	0.025	0.022	63.2%	-0.046	1.205	92.3/89
12356	1.009	0.025	0.025	0.022	62.1%	-0.028	1.160	92.2/89
12346	1.011	0.025	0.024	0.022	65.1%	-0.004	1.160	92.2/89
12345	1.010	0.025	0.024	0.021	63.5%	-0.012	1.153	91.9/89
1	1.014	0.055	0.056	0.051	64.7%	-0.011	1.184	17.2/17
2	1.011	0.055	0.051	0.048	65.2%	0.011	1.204	17.1/17
3	1.010	0.050	0.047	0.043	65.8%	0.012	1.172	17.4/17
4	1.006	0.048	0.043	0.040	65.1%	-0.023	1.208	17.5/17
5	1.003	0.047	0.043	0.040	64.7%	-0.052	1.218	17.2/17
6	1.005	0.053	0.050	0.046	64.8%	-0.030	1.274	17.4/17

Table 5.9: BAO fit results for the 1,000 lognormal mocks using the Planck template: mean and standard deviation of α, $\langle\alpha\rangle$ and $\sigma_{\rm std}$; width of the distribution, σ_{68}; mean error and fraction of mocks enclosed within it, $\langle\sigma_\alpha\rangle$ and f. enc.$\langle\alpha\rangle$; mean and the standard deviation of the pull statistics, $\langle d_{\rm norm}\rangle$ and $\sigma_{d_{\rm norm}}$; and mean χ^2 divided by the degrees of freedom. The results are shown for the cases of fitting the 6 redshift bins simultaneously, fitting removing one at a time and fitting each of them individually. **Top:** uncontaminated mocks. **Bottom:** contaminated mocks. We find that there is no difference in the BAO fit results between the uncontaminated and the contaminated mocks: both sets of mocks yield to the same distance measurement, $\langle\alpha\rangle$, with the same precision, $\langle\sigma_\alpha\rangle$. All the other columns are also consistent.

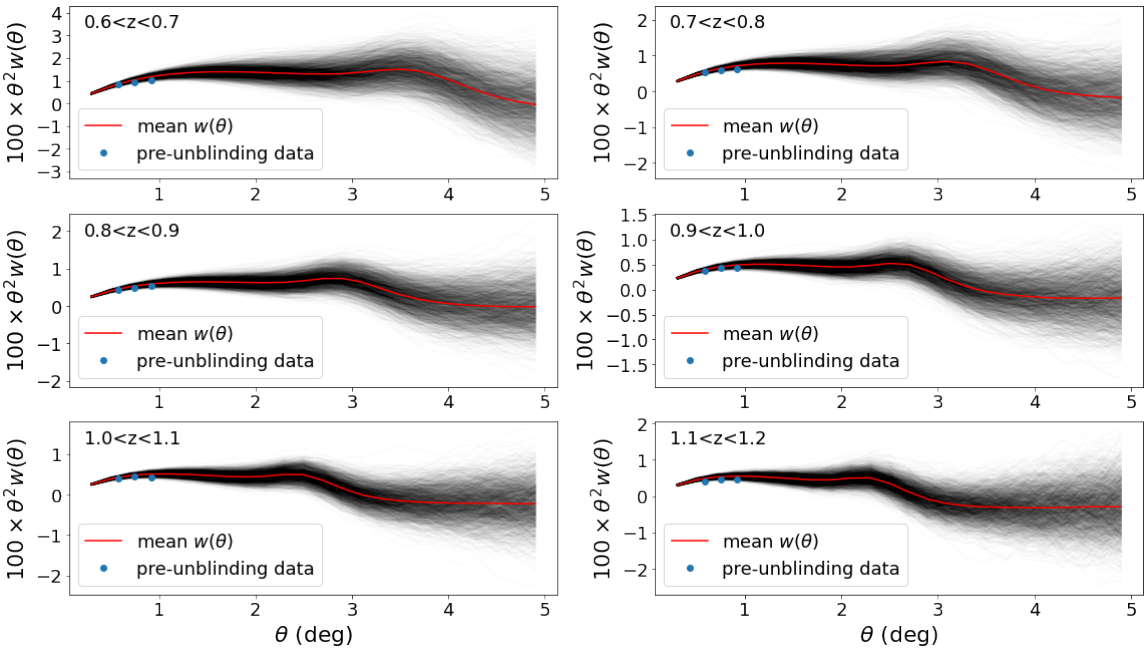

Figure 5.10: Angular correlation functions of the 1,952 COLA mocks (black curves) and mean $w(\theta)$ (red curve) for the 6 tomographic redshift bins. We also show the three pre-unblinding clustering measurements at low angles used as input to the COLA code. As we already discussed, these measurements at such low scales do not compromise our blinding protocol, since they contain no BAO information.

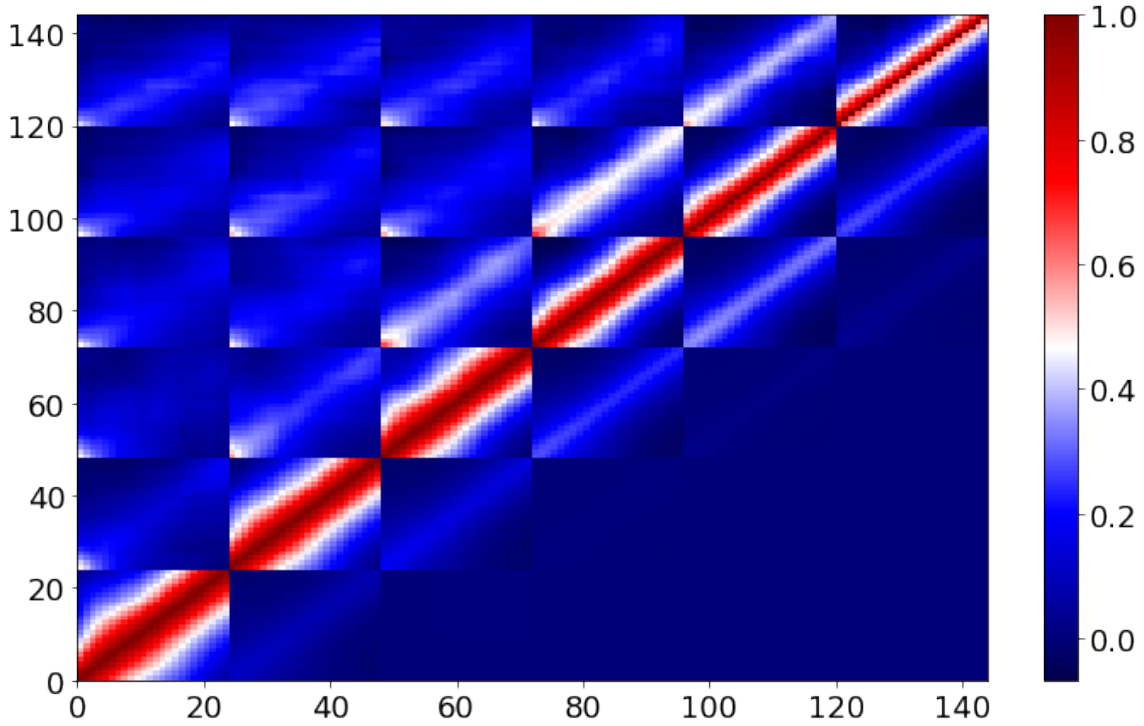

Figure 5.11: Normalized covariance matrix of the Y6 COLA mocks (upper-triangle region) and CosmoLike (lower-triangle region). In the case of CosmoLike, we show the one computed for Mice cosmology. We find that, as expected, their structure is very similar in the regions close to the diagonal, but they differ as we move to cross-correlations between more distant redshift bins. In the case of CosmoLike, the cross-correlations between non-contiguous redshift bins are negligible, whereas in the case of COLA they are not.

mocks, we can run the BAO fits on them. In Table 5.10 we display the BAO fit results on the mocks using the Mice template for the cases of fitting all the redshift bins simultaneously, removing one of them at a time and fitting each of them individually (the table above shows the results using the COLA covariance, whereas the one below shows the results using CosmoLike). The results displayed in this table are consistent with what we expected: the mean α is close to 1 (since the underlying cosmology of the mocks is the same as the one used for the template); the standard deviation of α is similar to the width of its distribution and the mean error; the fraction of mocks enclosed within $\langle\alpha\rangle \pm \langle\sigma_\alpha\rangle$ is close to 68%; and the mean χ^2 is very similar to the degrees of freedom for each case. However, when using the CosmoLike covariance matrix to fit the mocks we find that the χ^2 for the best-fit α are quite small compared to the degrees of freedom (this happens systematically for all the different entries of Table 5.10, CosmoLike covariance case). After performing several tests trying to understand the reason for this, we concluded that the Y6 COLA mocks have a slightly different covariance structure compared to that of the Y3 COLA mocks, likely due to the repetitions of the boxes used to create the simulations. However, this barely impacts the error-bar of the measurement, and we still use CosmoLike when running the BAO-fitting pipeline on the data.

The results displayed in Table 5.10 can be directly compared to the ones we obtained for the Y3 COLA mocks, see Table 4.3. The relative error we obtained in the Y3 for the case of fitting all the redshift bins simultaneously was $\langle\sigma_\alpha\rangle/\langle\alpha\rangle \sim 2.4\%$, whereas for the Y6 we find that it has a value of $\sim$1.9%, i.e., the error in the determination of the position of the BAO peak decreases by a $\sim$25% according to these results. This improvement is consistent with what we expected from the forecasts, see Table 5.4, where we found a decrease in the error of about 25% with respect to the Y3.

5.3 BAO Measurements on the Y6 Data (Blind)

The Y6 BAO analysis is ongoing, and it is still on its blinded phase. Therefore, in this section we do not show the clustering measurements on the data or the results of the BAO fits. However, we do show the blinded $w(\theta)$, and finish the chapter with a forecast of what we expect to measure for the Y6.

5.3.1 Clustering Signal of the Y6 BAO Sample (Blind)

In order to run the BAO fits on the data, we first need to measure the clustering signal of the Y6 BAO sample for all the redshift bins, i.e., the angular correlation functions. Here we do measure the $w(\theta)$ of the data, but we keep it blinded. Since we want to test the robustness of our clustering measurement, and also as a sanity check, in this section we show the comparison of the $w(\theta)$ computed with two different codes: TreeCorr and CUTE, i.e., the $\Delta w(\theta)$ between both. This allows us to test the

Bins	$\langle\alpha\rangle$	$\sigma_{\rm std}$	σ_{68}	$\langle\sigma_\alpha\rangle$	f. enc.$\langle\alpha\rangle$	$\langle d_{\rm norm}\rangle$	$\sigma_{d_{\rm norm}}$	$\langle\chi^2\rangle$/dof	p-value
COLA covariance									
All	1.006	0.019	0.019	0.018	67.8%	0.001	1.046	109.6/107	0.41
23456	1.007	0.020	0.020	0.019	66.8%	0.001	1.043	91.1/89	0.42
13456	1.007	0.021	0.020	0.020	68.2%	-0.004	1.046	91.1/89	0.42
12456	1.007	0.022	0.021	0.021	67.2%	-0.005	1.053	91.2/89	0.41
12356	1.005	0.021	0.021	0.020	66.9%	0.001	1.044	91.2/89	0.41
12346	1.006	0.021	0.021	0.020	67.2%	0.002	1.051	91.2/89	0.42
12345	1.006	0.020	0.020	0.019	66.6%	-0.003	1.044	91.1/89	0.42
1	1.002	0.048	0.045	0.049	71.6%	-0.023	1.030	17.2/17	0.44
2	1.000	0.046	0.044	0.043	67.3%	-0.004	1.117	17.2/17	0.44
3	1.004	0.041	0.038	0.038	68.8%	-0.001	1.094	17.3/17	0.44
4	1.010	0.040	0.037	0.037	67.9%	-0.011	1.091	17.3/17	0.43
5	1.007	0.041	0.038	0.038	68.6%	-0.014	1.063	17.3/17	0.43
6	1.008	0.049	0.046	0.047	68.9%	-0.024	1.102	17.4/17	0.43
CosmoLike covariance									
All	1.006	0.020	0.020	0.019	65.5%	-0.009	1.069	76.3/107	0.99
23456	1.006	0.021	0.021	0.020	67.4%	-0.010	1.038	59.0/89	0.99
13456	1.007	0.022	0.021	0.021	67.1%	-0.013	1.052	63.2/89	0.98
12456	1.006	0.023	0.022	0.021	64.7%	-0.017	1.088	64.2/89	0.98
12356	1.005	0.022	0.021	0.021	66.1%	-0.013	1.055	63.4/89	0.98
12346	1.005	0.022	0.021	0.020	65.3%	-0.004	1.079	65.0/89	0.97
12345	1.005	0.021	0.021	0.020	66.0%	-0.009	1.065	66.8/89	0.96
1	1.002	0.048	0.046	0.045	67.6%	-0.029	1.115	16.0/17	0.52
2	1.000	0.046	0.044	0.042	66.3%	-0.022	1.132	12.1/17	0.80
3	1.004	0.041	0.039	0.040	70.0%	-0.007	1.030	11.3/17	0.84
4	1.010	0.040	0.037	0.038	69.1%	-0.019	1.065	12.4/17	0.78
5	1.007	0.041	0.038	0.042	72.5%	-0.026	0.983	10.9/17	0.86
6	1.007	0.048	0.046	0.056	76.1%	-0.025	0.893	8.9/17	0.94

Table 5.10: BAO fit results for the 1,952 COLA mocks using the Mice template: mean and standard deviation of α, $\langle\alpha\rangle$ and $\sigma_{\rm std}$; width of the distribution, σ_{68}; mean error and fraction of mocks enclosed within it, $\langle\sigma_\alpha\rangle$ and f. enc.$\langle\alpha\rangle$; mean and the standard deviation of the pull statistics, $\langle d_{\rm norm}\rangle$ and $\sigma_{d_{\rm norm}}$; and mean χ^2 divided by the degrees of freedom. The results are shown for the cases of fitting the 6 redshift bins at a time, fitting removing one of them and fitting each of them individually. **Top:** using the COLA covariance. **Bottom:** using the CosmoLike covariance.

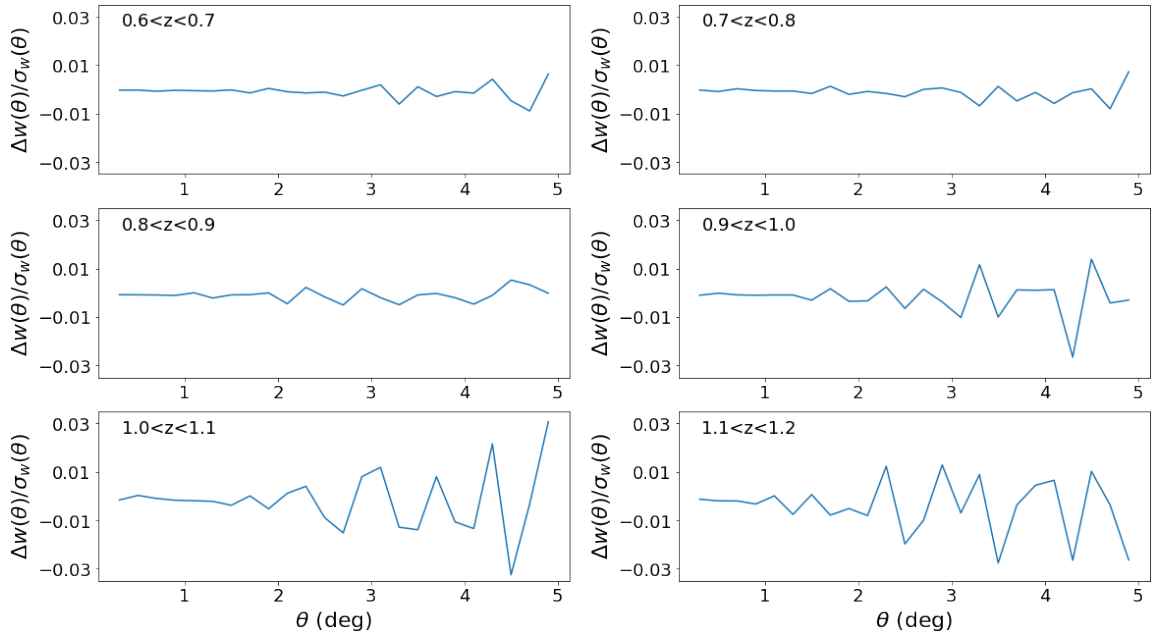

Figure 5.12: Blind $\Delta w(\theta)$ as measured with two independent codes (TreeCorr and CUTE). The difference is normalized by the square root of the diagonal of the CosmoLike covariance.

robustness of the measurement by comparing the results of two independent codes while keeping it blinded (since we only show differences in $w(\theta)$, and not the absolute values). In Fig. 5.12 we show this comparison as a function of θ and as a function of the redshift bin. We divide $\Delta w(\theta)$ by the square root of the diagonal of the CosmoLike covariance, and find that all the points are well within 0.05σ from zero, i.e., both measurements are extremely similar. In fact, the χ^2 between them is 0.06 for 144 degrees of freedom. We conclude that both codes give the same clustering measurements, and the difference between them is at the level of numerical precision.

5.3.2 Forecast for the Combination with 3×2pt

In this section we estimate the increase in precision in the cosmological parameters inferred by 3×2pt when combining with BAO, using a synthetic datavector for 3×2pt and the BAO likelihood of the mean of the COLA mocks. The 3×2pt datavector is obtained from the properties of the Y6 samples (area, number densities and redshift distributions for both lenses and sources), and we generate it using CosmoSIS. Since the BAO likelihood we use comes from the mean of the COLA mocks, we generate the 3×2pt datavector for Mice cosmology so that both likelihoods have the same underlying cosmology. We run chains allowing all the parameters of the 3×2pt analysis to vary: a total of 31, including cosmological parameters and photo-z-related param-

eters for lenses and sources. We run a total of two chains: one for 3×2pt and another one 3×2pt+BAO. The results are shown in Fig. 5.13, in which we show constraints in Ω_m, h and σ_8, marginalizing over all the other parameters. We find that the contours shrink when including the BAO likelihood. The vertical and horizontal dashed lines represent the input parameters (Mice cosmology, see Table 4.2), which are in very good agreement with the results of the chains. For 3×2pt alone we find

$$\begin{cases} \Omega_m = 0.250^{+0.013}_{-0.012}, \\ h = 0.727^{+0.107}_{-0.091}, \\ \sigma_8 = 0.798^{+0.025}_{-0.025}, \end{cases} \tag{5.41}$$

whereas for 3×2pt+BAO the results are

$$\begin{cases} \Omega_m = 0.251^{+0.011}_{-0.010}, \\ h = 0.724^{+0.106}_{-0.081}, \\ \sigma_8 = 0.795^{+0.021}_{-0.021}. \end{cases} \tag{5.42}$$

The Hubble parameter cannot be constrained with the combination of 3×2pt and BAO alone. The constraining power in Ω_m improves by ∼18% when combining with BAO, whereas in σ_8 improves by ∼16%. The constraining power in every other parameter improves as well, e.g., in the linear galaxy bias it increases by 7-10% depending on the redshift bin. The BAO measurement plays an important role in the whole DES picture, since it allows for more precise estimation of cosmological parameters.

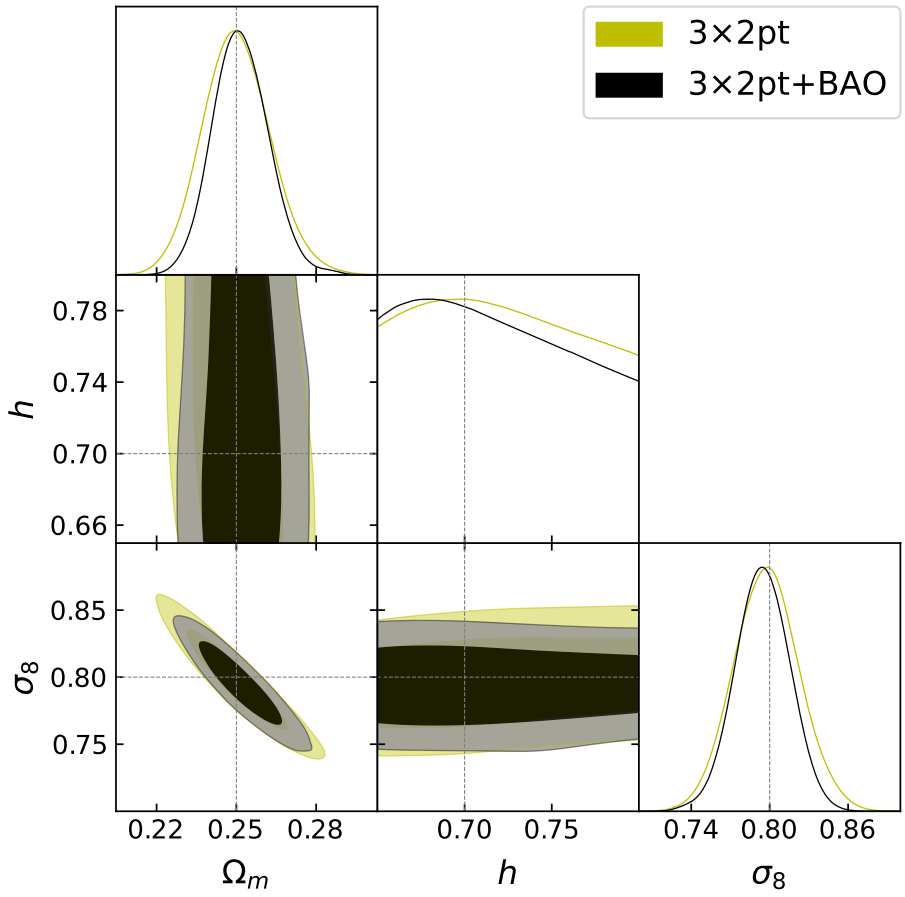

Figure 5.13: Marginalized constraints in Ω_m, h and σ_8 for 3×2pt alone and its combination with BAO. The 3×2pt datavector was generated with CosmoSIS for Mice cosmology, with the same properties as the official Y6 lens and source samples. The BAO likelihood used here is the one obtained when fitting the mean of the Y6 COLA mocks. The vertical and horizontal lines represent the input cosmological parameters.

Chapter 6

HOD Systematics in DESI Y1

Systematics are a major concern in the analyses of large galaxy surveys, but more in particular when we enter the very-high-precision era of Stage-IV surveys, such as the Dark Energy Spectroscopic Instrument (DESI). Systematic errors can arise from a variety of sources, including instrumental effects, calibration uncertainties and astrophysical effects, and they can impact the accuracy and precision of the measured BAO signal. DESI is designed to mitigate systematics through different techniques, including careful instrument calibration, robust data reduction and analysis procedures, and the use of multiple complementary probes of cosmology. For example, DESI uses "spectro-photometric" calibration, which combines spectroscopic and photometric data to improve the accuracy of the flux calibration while including several validation tests to assess the impact of potential systematic effects.

In this chapter we study the impact of the galaxy-halo connection in the BAO measurement, in the context of the DESI Y1 analysis. There are several models to describe the galaxy-halo connection, but we focus on the halo occupation distribution (HOD) approach. HOD systematics arise because the distribution of galaxies within halos can affect the clustering of galaxies on large scales, which is where the BAO feature is detected (at around $\sim$150 Mpc). Typically, HOD models affect more importantly the clustering on small scales. However, since DESI is a very high-precision instrument (it is able to measure the BAO feature with a precision at the sub-percent level), it is necessary to deal with these systematics. DESI analyses several different dark matter tracers, as we already discussed in Chapter 3, but here we focus on luminous red galaxies (LRGs).

6.1 The Galaxy-Halo Connection

The growth and spatial distribution of galaxies are closely related to the growth and spatial distribution of dark matter halos [126]. This is the so-called "galaxy-halo connection", which refers to the full multivariate distribution of properties of halos and the galaxies that form inside them. In the modern theory of structure

formation, dark matter halos are the basic unit into which matter collapses. They can be thought of as gravitationally bounded regions that decoupled from the Hubble flow and collapsed.

The galaxy-halo connection started as a concept together with the framework of cold dark matter models [126]. The study of this connection helps us answering some of the most fundamental questions in cosmology and astrophysics, such as understanding the physics of galaxy formation, probing the properties and spatial distribution of dark matter and inferring cosmological parameters. The galaxy-halo connection heavily relies on the establishment of the ΛCDM model. The high precision measurement of the parameters of this model by experiments as Planck and DES allows robust prediction for the growth of structure and the formation and evolution of dark matter halos [127].

6.1.1 Models of the Galaxy-Halo Connection

The connection between galaxies and their dark matter halos is an active area of research in astrophysics, and there are several models that attempt to describe this relationship. Among these methods we find:

- Halo occupation distribution (HOD) models [128, 129]. In these kind of models, the probability of a galaxy residing in a given halo is described by a statistical distribution function, which depends on the mass of the halo. The HOD model also includes a prescription for the distribution of central and satellite galaxies within a halo.

- Subhalo abundance matching (SHAM) models [130, 131]. In these kind of models, the observed abundance of galaxies is matched to that of subhalos (i.e., smaller halos within a larger halo) as predicted by simulations. SHAM models assume that each galaxy occupies a subhalo with a certain mass, and the relationship between galaxy luminosity and subhalo mass is calibrated using observations.

- Hydrodynamic simulations [132]. These models use simulations that incorporate the effects of gas dynamics, star formation, and feedback from SN and other sources to model the formation and evolution of galaxies within their dark matter halos. Hydrodynamic simulations can provide detailed predictions for the spatial distribution and kinematics of galaxies within their halos.

- Empirical models [133]. These models use empirical relationships between observable properties of galaxies (such as luminosity, color, or stellar mass) and their host halos, which are calibrated using observations of galaxy clustering and other large-scale structure measurements.

The HOD formalism is able to produce galaxy mocks that accurately reproduce the observed clustering, which allows us to run robustness tests of cosmology

pipelines [134, 135]. Also, HOD studies have proven to be successful in characterizing magnitude-limited samples in past galaxy redshift surveys [136, 137]. This, together with their flexibility to constrain cosmology from highly non-linear scales which are inaccessible for standard analytical methods [138, 139], are the reasons why in this chapter we focus on the HOD modeling approach in order to study the impact of the galaxy-halo connection in the measurement of the BAO signal.

6.1.2 The Halo Occupation Distribution (HOD)

Huge scientific advances in our understanding of the galaxy-halo connection have been driven with the use of empirical models that connect observations from large galaxy surveys to the predictions of the properties of dark matter halos in cosmological simulations. We generally assume that the basic properties of the halos are known for a given cosmological model, and they can be predicted using an N-body simulation[1] [126]. For instance, in order to predict clustering statistics one needs to know the abundance of halos, their clustering properties, the radial distribution of matter and the velocity distribution of dark matter within halos. In most of the discussion below we assume these predictions are made with gravity-only N-body simulations.

HODs are a widely used description for the galaxy-halo connection. They describe the bias between galaxies and dark matter by specifying

1. the probability distribution that a halo of mass M contains N galaxies of a particular type, $P(N|M)$.

2. the relative spatial distribution of galaxies and dark matter within halos.

3. the relative velocity distribution of galaxies and dark matter within halos.

This PDF, $P(N|M)$, is usually quantified separately for the central galaxies of the halos (using a Bernoulli distribution) and the satellite galaxies that orbit within them (using a Poisson distribution) [126]. Under these assumptions, the standard HOD is fully characterized by its mean occupation number[2], $\langle N \rangle(M) \equiv \langle N|M \rangle$. However, the HOD can also be a function of properties other than the halo mass, i.e., it can also depend on secondary halo properties, an effect commonly referred to as galaxy assembly bias [126]. If the HOD at a given halo mass M is statistically independent of the halo's large scale environment, then this description for the galaxy bias is essentially complete, i.e., given the HOD and the halo population predicted by our cosmological model, we can compute any galaxy clustering statistic on any physical scale [129]. Therefore, empirical determinations of the HOD for different types of

[1] The reason for this is that the physics of dark matter halos is dominated by the effect of gravity, and no detailed knowledge of other interactions is needed to describe it.

[2] The mean occupation number, or mean HOD, encapsulates the average distribution of a given type of galaxy hosted per halo of a certain mass M, and it depends on the properties of the selected galaxies.

galaxies would summarize everything that the observed galaxy clustering has to say about galaxy formation [126, 129].

The connection between modern halo occupation models and measurements of galaxy clustering is well constrained for a wide range of galaxy samples. The functional form of the HOD for mass- or luminosity-selected galaxies is generally assumed to be similar to that of the dark matter subhalos within their hosts.

6.1.3 HOD Models for LRGs in DESI

Here we describe the different types of HOD models used for the LRG tracer in the context of the DESI Y1 analysis. The fiducial HOD model is the one described in [140], hereafter denoted as Zheng07. In Zheng07, the mean HOD for central galaxies is given by

$$\langle N_{\rm cen}^{\rm LRG}\rangle(M) = \frac{1}{2}{\rm erfc}\left[\frac{\log_{10}(M_{\rm cut}/M)}{\sqrt{2}\sigma}\right], \tag{6.1}$$

whereas the mean HOD for satellite galaxies is

$$\langle N_{\rm sat}^{\rm LRG}\rangle(M) = \left(\frac{M-\kappa M_{\rm cut}}{M_1}\right)^{\alpha}\langle N_{\rm cen}^{\rm LRG}\rangle(M). \tag{6.2}$$

$M_{\rm cut}$ and $\kappa M_{\rm cut}$ are the minimum halo mass to host a central or a satellite galaxy, respectively; M_1 is the typical halo mass that hosts one satellite galaxy; σ is the steepness of the transition from 0 to 1 in the number of central galaxies; and α is the power law index on the number of satellite galaxies. The actual number of central galaxies per halo is drawn from a Bernoulli distribution with mean $\langle N_{\rm cen}^{\rm LRG}\rangle(M)$, whereas the actual number of satellite galaxies is drawn from a Poisson distribution with mean $\langle N_{\rm sat}^{\rm LRG}\rangle(M)$. Central galaxies are assigned to the center of mass of the largest sub-halo, with the velocity vector also set to that of the center of mass of the largest sub-halo. Satellite galaxies are assigned to particles of the halo with equal weights. Since a halo contains necessarily either zero or one central galaxies, the probability distribution $P(N_{\rm cen}^{\rm LRG}|\langle N_{\rm cen}^{\rm LRG}\rangle(M))$ is a nearest-integer distribution with

$$P(1) = 1 - P(0) = \langle N_{\rm cen}^{\rm LRG}\rangle(M). \tag{6.3}$$

Besides Zheng07, we also have two alternative HOD models: "baseline" and "extended" [141]. These are the ones we use in this thesis to estimate the contribution of HOD systematics to the BAO measurement. The baseline model is given by Zheng07 but including velocity bias ($\alpha_{\rm cen}$ and $\alpha_{\rm sat}$ for central and satellite galaxies, respectively) and a completeness factor. The extended model is simply the baseline model but also including the environment-based assembly bias ($B_{\rm cen}$ and $B_{\rm sat}$ for central and satellite galaxies, respectively), and a satellite radial profile bias parameter. Both alternative models are based on the fiducial one, Zheng07, and they are further described in [141]. These models are sufficiently varied for the study of HOD

systematics in the context of the DESI Y1 analysis for LRGs, see Yuan et al. in prep. For other tracers such as ELGs, the variety of HOD models considered is richer, as described in García-Quintero et al. in prep.

6.2 The AbacusSummit Simulations

The simulations we use in DESI in order to study the effect of the HOD modeling in the BAO measurement are the AbacusSummit simulations [142, 143]. The AbacusSummit project aims to generate a large suite of simulations to study the large-scale structure of the Universe and its evolution over time. These simulations are designed to be high-resolution and large-volume, allowing for detailed studies of cosmological parameters, as well as the statistical properties of the cosmic web, such as the distribution of matter, dark matter halos, and voids. They are run on state-of-the-art supercomputers and utilize advanced computational techniques to accurately model the evolution of the universe from the early Universe to the present day.

The AbacusSummit simulations we use in this chapter are cubic boxes of 2 Gpc/h size with about $\sim$3 million galaxies each, i.e., the number density of these boxes is, approximately,

$$n_{\rm gal} = \frac{N_{\rm gal}}{l_{\rm box}^3} = \frac{3\times10^6 \text{ gal}}{(2\times10^3 \text{ Mpc}/h)^3} = 3.75\times10^{-4} h^3/\text{Mpc}^3, \tag{6.4}$$

which has the same value as that of the Y1 LRG catalog. Cubic box mocks simulate a cubic volume of the Universe without explicit modeling of a specific survey footprint, whereas sky-cut simulations, such as the COLA mocks we used in Chapters 4 and 5, model a specific survey footprint or observational mask in the sky [144]. These boxes have periodic boundary conditions, where objects leaving one edge of the box re-enter from the opposite edge. Cubic box mocks are often used for testing and calibrating data analysis methods, studying the impact of astrophysical and cosmological parameters, and performing forecast studies. They are valuable tools for validating and optimizing data analysis techniques, assessing systematic biases, and estimating uncertainties in cosmological measurements, which is our main goal. The cosmology of the simulations, which hereafter we refer to as "DESI fiducial cosmology", is given by ΛCDM with Planck-2018 [14] cosmological parameters, explicitly $\Omega_m = 0.3153$, $h = 0.6736$, $\Omega_b = 0.0493$, $\sigma_8 = 0.8111$ and $n_s = 0.9649$ (the same ones displayed in Table 1.1).

6.2.1 AbacusHOD

The HOD model of the AbacusSummit simulations is specified with the AbacusHOD model. AbacusHOD is designed to generate mock galaxy catalogs that reproduce the observed clustering properties of galaxies in large-scale structure surveys. The framework is based on an extended multi-tracer approach that combines

Mock label	Model
HOD A0	Baseline (best-fit parameters) .....
HOD A1, A2, A3	Baseline (3σ variations from best fit) .
HOD B0	Extended (best-fit parameters) .....
HOD B1, B2, B3	Extended (3σ variations from best fit)

Table 6.1: Different HOD models considered for the estimation of HOD systematics in the context of the DESI Y1 analysis. We have two alternative HOD models: baseline and extended. The best-fit baseline and extended HODs are labeled as A0 and B0, respectively. HODs A1, A2 and A3 are 3σ random variations of the best-fit baseline model, whereas HODs B1, B2, B3 are 3σ random variations of the best-fit extended model.

multiple tracers of the underlying dark matter distribution, such as galaxies of different luminosities or colors, to model the HOD in a more realistic and flexible way. It is designed specifically for next-generations large-galaxy surveys, such as DESI, and takes advantage of the volume and precision offered by the AbacusSummit simulations. The AbacusHOD code is publicly available as a part of abacusutils[3].

A total of 25 AbacusSummit cubic simulations were generated for each HOD. As we already mentioned, we have two different alternative HOD models for LRGs, namely the baseline and the extended models. The fits are done by optimizing them against the Fuji v3.0[4] $\xi(r_\parallel, r_\perp)$ measurement of the LRG main sample between 0.1-30 Mpc/h, using the mean number density within that redshift bin. For redshift bins, $0.6 < z < 0.8$ was used for the $z = 0.8$ snapshot[5], which is the one we consider here. These fits to the Fuji v3.0 data are fully described in Yuan et al. in prep.

The best-fit parameters for both HOD models are then used to create the HOD mocks. In Table 6.1 we display the different sets of HOD mocks used for LRGs, together with their respective labels. We denote the best-fit HOD baseline model as HOD A0, and the best-fit extended one as HOD B0. HODs A1, A2 and A3 are random variations within 3σ of the best-fit parameters of HOD A0, whereas HODs B1, B2 and B3 are analogous variations but of HOD B0. Since we have 8 HOD models, we end up with a total of 25×8 AbacusSummit simulations; however, each set of 25, corresponding to a different HOD model, is analyzed independently. The creation of the mock catalogs is further described in Yuan et al. in prep.

[3] https://github.com/abacusorg/abacusutils

[4] The data corresponds to DESI's One-Percent Survey, which was observed as the third and final phase of its Survey Validation in April and May of 2021.

[5] The simulation output is organized into discrete redshift snapshots, and for the results shown in this thesis, the $z = 0.8$ was the one considered.

6.2.2 Reconstruction

We already introduced the concept of reconstruction in section 2.5.3.1. Density field reconstruction is a technique used in cosmology in order to infer the initial density fluctuations of the Universe from its observed large-scale structure [145, 146]. It is based on the idea that the distribution of galaxies is not only influenced by the initial density fluctuations present at the time of the Big Bang, but also by the subsequent evolution of the Universe, including the growth of cosmic structures due to gravity.

Density field reconstruction is widely used in the context of BAO studies in spectroscopic galaxy surveys, since it allows for a better detection of the BAO feature because of a sharpening of the BAO peak [146]. The reconstruction of our AbacusSummit simulations is carried out using the Python package Pyrecon, which is part of Cosmodesi[6]. Cosmodesi is the main scientific repository of DESI, and encapsulates several different software pipelines, such as Pycorr and Pypower, which are used to compute the clustering measurements in configuration and Fourier spaces, respectively, in both simulations and data. Pyrecon has the option to run the reconstruction with several different algorithms, such as MultiGrid, based on Martin J. White's code[7], and iterative fast Fourier transform (IFFT), based on Julian E. Bautista's code[8]. For the HOD systematics studies, MultiGrid was the approach chosen, and it was run with a number of grid points of $n_{\rm mesh} = 512$ and a smoothing scale of 10 Mpc/h. These settings were optimized by the optimal reconstruction analysis team, Paillas et al. in prep.

6.2.3 Clustering Measurements

The clustering measurements of the AbacusSummit simulations are computed using the Python packages Pycorr and Pypower for configuration and Fourier space, respectively, being these two codes part of Cosmodesi. Originally, the AbacusSummit cubic boxes are divided into 64 smaller boxes of size 500 Mpc/h, and so we have to concatenate them all to create a full mock. Also, by default, the positions in the files are in real space, which means we have to add the effect of the peculiar velocities of the galaxies along the LOS to transform then into redshift-space positions. This is effectively done by shifting the z-coordinate (hereafter, z_c) as

$$z_c \to z_c + \frac{v_{z_c}}{a(0.8)H(0.8)}, \tag{6.5}$$

where v_{z_c} is the velocity of each galaxy along the z-axis. The scale factor and the Hubble function are evaluated at redshift 0.8 since, as we mentioned, we use the $z = 0.8$ snapshot. Because of the previous transformation, some galaxies might

[6] https://github.com/cosmodesi
[7] https://github.com/martinjameswhite/recon_code
[8] https://github.com/julianbautista/eboss_clustering/blob/master/python/recon.py

Setting	Configuration space	Fourier space
Edges	$s \in [0, 200]$ Mpc/h, $\mu \in [-1, 1]$	$k \in [0, \pi n_{\rm mesh}/l_{\rm box}]$ h/Mpc
Bin size	$\Delta s = 1$ Mpc/h, $\Delta\mu = 0.0083$	$\Delta k = 0.001$ h/Mpc

Table 6.2: Default settings for Pycorr (i.e., configuration space) and Pypower (i.e., Fourier space) when computing the clustering measurements of the AbacusSummit mocks. The upper limit of k is given by the Nyquist frequency, eq. (6.6). By default, the value of $n_{\rm mesh}$ we use is 512, which means $k_{\rm max} = 0.804$ h/Mpc.

end up outside of the cubic box of size 2 Gpc/h. However, because of the periodic boundary conditions of our simulations, we simply move them back to the box. Once we have created the full redshift-space catalogs, we run Pycorr and Pypower on them. In Table 6.2 we show the default settings used when computing the clustering measurements. For configuration space, by default we choose s to be within 0 and 200 Mpc/h, with a total of 201 bins ($\Delta s = 1$ Mpc/h); and μ to be between -1 and 1, with 241 bins ($\Delta\mu = 0.0083$). For Fourier space, we choose k to be within 0 and the Nyquist frequency, which is defined as

$$k_{\rm Nyquist} = \frac{2\pi}{l_{\rm box}} \frac{n_{\rm mesh}}{2}, \tag{6.6}$$

where $l_{\rm box}$ is the size of the cubic box ($l_{\rm box} = 2$ Gpc/h) and $n_{\rm mesh}$ is the number of mesh points or grid points used in to discretize the continuous field of matter density fluctuations (i.e., the volume of the box is discretized in $n^3_{\rm mesh}$ smaller regions). The Nyquist frequency represents the highest frequency that can be accurately measured without introducing aliasing effects, which distort the signal. It corresponds to the wavenumber that is half the size of the survey or simulation volume. The default value we use for $n_{\rm mesh}$ is 512.

In Fig. 6.1 we show the configuration-space clustering measurements, $\xi_\ell(s)$, for the alternative HOD models A0 and B0. For both HOD models we show the mean of the correlation functions for the 25 AbacusSummit simulations, for pre and post-reconstruction catalogs. We find that both models give rise to a very similar clustering signal for the three multipoles. We also find that reconstruction sharpens the BAO peak, as expected, and that the $\ell = 4$ multipole is very noisy compared to the other two, which is the main reason why we exclude it from the BAO fits. On the other hand, in Fig. 6.2 we show the Fourier-space clustering measurements, $P_\ell(k)$, also for the alternative HOD models A0 and B0. The results are consistent with those we just discussed in the case of configuration space.

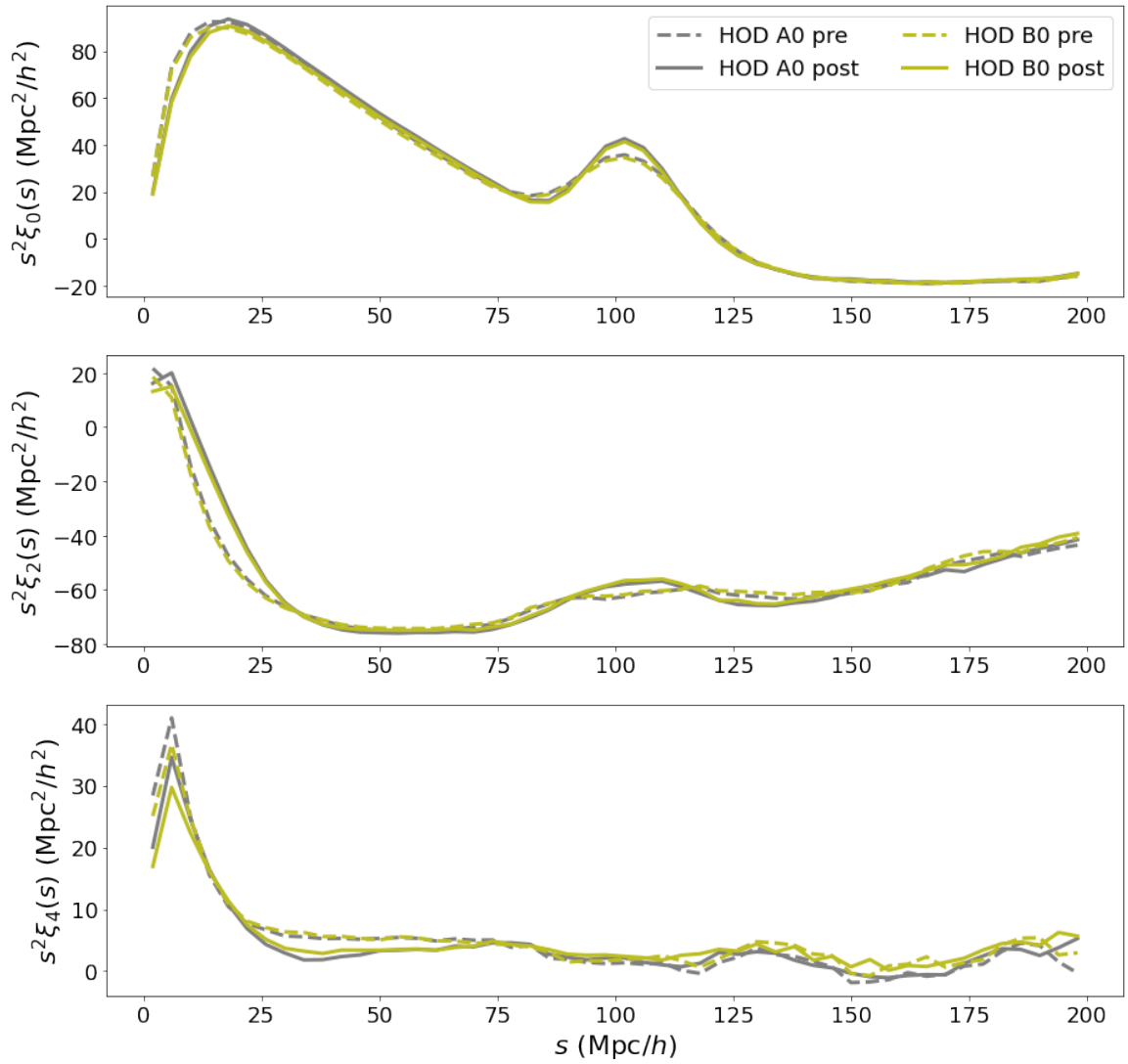

Figure 6.1: Multipoles of the correlation function (from top to bottom, $\ell = 0,\ 2,\ 4$) for HOD models A0 and B0, as measured with Pycorr. The plots shown here correspond to the mean of the 25 AbacusSummit simulations. Pre and post-reconstruction results are shown (solid and dashed lines, respectively). We find that both HOD models give rise to a very similar clustering signal, for both pre and post-reconstruction measurements. We also find that, as expected, reconstruction sharpens the BAO peak.

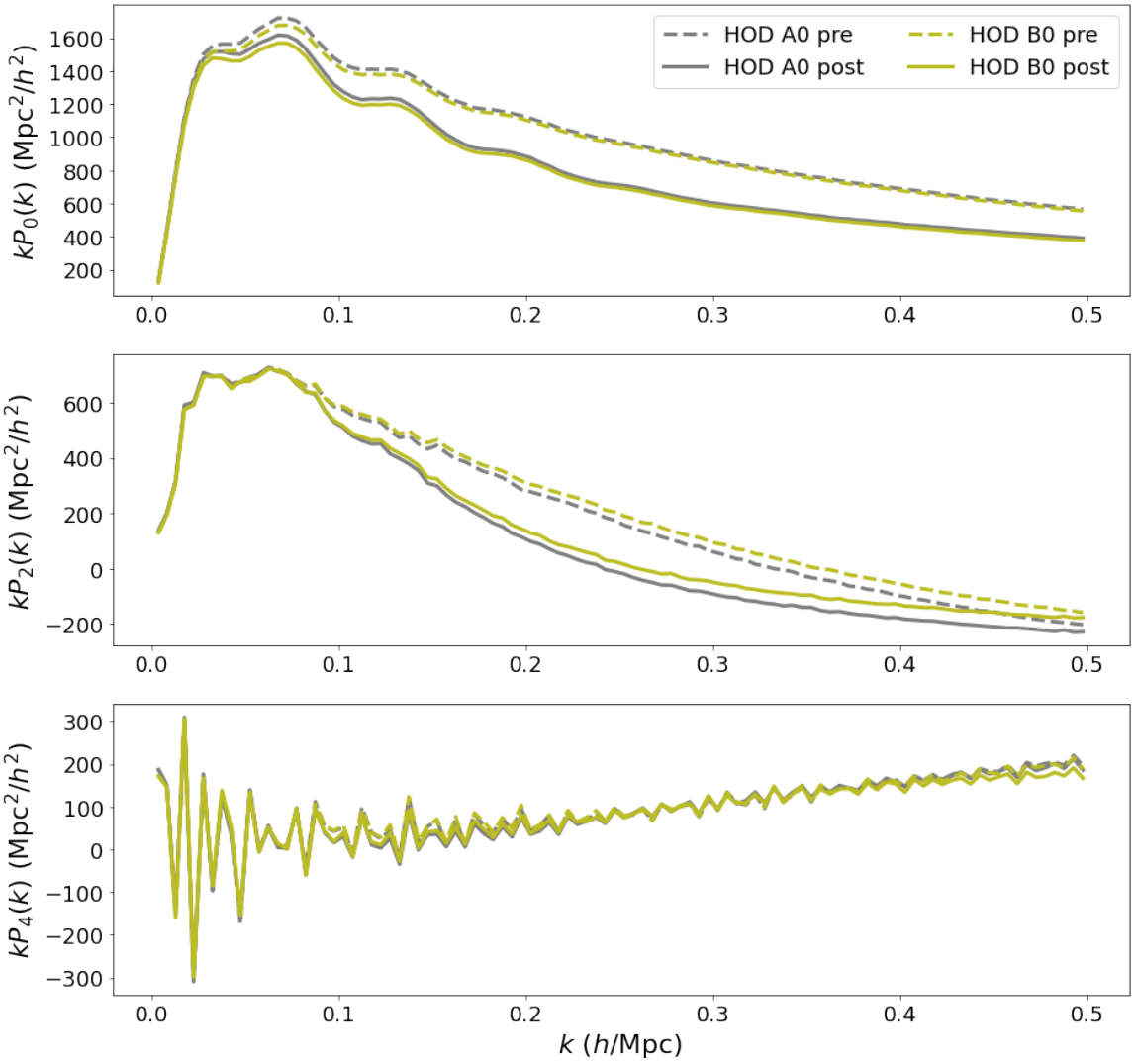

Figure 6.2: Multipoles of the power spectrum (from top to bottom, $\ell = 0,\ 2,\ 4$) for HOD models A0 and B0, as measured with Pypower. The plots shown here correspond to the mean of the 25 AbacusSummit simulations. Pre and post-reconstruction results are shown (solid and dashed lines, respectively). We find that both HOD models give rise to a very similar clustering signal, for both pre and post-reconstruction measurements.

6.3 BAO Measurements on Simulations

The BAO fits are performed in a similar way to the methodology we described in Chapter 4. Here we also use a template-based method, but in this case we have two α parameters, one along the LOS ($\alpha_\parallel$) and another one across the LOS ($\alpha_\perp$), as we already explained in section 2.5.3.2. The fiducial code for performing BAO fits in DESI is Barry [147], which is publicly available at its code repository[9]. Barry is a modular code which contains datasets, model fitting tools, and model implementations incorporating different descriptions of non-linear physics and algorithms for isolating the BAO feature. These were evaluated for bias, correlation and fitting strength using two-point correlation functions and power spectra from simulations developed for the SDSS Data Release 12 [147]. Barry has several different models which can be used to run the BAO fits, and all of them can be compactly written as

$$P_{\ell,\mathrm{model}}(k,\mathbf{\Theta}) = \int_{-1}^{1} d\mu P_{\mathrm{ref}}\left[k'(k,\mu),\mu'(\mu)\right]\mathcal{L}_\ell(\mu) + \mathcal{D}(k), \tag{6.7}$$

where $\mathbf{\Theta}$ denotes the set of free parameters of the model. The reference power spectrum, $P_{\mathrm{ref}}(k,\mu)$, can be expressed as

$$P_{\mathrm{ref}}(k,\mu) = \left[\mathcal{A}(k,\mu) + \mathcal{B}(k,\mu)P_m^{\mathrm{nw}}(k)\right]\left[1 + \left(\frac{P_m^{\mathrm{lin}}(k)}{P_m^{\mathrm{nw}}(k)} - 1\right)\mathcal{C}^2(k,\mu)\right], \tag{6.8}$$

where $k'(k,\mu)$ and $\mu'(\mu)$ are given by eqs. (2.137) and (2.138), respectively (which we already deduced in Chapter 2). The factor $\mathcal{B}(k,\mu)$ that appears in the expression of $P_{\mathrm{ref}}(k,\mu)$ is some function that applies to the no-wiggle power spectrum, such as linear bias correction, Kaiser factor or Fingers-of-God term [147]; the function $\mathcal{C}^2(k,\mu)$ is the propagator that determines the evolution of the BAO feature; and $\mathcal{A}(k,\mu)$ and $\mathcal{D}(k)$ are extra additive terms, typically polynomial functions (the broadband-term parameters). By setting

$$\begin{cases} \mathcal{A}(k,\mu) = 0, \\ \mathcal{B}(k,\mu) = \left(b + f\mu^2\right)^2, \\ \mathcal{C}^2(k,\mu) = \exp\left(-k^2\Sigma_{\mathrm{tot}}^2(\mu)\right), \end{cases} \tag{6.9}$$

the reference power spectrum of eq. (6.8) becomes exactly the same expression we used in Chapter 4, eq. (4.29).

Out of all the different models that Barry incorporates, the one we use here is the one presented in [45], hereafter denoted as Beutler17. This model includes the Fingers-of-God term to account for non-linear redshift space distortions, which are

[9] https://github.com/Samreay/Barry

also somewhat accounted for by the polynomial terms (the broadband-term parameters). This model is implemented by setting

$$\begin{cases} \mathcal{B}(k,\mu) = b^2 \left(1 + \frac{1}{2} k \Sigma_s^2\right)^{-2}, \\ C^2(k,\mu) = \exp\left(-k^2 \Sigma_{\rm tot}^2(\mu)\right), \end{cases} \tag{6.10}$$

where b is the linear galaxy bias, Σ_s^2 is the damping scale for the Fingers-of-God and $\Sigma_{\rm tot}(\mu)$ allows for a smooth dampening of the BAO feature at small scales (and is given by eq. (5.10)). For the BEUTLER17 model we also set

$$\mathcal{A}(k,\mu) = \sum_{i=1}^{n} a_i k^{2-i}. \tag{6.11}$$

For the fits in configuration space, the power spectrum of our model is Hankel-transformed into the two-point correlation function.

6.3.1 Default Settings

The default settings for BARRY are displayed in Table 6.3. The templates we use for all the fits in this chapter are computed for the same cosmological parameters as the underlying cosmology of the mocks, i.e., the DESI fiducial cosmology (see section 6.2). Therefore, we expect the best-fit α to be consistent with 1, in both the parallel and the perpendicular directions. For both configuration and Fourier spaces, we fit the monopole and the hexadecapole ($\ell = 0$ and $\ell = 2$, respectively). The covariance matrices used for the fits include covariances between s or k bins, and also cross-covariances between multipoles. For the computation of configuration-space covariances, we use the RASCALC[10] code [148, 149], whereas for Fourier space, we use COVAPT[11] [150]. Several covariance matrices were generated: a total of 2 for the case of configuration space, one for HODs A0, A1, A2, A3 (corresponding to the baseline model) and another one for HODs B0, B1, B2, B3 (corresponding to the extended model); and a total of 8 for the case of Fourier space, one for each HOD model independently. The reason for this is that configuration-space covariances are much more time-consuming than Fourier-space ones. The same set of covariance matrices were also computed for post-reconstruction measurements, for configuration and Fourier spaces. Regarding the number of BAO chains run, we have a total of 8 HOD models and a set of 25 ABACUSSUMMIT simulations for each model. We also fit the mean of the mocks, which means we effectively have 26 realizations for each HOD model. Since we run the BAO fits in configuration and Fourier spaces, for pre and post-reconstruction cases, we end up with a total of $8 \times 26 \times 2 \times 2 = 832$ chains.

[10] https://github.com/oliverphilcox/RascalC
[11] https://github.com/JayWadekar/CovaPT

Setting	Configuration space - $\xi_\ell(s)$	Fourier space - $P_\ell(k)$
Edges	$s \in [50, 150]$ Mpc/h	$k \in [0.02, 0.3]$ h/Mpc
Bin size	$\Delta s = 4$ Mpc/h	$\Delta k = 0.005$ h/Mpc
Multipoles	$\ell = 0,\ 2$	$\ell = 0,\ 2$
Template	BEUTLER17	BEUTLER17
Cosmology	DESI fiducial cosmology	DESI fiducial cosmology
# of bb-term par.	4 ($i = -1, 0, 1, 2$)	6 ($i = -1, 0, 1, 2, 3, 4$)
Non-linear BAO .	Gaussian prior $\Sigma_{\parallel,\perp}$, Σ_s	Gaussian prior $\Sigma_{\parallel,\perp}$, Σ_s

Table 6.3: Default settings to perform the BAO fits in configuration and Fourier spaces. The scales for the fits are well within the limits in which we computed the clustering measurements, see Table 6.2; however, the original clustering measurements were re-binned (for both configuration and Fourier spaces). The Gaussian priors on $\Sigma_{\parallel,\perp}$ and Σ_s have 2 Mpc/h of width. The center of the Gaussian prior is set to 9.6 and 5.4 Mpc/h for $\Sigma_\parallel$ (pre and post-reconstruction fits, respectively); 5.0 and 1.8 Mpc/h for $\Sigma_\perp$ (pre and post-reconstruction fits, respectively); and 0 Mpc/h for Σ_s. These values were optimized by the BAO systematics group, Howlett et al. in prep.

6.3.2 BAO Fit Results

In Table 6.4 we show the BAO fit results for the different alternative HOD models considered (those listed in Table 6.1). All the quantities displayed in this table are averages over the 25 ABACUSSUMMIT realizations. The upper table shows the configuration-space results, whereas the lower one shows the Fourier-space ones. We find that there is a good agreement between configuration and Fourier-space results. All the average α are consistent with 1 within the error-bars, for both parallel and perpendicular directions. However, pre-reconstruction $\alpha_\parallel$ is slightly biased with respect to 1 compared to the post-reconstruction results. This is due to the non-linear evolution of the BAO peak, which is mitigated with the density field reconstruction technique. The error-bars in both $\alpha_\parallel$ and $\alpha_\perp$ are reduced after reconstruction, and the same happens to the standard deviations. Reconstruction effectively sharpens the BAO peak, making it easier to measure, while providing a more precise distance measurement as well. For both configuration and Fourier-space results, we find that the correlation between parallel and perpendicular decreases (in absolute value) after the reconstruction. Finally, we find that the average χ^2 are, in general, very similar to the degrees of freedom for all the different HODs, which further ensures the high quality of the results.

In Fig. 6.3 we show the scatter plots for the best-fit α of HOD A0 versus HOD B0, as an example. We include $\alpha_\parallel$ (left) and $\alpha_\perp$ (right), for configuration (top) and Fourier (bottom) spaces. Each point in these scatter plots represents an individual mock realization, out of the total of 25. Blue points indicate best-fit α for pre-

case	$\langle\alpha_\parallel\rangle$	std($\alpha_\parallel$)	$\langle\sigma_\parallel\rangle$	$\langle\alpha_\perp\rangle$	std($\alpha_\perp$)	$\langle\sigma_\perp\rangle$	$\langle\rho_{\parallel,\perp}\rangle$	$\langle\chi^2\rangle$/dof
Configuration space								
HOD A0 pre	1.0124	0.0185	0.0220	1.0001	0.0088	0.0099	-0.45	37.3/37
HOD A1 pre	1.0118	0.0183	0.0222	1.0004	0.0084	0.0099	-0.44	36.3/37
HOD A2 pre	1.0121	0.0187	0.0228	1.0000	0.0085	0.0103	-0.45	36.5/37
HOD A3 pre	1.0117	0.0182	0.0226	1.0010	0.0079	0.0102	-0.44	36.5/37
HOD B0 pre	1.0134	0.0180	0.0224	1.0000	0.0080	0.0101	-0.45	37.0/37
HOD B1 pre	1.0121	0.0189	0.0220	1.0002	0.0084	0.0098	-0.44	37.7/37
HOD B2 pre	1.0137	0.0184	0.0223	0.9995	0.0084	0.0100	-0.45	37.4/37
HOD B3 pre	1.0130	0.0181	0.0226	0.9996	0.0086	0.0101	-0.44	37.0/37
HOD A0 post	0.9991	0.0118	0.0100	0.9981	0.0052	0.0058	-0.42	35.5/37
HOD A1 post	0.9984	0.0123	0.0100	0.9982	0.0051	0.0058	-0.42	36.7/37
HOD A2 post	0.9982	0.0121	0.0102	0.9982	0.0051	0.0060	-0.43	33.7/37
HOD A3 post	0.9986	0.0118	0.0102	0.9983	0.0050	0.0059	-0.43	34.5/37
HOD B0 post	0.9992	0.0113	0.0101	0.9989	0.0045	0.0058	-0.42	38.7/37
HOD B1 post	0.9978	0.0112	0.0101	0.9992	0.0052	0.0057	-0.43	38.5/37
HOD B2 post	0.9999	0.0110	0.0101	0.9987	0.0043	0.0058	-0.42	38.5/37
HOD B3 post	0.9987	0.0111	0.0100	0.9990	0.0050	0.0058	-0.42	39.6/37
Fourier space								
HOD A0 pre	1.0104	0.0186	0.0198	1.0006	0.0083	0.0092	-0.43	94.6/93
HOD A1 pre	1.0101	0.0184	0.0199	1.0009	0.0081	0.0091	-0.44	94.4/93
HOD A2 pre	1.0099	0.0186	0.0198	1.0005	0.0083	0.0092	-0.43	94.6/93
HOD A3 pre	1.0099	0.0187	0.0200	1.0013	0.0078	0.0091	-0.43	94.8/93
HOD B0 pre	1.0110	0.0181	0.0202	1.0007	0.0079	0.0093	-0.44	95.5/93
HOD B1 pre	1.0095	0.0195	0.0203	1.0009	0.0083	0.0094	-0.43	94.6/93
HOD B2 pre	1.0113	0.0184	0.0200	1.0003	0.0084	0.0092	-0.44	94.4/93
HOD B3 pre	1.0105	0.0179	0.0201	1.0005	0.0083	0.0092	-0.43	95.6/93
HOD A0 post	0.9985	0.0118	0.0096	0.9986	0.0053	0.0056	-0.42	93.5/93
HOD A1 post	0.9982	0.0121	0.0095	0.9987	0.0051	0.0056	-0.42	93.8/93
HOD A2 post	0.9979	0.0119	0.0096	0.9986	0.0051	0.0056	-0.42	94.1/93
HOD A3 post	0.9985	0.0116	0.0097	0.9988	0.0049	0.0056	-0.42	93.0/93
HOD B0 post	0.9987	0.0114	0.0097	0.9994	0.0049	0.0056	-0.43	93.9/93
HOD B1 post	0.9974	0.0111	0.0096	0.9997	0.0053	0.0057	-0.43	94.0/93
HOD B2 post	0.9996	0.0113	0.0095	0.9993	0.0046	0.0056	-0.42	93.6/93
HOD B3 post	0.9985	0.0109	0.0096	0.9995	0.0049	0.0056	-0.42	94.8/93

Table 6.4: BAO fit results for the 25 AbacusSummit simulations, for the alternative HOD models of Table 6.1 (pre and post-reconstruction). The average α, its standard deviation and its average error are shown for both the parallel and the perpendicular components. The average correlation between parallel and perpendicular and the average χ^2 divided by the degrees of freedom of the fits are also shown. **Top table:** configuration space. **Bottom table:** Fourier space.

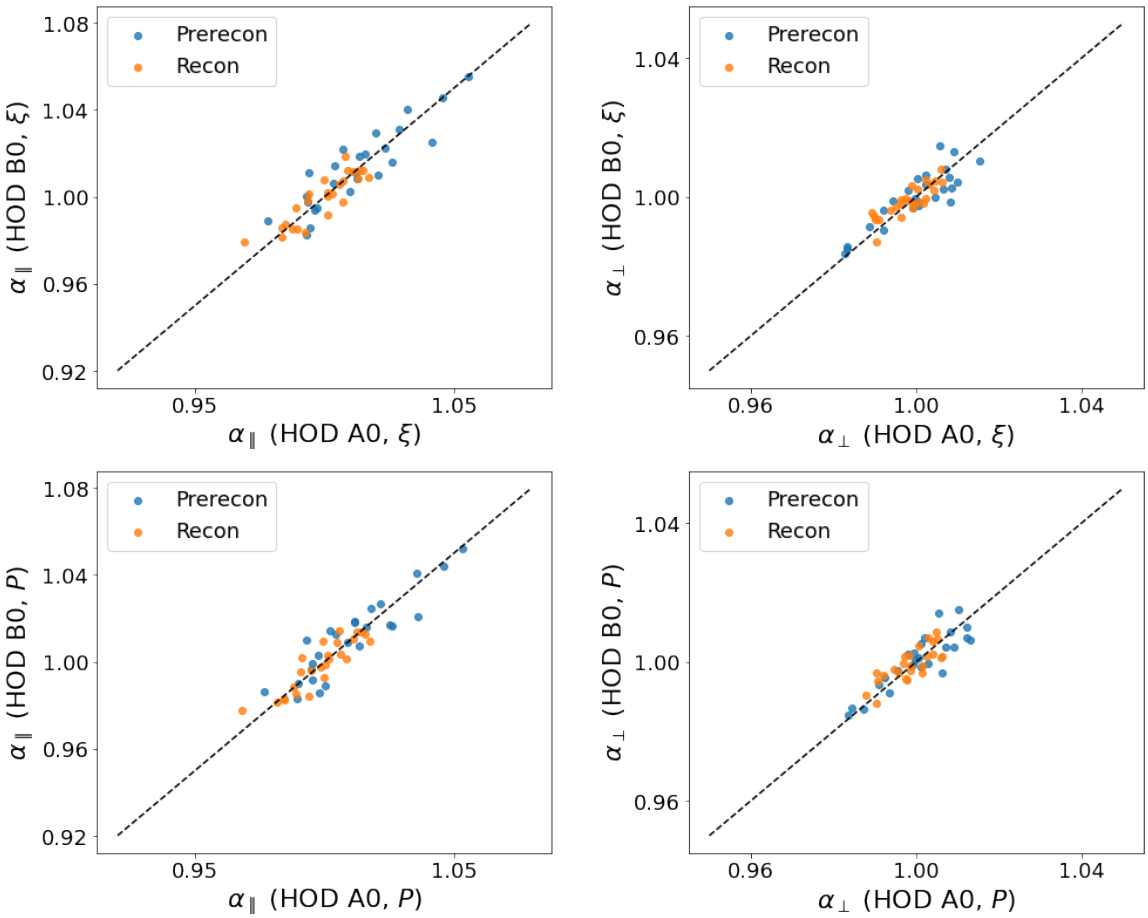

Figure 6.3: Scatter plots of the best-fit α of HOD A0 versus HOD B0. **Top-left panel:** $\alpha_{\parallel}$ configuration space. **Top-right panel:** $\alpha_{\perp}$ configuration space. **Bottom-left panel:** $\alpha_{\parallel}$ Fourier space. **Bottom-right panel:** $\alpha_{\perp}$ Fourier space. Blue points represent pre-reconstruction measurements, whereas orange ones represent post-reconstruction.

reconstruction, whereas orange points are for post-reconstruction. We find that the individual realizations scatter around the diagonal for all the cases shown in Fig. 6.3. However, for post-reconstruction we find that the scatter gets reduced significantly, particularly in the case of $\alpha_{\parallel}$. This is consistent with the results displayed in Table. 6.4. We also find that the results in configuration space are consistent with those of Fourier space, reinforcing the robustness of the measurements.

6.3.3 Error Budget for HOD Systematics

In this section we estimate the error budget related to HOD systematics for the DESI Y1 analysis. For any pair of HOD models i and j, we define $\sigma^{i,j}_{\rm sys}$ as

$$\sigma^{i,j}_{\rm sys} = \max\left(|\langle\Delta\alpha_{i,j}\rangle|, {\rm std}(\Delta\alpha_{i,j})/\sqrt{N_{\rm mocks}}\right), \tag{6.12}$$

where $\Delta\alpha_{i,j}$ is an array of length $N_{\rm mocks} = 25$ representing the difference between the best-fit α of HOD models i and j. The average and the standard deviation in the previous expression are computed over the 25 mock realizations. In Fig. 6.4 we show the heat-maps of $\sigma^{i,j}_{\rm sys}$ as computed for every pair of HOD models considered. We separate the heat-maps into parallel and perpendicular components, but in a single heat-map we include pre-reconstruction (upper-triangle) and post-reconstruction (lower-triangle) results. We include $\alpha_\parallel$ (left) and $\alpha_\perp$ (right), for configuration (top) and Fourier (bottom) spaces. We find that the heat-maps for configuration and Fourier spaces are consistent. The systematics error budget, $\sigma_{\rm sys}$, can be estimated from these heat-maps as the maximum value of $\sigma^{i,j}_{\rm sys}$, i.e.,

$$\sigma_{\rm sys} = \max_{i,j}(\sigma^{i,j}_{\rm sys}). \tag{6.13}$$

For post-reconstruction, we find that the pair of HODs with the largest $\sigma_{{\rm sys},\parallel}$ is (HOD B1, HOD B2), for both configuration and Fourier spaces, with a value of

$$\sigma_{{\rm sys},\parallel} = 0.0022. \tag{6.14}$$

On the other hand, the pair of HODs with the largest $\sigma_{{\rm sys},\perp}$ is (HOD A0, HOD B1), with

$$\sigma_{{\rm sys},\perp} = 0.0011. \tag{6.15}$$

Since both $\alpha_\parallel$ and $\alpha_\perp$ are very close to 1, these results represent a $\sim$0.2% and $\sim$0.1% effect, respectively. Taking a look at the post-reconstruction results displayed in Table 6.4, we expect a statistical uncertainty of about 0.0120 in $\alpha_\parallel$, and of about 0.0050 in $\alpha_\perp$ for the DESI Y1 LRG BAO measurements. Therefore, the contribution of the HOD systematics to the total error-bar of the measurement is

$$\sigma_{{\rm sys},\parallel} \approx \frac{1}{5}\sigma_{{\rm stat},\parallel}, \quad \sigma_{{\rm sys},\perp} \approx \frac{1}{5}\sigma_{{\rm stat},\perp}, \tag{6.16}$$

i.e., of about $0.2\sigma_{\rm stat}$ for both parallel and perpendicular components. The contribution of HOD systematics is relatively large compared to the statistical error; however, we know these error budgets we just obtained are overestimated. The reason is that they have a contribution coming from stochasticity,

$$\sigma^2_{\rm sys} = \sigma^2_{\rm HOD} + \sigma^2_{\rm stochastic}, \tag{6.17}$$

i.e., part of the $\sigma_{\rm sys}$ we just estimated is not due to the HOD modeling itself ($\sigma_{\rm HOD}$ contribution), but rather to the stochasticity when assigning galaxies into halos ($\sigma_{\rm stochastic}$ contribution). Therefore, the $0.2\sigma_{\rm stat}$ obtained in this manuscript is not the final estimation of the HOD systematics uncertainty, but rather an upper limit.

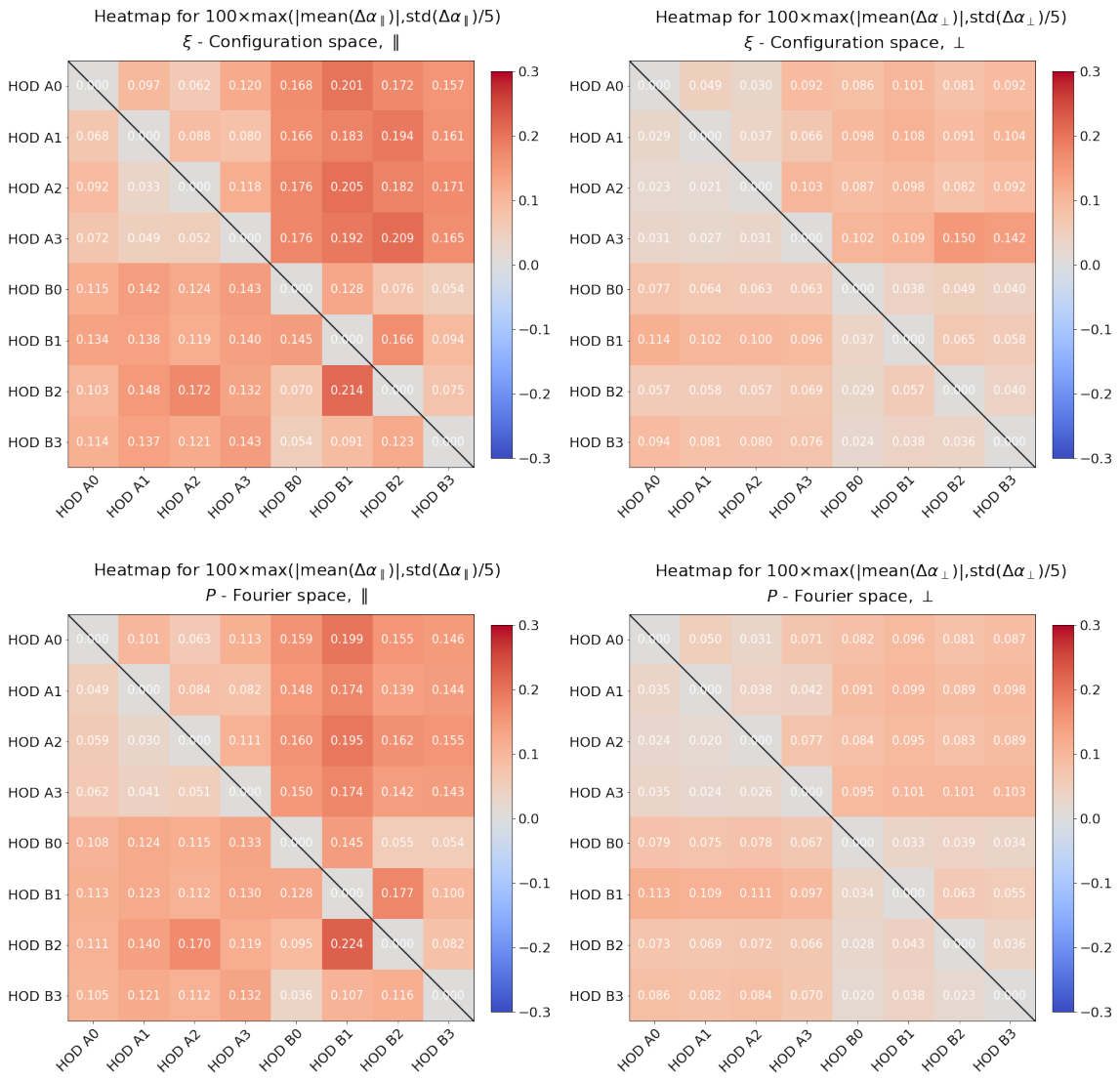

Figure 6.4: Heat-maps showing $100 \times \sigma^{i,j}_{\rm sys}$, as defined by eq. (6.12), for the different HOD models considered. Upper-triangle regions represents pre-reconstruction results, whereas lower-triangle regions represent post-reconstruction. **Top-left panel:** Configuration-space $\alpha_{\parallel}$. **Top-right panel:** Configuration-space $\alpha_{\perp}$. **Bottom-left panel:** Fourier-space $\alpha_{\parallel}$. **Bottom-right panel:** Fourier-space $\alpha_{\perp}$.

6.3.4 Implications and Future Work

The DESI Y1 analysis is still ongoing, and we will eventually estimate the contribution of stochasticity in order for us to be able to account for it in the final error budget, i.e.,

$$\sigma_{\rm HOD} = \sqrt{\sigma^2_{\rm sys} - \sigma^2_{\rm stochastic}}. \tag{6.18}$$

In Fig. 6.5 we show the expansion rate of the Universe as a function of redshift, i.e., the evolution of the Hubble parameter. We include the theoretical prediction for $H(z)$ of ΛCDM with Planck-2018 cosmological parameters (black line), and also the forecasts for the error-bars of the measurements for the different dark matter tracers of DESI (BGS, LRGs, ELGs and QSOs). The forecasts for LRGs are the ones which yield to the smallest error-bars, which is another motivation to accurately estimate its systematics and keep them well under control. The DESI Y1 analysis is the first cosmological analysis ever in which systematics related to the HOD modeling have been studied in the history of large galaxy surveys. We have found an upper limit for these systematics of about 0.2% and 0.1% in the parallel and perpendicular directions with respect to the line of sight, respectively. This is a very small effect, which is the reason why they had never been studied for previous galaxy surveys (which had error-bars of, typically, $\sim$2-3%). However, because of DESI's sub-percent level of precision, we have shown these numbers represent a systematic contribution with an upper limit of about $0.2\sigma_{\rm stat}$, i.e., they have to be carefully accounted for in the final error-bars of our measurements. Incoming work will allow us to obtain the contribution of HOD systematics with higher precision for the DESI Y1 analysis, removing the stochastic contribution of our $0.2\sigma_{\rm stat}$ estimation. Furthermore, even more precise studies, e.g., increasing the variety of the HOD models considered, will have to be performed for future DESI analyses, since the precision of the BAO measurement will be even higher.

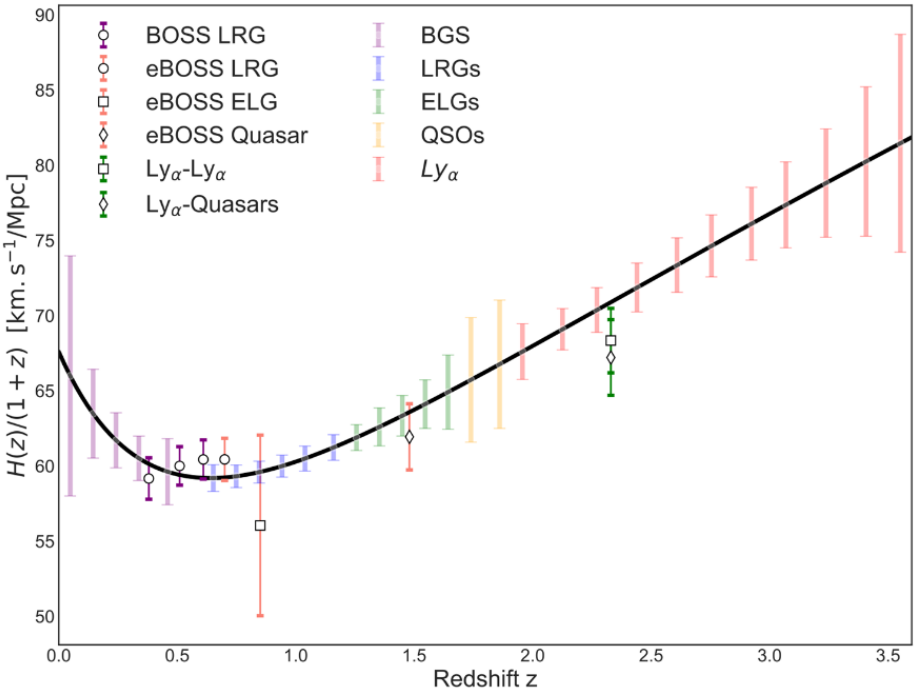

Figure 6.5: Expansion rate of the Universe as a function of redshift, i.e., evolution of the Hubble parameter. In particular, we plot $H(z)$ for a ΛCDM model with Planck-2018 cosmological parameters (black line), together with the expected error-bars for the different dark matter tracers studied by DESI (BGS, LRGs, ELGs and QSOs). Different measurements of $H(z)$ from BOSS and eBOSS are also included as points with error-bars. Original figure taken from [151].

Chapter 7

Conclusions

The main goal of this thesis is the precise measurement of the expansion history of the Universe using the BAO cosmological probe. This thesis was developed within two of the most important large galaxy surveys up to date: the Dark Energy Survey and the Dark Energy Spectroscopic Instrument. In the case of DES, we studied the BAO feature as measured from the two-point angular correlation function, for both the DES Y3 and Y6 analyses. We developed a BAO-fitting pipeline which was tested and validated using the fiducial set of simulations of the DES Y3 analysis, the Y3 COLA mocks, and then run on the Y3 data. It was also run on the Y6 COLA mocks, and will eventually be run on the Y6 data (which is still blinded). In the case of DESI, we studied the effect of HOD systematics in the BAO measurement for the DESI Y1 analysis.

One of the main results of this thesis is the template-based BAO-fitting pipeline we developed to measure the BAO distance scale from the angular correlation function. Since it is a template-based algorithm, the position of the BAO peak is given in terms of the BAO shift parameter, α, between the data and a template $w(\theta)$. For the modeling of the template $w(\theta)$, we took into account redshift-space distortions and also non-linear features in the evolution of the BAO peak. This template can be computed for any set of redshift distributions and for any set of cosmological parameters, but in this manuscript we show its results for two particular cosmologies: Mice and Planck. The pipeline has the option to compute the template C_ℓ as well, i.e., the Hankel transform of the $w(\theta)$, which is used for the Fourier space analysis; and also the possibility to compute the template $w(\theta)$ with no BAO feature at all, which can be used to compute the significance of the BAO measurement. We also included the option to fit for the linear galaxy bias by splitting the computation of the template $w(\theta)$ into three parts, which is particularly useful to measure the amplitude of the galaxy clustering signal of our data. Besides all these settings related to the template, it also has several other settings that can be adjusted to run fits with different configurations, such as the number of redshift bins included in the fit, the number of broadband-term parameters fitted, the scale-cuts for the fits, the use of different covariance matrices, and even the possibility to run on different sets

of simulations.

Our BAO-fitting pipeline was validated using the Y3 COLA mocks. We showed that its results are robust against variations in the template cosmology, and also against variations in several other settings. For the first test, we run the pipeline using templates computed for Mice and Planck cosmologies independently. We found that the best-fit α (evaluated at $z_{\rm eff} = 0.835$) is consistent with the expected value for both templates, being $\langle\alpha^{\rm Planck}\rangle/\langle\alpha^{\rm Mice}\rangle = 0.961$ the result from the fits and $\alpha^{\rm Planck}/\alpha^{\rm Mice} = 0.960$ the theoretical expectation. For the second test, we run the pipeline varying the number of redshift bins included in the fits, the number of broadband-term parameters fitted, the minimum θ and the covariance matrix, and we found consistent results for the different cases. Once validated, we used our pipeline to run the BAO measurements on the DES Y3 data, for which we obtained a distance measurement of $d_M(0.835)/r_d = 18.94 \pm 0.48$. This result is consistent with Planck at the 2.5σ level, and also consistent with the results of the official DES Y3 analysis, and represents the most precise measurement from a photometric galaxy survey up to date, with a relative error of 2.6%. We also combined our likelihood with the other main DES Y3 likelihoods, namely 3×2pt and SN. We found tighter constraints in h, Ω_m and σ_8 after the combination: the posterior in h is more symmetrical, with a gain in constraining power of ∼20%; the error in Ω_m is reduced by ∼25%; and the constraining power in σ_8 improves by ∼16%.

Another key result of this thesis is the development of an algorithm to optimize the DES Y6 BAO sample. Using the optimal sample, we forecast an increase in precision of about 25% in the BAO measurement with respect to the Y3 analysis. This optimization was carried out by adding two free parameters, a and b, to the i-magnitude limit in the sample selection: $i < a + bz_{\rm ph}$. These parameters were optimized by running forecasts in a grid of 10,000 samples, and we found the optimal values to be $a = 19.64$ and $b = 2.894$. The optimization was also run adding one more redshift bin with respect to the Y3, from $z = 1.1$ to $z = 1.2$, and we found the same values for the optimal parameters, and a slightly increase in the forecasted precision. We also validated the photometric redshifts of the optimal BAO sample, computed with DNF, using external spectroscopic samples, namely VIPERS and BOSS+eBOSS (these two were used to compute the clustering-z measurements). For the fiducial analysis on the data, we decided to use the redshift distributions estimated with DNF but modified in order to match the properties of those of VIPERS and clustering-z. For this modification we developed a shift and stretch algorithm, which allowed us to use the DNF estimations as templates for the shape of the final redshift distributions. We explicitly tested the impact of the redshift distribution in the BAO measurement by running our BAO-fitting pipeline to fit a template $w(\theta)$ computed with a given redshift distribution with a template computed with a different one, but for the same underlying cosmology. We found this effect to be negligible, at the level of 0.25σ (statistical) for the most extreme case.

Our BAO-fitting pipeline was also run on the Y6 COLA mocks, and we found that the increase in precision with respect to the Y3 mocks is consistent with what

we expected from forecasts, i.e., an increase of about 25%. Since the Y6 analysis was ongoing and still on its blinded phase at the moment of writing this dissertation, in this manuscript we did not show the BAO-fit results on the Y6 data. However, we did show the blinded measurement of the $w(\theta)$ by comparing the results of two independent codes, i.e., we plotted $\Delta w(\theta)$, and found that both of them yielded to the same clustering measurements. To finalize, we ran chains combining a Y6-like BAO likelihood (from the Y6 COLA mocks) with a likelihood coming from a synthetic datavector of 3×2pt. We obtained that we expect an increase in precision of ∼18% in Ω_m and of ∼16% in σ_8 after the combination with respect to 3×2pt alone.

Besides all the different analyses performed within DES, we also studied the effect of HOD systematics for the BAO measurement in the context of the DESI Y1 analysis for LRGs. This is the first cosmological analysis ever in which systematics related to the HOD modeling have been estimated in the history of large galaxy surveys. We have found an upper limit for these systematics of about 0.2% and 0.1% in $\alpha_{\parallel}$ and $\alpha_{\perp}$, respectively. Because of DESI's sub-percent level of precision, these numbers represent a systematic contribution of about $\leq 0.2\sigma_{\rm stat}$, i.e., they have to be carefully accounted for in the final error-bars of our measurements, since they are not negligible. Incoming work will allow us to obtain the contribution of HOD systematics with higher precision for the DESI Y1 analysis, removing the stochastic contribution of our $0.2\sigma_{\rm stat}$ estimation. Furthermore, even more precise studies, e.g., increasing the variety of the HOD models considered, will have to be performed for future DESI analyses, since the precision of the BAO measurement will be even higher.

With all the work presented in this thesis, we have confirmed that the BAO probe is a particularly suitable way to measure the expansion history of the Universe, being extremely robust against observational systematics. Another key aspect of the BAO is that it can be used as a consistency test with the high-redshift results coming from Planck. We have found that template-based methods are excellent algorithms to measure the BAO feature from the galaxy clustering signal, for both photometric and spectroscopic datasets. The BAO-fitting pipeline we developed could be used for BAO analyses of future photometric galaxy surveys, even though these kind of surveys are specifically designed to study weak lensing, typically via the 3×2pt cosmological probe. However, we have shown that the BAO improves the parameter estimation of 3×2pt, increasing the constraining power in Ω_m and σ_8. Our pipeline can also be used to generate the template $\xi_\ell(s)$ or $P_\ell(k)$ for BAO analyses of spectroscopic surveys by slightly changing the code, and part of the algorithm can be adapted to run 2-dimensional BAO fits. On the other hand, the algorithm we developed to optimize the sample selection for the Y6 can be used for future cosmological analyses as well, since it can be modified to run forecasts of different natures. As a concluding remark, the results of the DES Y6 BAO analysis will be the most accurate for a photometric redshift survey ever, with an error in the distance measurement of ∼2%, which will also be the most precise measurement at redshift ∼0.85 up to date, competitive and even better than those of spectroscopic redshift surveys. Also, the results of the DESI Y1 analysis will measure the expansion history with an even higher and

unprecedented precision, which is the reason why HOD systematics have to be very carefully estimated. We forecast statistical errors of about 1.2% and 0.5% in $\alpha_{\parallel}$ and $\alpha_{\perp}$, respectively, for the DESI Y1 analysis, and the results to be released by the end of 2023.

Appendix A

Appendix: Large-Scale Structure

A.1 Random Fields and Gaussian Perturbations

Inhomogeneities in the matter distribution of the Universe give us information about its origin and evolution, and allow us to study the nature of its components. Inflationary models predict density perturbations that are generated by Gaussian quantum fluctuations of a scalar field. Therefore, the density perturbations are random fields with Gaussian distribution, and they are completely described by the matter power spectrum $P_m(k)$, which is given by

$$\langle \delta_c(\boldsymbol{k}) \delta_c^*(\boldsymbol{k}') \rangle = P_m(k) \delta^3(\boldsymbol{k} - \boldsymbol{k}'). \tag{A.1}$$

A.1.1 Gaussian Perturbations: Discrete Limit

Let's write the expansion of a function $g(\boldsymbol{x})$ in Fourier modes as

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k}} g_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{A.2}$$

The $g_{\boldsymbol{k}}$ correspond to Gaussian random variables, i.e., their probability distribution function, $P(g_{\boldsymbol{k}})$, is Gaussian. These Gaussian random variables are, in general, complex numbers. Therefore, we can write

$$g_{\boldsymbol{k}} = \alpha_{\boldsymbol{k}} + i\beta_{\boldsymbol{k}}. \tag{A.3}$$

However, even though $g_{\boldsymbol{k}}$ are, in general, complex, $g(\boldsymbol{x})$ must be real, which implies

$$g_{\boldsymbol{k}}^* = g_{-\boldsymbol{k}}. \tag{A.4}$$

Also, since $P(g_{\boldsymbol{k}})$ is Gaussian, it can be written as

$$\begin{aligned} P(g_{\boldsymbol{k}}) &= \frac{1}{2\pi\sigma_{\boldsymbol{k}}^2} \exp\left(-\frac{1}{2}\frac{|g_{\boldsymbol{k}}|^2}{\sigma_{\boldsymbol{k}}^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{\boldsymbol{k}}} \exp\left(-\frac{1}{2}\frac{\alpha_{\boldsymbol{k}}^2}{\sigma_{\boldsymbol{k}}^2}\right) \frac{1}{\sqrt{2\pi}\sigma_{\boldsymbol{k}}} \exp\left(-\frac{1}{2}\frac{\beta_{\boldsymbol{k}}^2}{\sigma_{\boldsymbol{k}}^2}\right). \end{aligned} \tag{A.5}$$

By definition, the mean value of any function $f(g_{\boldsymbol{k}})$ with probability distribution $P(g_{\boldsymbol{k}})$ can be computed as

$$\langle f(g_{\boldsymbol{k}})\rangle = \int_{-\infty}^{\infty} d\alpha_{\boldsymbol{k}} \int_{-\infty}^{\infty} d\beta_{\boldsymbol{k}} f(g_{\boldsymbol{k}}) P(g_{\boldsymbol{k}}), \tag{A.6}$$

which leads to

$$\begin{cases} \langle g_{\boldsymbol{k}}\rangle = 0, \\ \langle |g_{\boldsymbol{k}}|^2\rangle = 2\sigma_{\boldsymbol{k}}^2. \end{cases} \tag{A.7}$$

This is an expected result, since, by construction, the mean value of any Gaussian random fluctuation is zero because of the definition of $P(g_{\boldsymbol{k}})$, whereas the variance is given by that of the Gaussian distribution. We also assume isotropy and homogeneity:

- Isotropy. Isotropy implies that there is no privileged direction in space, which in turn translates into

$$\sigma_{\boldsymbol{k}} = \sigma_k. \tag{A.8}$$

- Homogeneity. This implies that $g_{\boldsymbol{k}}$ and $g_{\boldsymbol{k}'}$ are not correlated,

$$\begin{cases} \langle g_{\boldsymbol{k}} g^*_{\boldsymbol{k}'}\rangle = 0, \quad \boldsymbol{k} \neq \boldsymbol{k}', \\ \langle g_{\boldsymbol{k}} g^*_{\boldsymbol{k}}\rangle = \langle |g_{\boldsymbol{k}}|^2\rangle = 2\sigma_{\boldsymbol{k}}^2. \end{cases} \tag{A.9}$$

  Therefore,

$$\langle g_{\boldsymbol{k}} g^*_{\boldsymbol{k}'}\rangle = \langle |g_{\boldsymbol{k}}|^2\rangle \delta_{\boldsymbol{k}\boldsymbol{k}'} = 2\sigma_{\boldsymbol{k}}^2 \delta_{\boldsymbol{k}\boldsymbol{k}'}. \tag{A.10}$$

Let's go back to real space and assume we have a random fluctuation $g(\boldsymbol{x})$. By definition, the mean value of $g(\boldsymbol{x})$ is zero because it's a random fluctuation. On the other hand, the variance of $g(\boldsymbol{x})$ is not zero, but it does not depend on the position. Explicitly, $g(\boldsymbol{x})^2$ can be written as

$$g(\boldsymbol{x})^2 = g^*(\boldsymbol{x}) g(\boldsymbol{x}) = \sum_{\boldsymbol{k}} g^*_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \sum_{\boldsymbol{k}'} g_{\boldsymbol{k}'} e^{i\boldsymbol{k}'\cdot\boldsymbol{x}}. \tag{A.11}$$

Taking the mean value in the previous expression, we find that the variance of $g(\boldsymbol{x})$ is given by

$$\langle g(\boldsymbol{x})^2\rangle = \sum_{\boldsymbol{k}} \sum_{\boldsymbol{k}'} \langle g_{\boldsymbol{k}'} g^*_{\boldsymbol{k}}\rangle e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{x}} = \sum_{\boldsymbol{k}} \langle |g_{\boldsymbol{k}}|^2\rangle = 2\sum_{\boldsymbol{k}} \sigma_{\boldsymbol{k}}^2. \tag{A.12}$$

Not only the mean of $g(\boldsymbol{x})$ does not depend on the position, but also the variance is independent. This is a consequence of the homogeneity of the initial perturbation.

A.1.2 Gaussian Perturbations: Continuous Limit

The objective here is to rewrite the results of the previous section for the continuous case. The Fourier transform of a function $g(\boldsymbol{x})$ is defined as

$$g(\boldsymbol{k}) = \frac{1}{(2\pi)^{3/2}} \int d^3\boldsymbol{x} g(\boldsymbol{x}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{A.13}$$

By definition of the Fourier transform, we already know that

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{k}} g_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \rightarrow g(\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}} \int d^3\boldsymbol{k} g(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{A.14}$$

We also know that $\Delta k = 2\pi/L$ in the discrete limit, which implies

$$\sum_{\boldsymbol{k}} \rightarrow \left(\frac{L}{2\pi}\right)^3 \int d^3\boldsymbol{k}. \tag{A.15}$$

Using eqs. (A.13) and (A.14), we easily find that

$$g_{\boldsymbol{k}} \rightarrow \frac{1}{(2\pi)^{3/2}} \left(\frac{2\pi}{L}\right)^3 g(\boldsymbol{k}). \tag{A.16}$$

In addition, it must be taken into account that

$$\delta_{\boldsymbol{k}\boldsymbol{k}'} \rightarrow \left(\frac{2\pi}{L}\right)^3 \delta^3(\boldsymbol{k} - \boldsymbol{k}'). \tag{A.17}$$

As we already showed before,

$$\langle g(\boldsymbol{x})^2 \rangle = \sum_{\boldsymbol{k}} \langle |g_{\boldsymbol{k}}|^2 \rangle. \tag{A.18}$$

Therefore, in the continuous limit,

$$\sum_{\boldsymbol{k}} \langle |g_{\boldsymbol{k}}|^2 \rangle \rightarrow \left(\frac{L}{2\pi}\right)^3 \int d^3\boldsymbol{k} \langle |g_{\boldsymbol{k}}|^2 \rangle = \int \frac{dk}{k} V \frac{k^3}{2\pi^2} \langle |g_{\boldsymbol{k}}|^2 \rangle = \int \frac{dk}{k} P_g(k), \tag{A.19}$$

where we defined the power spectrum of g as

$$P_g(k) \equiv V \frac{k^3}{2\pi^2} \langle |g_{\boldsymbol{k}}|^2 \rangle. \tag{A.20}$$

The square root of the power spectrum, $P_g(k)^{1/2}$, is the typical amplitude of g at the scale k. Explicitly, the variance of $g(\boldsymbol{x})$ is, finally, given by

$$\langle g(\boldsymbol{x})^2 \rangle = \int \frac{dk}{k} V \frac{k^3}{2\pi^2} \langle |g_{\boldsymbol{k}}|^2 \rangle = \int \frac{dk}{k} P_g(k). \tag{A.21}$$

All the previous results are physics independent, in the sense that all the assumptions we have made are purely mathematical, from the random nature of our fluctuations to the isotropy and homogeneity. Therefore, any generic field $g(\boldsymbol{x})$ that meets these properties, such as the matter density fluctuations $\delta(\boldsymbol{x})$, can be assigned a power spectrum $P_g(k)$.

A.2 The σ_8 Parameter

The σ_8 parameter is related to the concept of smoothing. Large cosmological surveys observe the density contrast δ_m as a function of the position, $\boldsymbol{x}$. This density contrast measures the average matter density at a given point with respect to the average measured everywhere in the Universe. Therefore, this function is extremely noisy, since it is very sensitive to changes from observing empty space to observing some matter distribution, such as a galaxy. This problem is solved by smoothing our density contrast at a given scale, R, by convolving δ_m with a smoothing function W_R,

$$\delta_m(\boldsymbol{x}, R) = \int_0^\infty d\boldsymbol{x}' W_R(|\boldsymbol{x} - \boldsymbol{x}'|)\delta_m(\boldsymbol{x}'). \tag{A.22}$$

Because of the convolution theorem, which states that the Fourier transform of the convolution is the product of the Fourier transforms, we easily find that

$$\delta_m(\boldsymbol{k}, R) = W_R(k)\delta_m(\boldsymbol{k}). \tag{A.23}$$

A typical choice for the real-space smoothing function, $W_R(|\boldsymbol{x} - \boldsymbol{x}'|)$, is a top-hat function,

$$W_R(|\boldsymbol{x} - \boldsymbol{x}'|) = \begin{cases} 1, & |\boldsymbol{x} - \boldsymbol{x}'| < R, \\ 0, & \text{otherwise.} \end{cases} \tag{A.24}$$

The Fourier transform of the top-hat function is none other than the spherical Bessel function of first order,

$$W_R(k) = 3\frac{j_1(kR)}{kR}. \tag{A.25}$$

Therefore, using eq. (A.23),

$$\delta_m(\boldsymbol{k}, R) = 3\frac{j_1(kR)}{kR}\delta_m(\boldsymbol{k}). \tag{A.26}$$

Computing the variance of δ_m, using eq. (2.50), we easily find that

$$\sigma_R^2 = \langle \delta_m(\boldsymbol{x}, R)^2 \rangle = \int \frac{dk}{k} V \frac{k^3}{2\pi^2} \langle |\delta_m(\boldsymbol{k}, R)|^2 \rangle \tag{A.27}$$

$$= \int \frac{dk}{k} V \frac{k^3}{2\pi^2} \left[3\frac{j_1(kR)}{kR}\right]^2 \langle |\delta_m(\boldsymbol{k})|^2 \rangle = \int dk \frac{k^2}{2\pi^2} \left[3\frac{j_1(kR)}{kR}\right]^2 P_m(k). \tag{A.28}$$

The previous expression, evaluated at $R = 8$ Mpc$/h$, gives us the definition of σ_8, eq. (2.59).

A.3 The Limber Approximation

The Limber approximation is widely used in the literature to relate projected angular clustering of galaxies to the spatial clustering of galaxies. It is used to simplify the expression for the calculation of either $w(\theta)$ or its Fourier transform, C_ℓ, by approximating the spherical Bessel function as

$$j_\ell(x) \approx \sqrt{\frac{\pi}{2\ell+1}}\delta(\ell+1/2-x), \tag{A.29}$$

where δ is the Dirac delta. Then, using the property of the Dirac delta that allows us to express $\delta(f(x))$ as

$$\delta(f(x)) = \sum_i \frac{\delta(x-x_i)}{|f'(x)|_{x=x_i}}, \quad \text{with } x_i \text{ such that } f(x_i)=0, \tag{A.30}$$

we find that

$$j_\ell(k\chi) \approx \sqrt{\frac{\pi}{2\ell+1}}\delta(\ell+1/2-k\chi) = \sqrt{\frac{\pi}{2\ell+1}}\delta\left(k-\frac{\ell+1/2}{\chi}\right)\frac{1}{\chi}. \tag{A.31}$$

This approximation is not accurate for small scales, i.e., for small values of θ or large values of ℓ, but it is particularly useful otherwise since it allows us to compute either $w(\theta)$ or C_ℓ as a single integral, instead of a triple one.

Bibliography

[1] Georges Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. In *Annales de la Société scientifique de Bruxelles*, volume 47, pages 49–59, 1927. 17

[2] Edwin Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929. 17, 34

[3] Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421, 1965. 17

[4] AA Starobinskii. Spectrum of relict gravitational radiation and the early state of the universe. *JETP Letters*, 30(11):682–685, 1979. 18

[5] Alan H Guth and S-HH Tye. Phase transitions and magnetic monopole production in the very early universe. *Physical Review Letters*, 44(10):631, 1980. 18

[6] Andreas Albrecht and Paul J Steinhardt. Cosmology for grand unified theories with radiatively induced symmetry breaking. *Physical Review Letters*, 48(17):1220, 1982. 18

[7] Andrei D Linde. Chaotic inflation. *Physics Letters B*, 129(3-4):177–181, 1983. 18

[8] Edmund J Copeland, Mohammad Sami, and Shinji Tsujikawa. Dynamics of dark energy. *International Journal of Modern Physics D*, 15(11):1753–1935, 2006. 19, 20

[9] Saul Perlmutter, Goldhaber Aldering, Gerson Goldhaber, RA Knop, Peter Nugent, Patricia G Castro, Susana Deustua, Sebastien Fabbro, Ariel Goobar, Donald E Groom, et al. Measurements of Ω and Λ from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, 1999. 19, 21, 90

[10] Robert J Scherrer and Michael S Turner. On the relic, cosmic abundance of stable, weakly interacting massive particles. *Physical Review D*, 33(6):1585, 1986. 21

[11] Albert Einstein. *Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie*. Gedruckt in der Reichsdruckerei, 1915. 22

[12] John D Barrow and Frank J Tipler. The anthropic cosmological principle. *Oxford: Clarendon Press*, 1986. 23

[13] JD Barrow. What is the principal evidence for the cosmological principle? *Quarterly Journal of the Royal Astronomical Society*, 30:163–167, 1989. 23

[14] Nabila Aghanim, Yashar Akrami, Mark Ashdown, J Aumont, C Baccigalupi, M Ballardini, AJ Banday, RB Barreiro, N Bartolo, S Basak, et al. Planck 2018 results-VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020. 28, 29, 32, 33, 34, 46, 53, 119, 138, 193

[15] Wayne Hu and Scott Dodelson. Cosmic microwave background anisotropies. *Annual Review of Astronomy and Astrophysics*, 40(1):171–216, 2002. 28

[16] Richard H Cyburt, Brian D Fields, Keith A Olive, and Tsung-Han Yeh. Big bang nucleosynthesis: Present status. *Reviews of Modern Physics*, 88(1):015004, 2016. 28

[17] Maxim Pospelov and Josef Pradler. Big bang nucleosynthesis as a probe of new physics. *Annual Review of Nuclear and Particle Science*, 60:539–568, 2010. 31

[18] Volker Springel, Carlos S Frenk, and Simon DM White. The large-scale structure of the Universe. *nature*, 440(7088):1137–1144, 2006. 30

[19] Pablo F De Salas, Stefano Gariazzo, Olga Mena, Christoph A Ternes, and Mariam Tórtola. Neutrino mass ordering from oscillations and beyond: 2018 status and future prospects. *Frontiers in Astronomy and Space Sciences*, 5:36, 2018. 34

[20] Bruce Bassett and Renée Hlozek. Baryon acoustic oscillations. *Dark energy: observational and theoretical approaches*, 246, 2010. 53

[21] Peter Meszaros. The behaviour of point masses in an expanding cosmological substratum. *Astronomy and Astrophysics*, 37:225–228, 1974. 59

[22] Phillip James Edwin Peebles. *The large-scale structure of the universe*, volume 98. Princeton university press, 2020. 60

[23] Dragan Huterer, David Kirkby, Rachel Bean, Andrew Connolly, Kyle Dawson, Scott Dodelson, August Evrard, Bhuvnesh Jain, Michael Jarvis, Eric Linder, et al. Growth of cosmic structure: Probing dark energy beyond expansion. *Astroparticle Physics*, 63:23–41, 2015. 60

[24] Eric V Linder. Cosmic growth history and expansion history. *Physical Review D*, 72(4):043529, 2005. 60

[25] Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models. *The Astrophysical Journal*, 538(2):473, 2000. 61, 63

[26] Robert E Smith, J A Peacock, A Jenkins, SDM White, CS Frenk, FR Pearce, Peter A Thomas, G Efstathiou, and HMP Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *Monthly Notices of the Royal Astronomical Society*, 341(4):1311–1332, 2003. 61

[27] M Crocce, FJ Castander, E Gaztañaga, P Fosalba, and J Carretero. The MICE Grand Challenge lightcone simulation–II. Halo and galaxy catalogues. *Monthly Notices of the Royal Astronomical Society*, 453(2):1513–1530, 2015. 63, 113

[28] T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila, D. Bacon, E. Baxter, K. Bechtol, M. R. Becker, G. M. Bernstein, S. Bhargava, S. Birrer, J. Blazek, A. Brandao-Souza, et al. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 105:023520, Jan 2022. 64, 149, 150, 172

[29] Martín Crocce, Anna Cabré, and Enrique Gaztañaga. Modelling the angular correlation function and its full covariance in photometric galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 414(1):329–349, 2011. 70, 71

[30] Stephen D Landy and Alexander S Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64–71, 1993. 71, 136

[31] Scott Dodelson and Fabian Schmidt. *Modern cosmology*. Academic press, 2020. 71

[32] Daniel J Eisenstein, Idit Zehavi, David W Hogg, Roman Scoccimarro, Michael R Blanton, Robert C Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, et al. Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *The Astrophysical Journal*, 633(2):560, 2005. 74, 75, 81

[33] Shaun Cole, Will J Percival, John A Peacock, Peder Norberg, Carlton M Baugh, Carlos S Frenk, Ivan Baldry, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, et al. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the Royal Astronomical Society*, 362(2):505–534, 2005. 74, 81

[34] Charles Alcock and Bohdan Paczyński. An evolution free test for non-zero cosmological constant. *Nature*, 281(5730):358–359, 1979. 76, 80

[35] Shadab Alam, Marie Aubert, Santiago Avila, Christophe Balland, Julian E Bautista, Matthew A Bershady, Dmitry Bizyaev, Michael R Blanton, Adam S Bolton, Jo Bovy, et al. Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory. *Physical Review D*, 103(8):083533, 2021. 76, 83, 142, 143

[36] Nick Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 07 1987. 76, 78

[37] JC Jackson. A critique of Rees's theory of primordial gravitational radiation. *Monthly Notices of the Royal Astronomical Society*, 156(1):1P–5P, 1972. 76

[38] Hector Gil-Maŕin, Julían E Bautista, Romain Paviot, Mariana Vargas-Magaña, Sylvain De La Torre, Sebastien Fromenteau, Shadab Alam, Santiago Ávila, Etienne Burtin, Chia Hsun Chuang, et al. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0. *Monthly Notices of the Royal Astronomical Society*, 498(2):2492–2531, 2020. 79, 80, 83, 114, 141

[39] Ashley J Ross, Will J Percival, and Marc Manera. The information content of anisotropic Baryon Acoustic Oscillation scale measurements. *Monthly Notices of the Royal Astronomical Society*, 451(2):1331–1340, 2015. 81

[40] Florian Beutler, Chris Blake, Matthew Colless, D Heath Jones, Lister Staveley-Smith, Lachlan Campbell, Quentin Parker, Will Saunders, and Fred Watson. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *Monthly Notices of the Royal Astronomical Society*, 416(4):3017–3032, 2011. 81

[41] David Parkinson, Signe Riemer-Sørensen, Chris Blake, Gregory B Poole, Tamara M Davis, Sarah Brough, Matthew Colless, Carlos Contreras, Warrick Couch, Scott Croom, et al. The WiggleZ dark energy survey: final data release and cosmological results. *Physical Review D*, 86(10):103518, 2012. 81

[42] Kyle S Dawson, David J Schlegel, Christopher P Ahn, Scott F Anderson, Éric Aubourg, Stephen Bailey, Robert H Barkhouser, Julian E Bautista, Alessandra Beifiori, Andreas A Berlind, et al. The Baryon oscillation spectroscopic survey of SDSS-III. *The Astronomical Journal*, 145(1):10, 2012. 82, 86

[43] Lauren Anderson, Éric Aubourg, Stephen Bailey, Florian Beutler, Vaishali Bhardwaj, Michael Blanton, Adam S Bolton, Jon Brinkmann, Joel R Brownstein, Angela Burden, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 441(1):24–62, 2014. 82

[44] Ashley J Ross, Florian Beutler, Chia-Hsun Chuang, Marcos Pellejero-Ibanez, Hee-Jong Seo, Mariana Vargas-Magana, Antonio J Cuesta, Will J Percival, Angela Burden, Ariel G Sánchez, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function. *Monthly Notices of the Royal Astronomical Society*, 464(1):1168–1191, 2017. 82

[45] Florian Beutler, Hee-Jong Seo, Ashley J Ross, Patrick McDonald, Shun Saito, Adam S Bolton, Joel R Brownstein, Chia-Hsun Chuang, Antonio J Cuesta, Daniel J Eisenstein, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space. *Monthly Notices of the Royal Astronomical Society*, 464(3):3409–3430, 2017. 82, 199

[46] Mariana Vargas-Magaña, Shirley Ho, Antonio J Cuesta, Ross O'Connell, Ashley J Ross, Daniel J Eisenstein, Will J Percival, Jan Niklas Grieb, Ariel G Sánchez, Jeremy L Tinker, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: theoretical systematics and Baryon Acoustic Oscillations in the galaxy correlation function. *Monthly Notices of the Royal Astronomical Society*, 477(1):1153–1188, 2018. 82

[47] Helion Du Mas Des Bourboux, James Rich, Andreu Font-Ribera, Victoria de Sainte Agathe, James Farr, Thomas Etourneau, Jean-Marc Le Goff, Andrei Cuceu, Christophe Balland, Julian E Bautista, et al. The Completed SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations with Lyα Forests. *The Astrophysical Journal*, 901(2):153, 2020. 82, 83, 141

[48] Éric Aubourg, Stephen Bailey, Julian E. Bautista, Florian Beutler, Vaishali Bhardwaj, Dmitry Bizyaev, Michael Blanton, Michael Blomqvist, Adam S. Bolton, Jo Bovy, Howard Brewington, J. Brinkmann, Joel R. Brownstein, Angela Burden, Nicolás G. Busca, William Carithers, Chia-Hsun Chuang, Johan Comparat, Rupert A. C. Croft, Antonio J. Cuesta, Kyle S. Dawson, Timothée

Delubac, Daniel J. Eisenstein, Andreu Font-Ribera, et al. Cosmological implications of baryon acoustic oscillation measurements. *Phys. Rev. D*, 92:123516, Dec 2015. 82

[49] Elisabete Da Cunha, Andrew M Hopkins, Matthew Colless, Edward N Taylor, Chris Blake, Cullan Howlett, Christina Magoulas, John R Lucey, Claudia Lagos, Kyler Kuehn, et al. The Taipan galaxy survey: scientific goals and observing strategy. *Publications of the Astronomical Society of Australia*, 34:e047, 2017. 82

[50] T. M. C. Abbott, M. Aguena, S. Allam, A. Amon, F. Andrade-Oliveira, J. Asorey, S. Avila, G. M. Bernstein, E. Bertin, A. Brandao-Souza, D. Brooks, D. L. Burke, J. Calcino, H. Camacho, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, K. C. Chan, et al. Dark Energy Survey Year 3 results: A 2.7% measurement of baryon acoustic oscillation distance scale at redshift 0.835. *Phys. Rev. D*, 105:043512, Feb 2022. 83, 106, 117, 118, 122, 129, 135, 142

[51] Julian E Bautista, Romain Paviot, Mariana Vargas Magaña, Sylvain de la Torre, Sebastien Fromenteau, Hector Gil-Marín, Ashley J Ross, Etienne Burtin, Kyle S Dawson, Jiamin Hou, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *Monthly Notices of the Royal Astronomical Society*, 500(1):736–762, 2021. 83, 114, 141

[52] Arnaud De Mattia, Vanina Ruhlmann-Kleider, Anand Raichoor, Ashley J Ross, Amélie Tamone, Cheng Zhao, Shadab Alam, Santiago Avila, Etienne Burtin, Julian Bautista, et al. The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1. *Monthly Notices of the Royal Astronomical Society*, 501(4):5616–5645, 2021. 83, 114, 141

[53] Jiamin Hou, Ariel G Sánchez, Ashley J Ross, Alex Smith, Richard Neveux, Julian Bautista, Etienne Burtin, Cheng Zhao, Román Scoccimarro, Kyle S Dawson, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the quasar sample in configuration space between redshift 0.8 and 2.2. *Monthly Notices of the Royal Astronomical Society*, 500(1):1201–1221, 2021. 83, 141

[54] Richard Neveux, Etienne Burtin, Arnaud de Mattia, Alex Smith, Ashley J Ross, Jiamin Hou, Julian Bautista, Jonathan Brinkmann, Chia-Hsun Chuang, Kyle S Dawson, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power

spectrum of the quasar sample between redshift 0.8 and 2.2. *Monthly Notices of the Royal Astronomical Society*, 499(1):210–229, 2020. 83, 141

[55] Giuseppe D Racca, René Laureijs, Luca Stagnaro, Jean-Christophe Salvignol, José Lorenzo Alvarez, Gonzalo Saavedra Criado, Luis Gaspar Venancio, Alex Short, Paolo Strada, Tobias Bönke, et al. The Euclid mission design. In *Space telescopes and instrumentation 2016: optical, infrared, and millimeter wave*, volume 9904, pages 235–257. SPIE, 2016. 83

[56] Catherine Heymans, Ludovic Van Waerbeke, Lance Miller, Thomas Erben, Hendrik Hildebrandt, Henk Hoekstra, Thomas D Kitching, Yannick Mellier, Patrick Simon, Christopher Bonnett, et al. CFHTLenS: the Canada–France–Hawaii telescope lensing survey. *Monthly Notices of the Royal Astronomical Society*, 427(1):146–166, 2012. 85

[57] Jelte TA de Jong, Gijs A Verdoes Kleijn, Konrad H Kuijken, and Edwin A Valentijn. The kilo-degree survey. *Experimental Astronomy*, 35(1):25–44, 2013. 85

[58] Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Stéphane Arnouts, Neta A Bahcall, Steven Bickerton, James Bosch, Kevin Bundy, Peter L Capak, James HH Chan, et al. The Hyper Suprime-Cam SSP survey: overview and survey design. *Publications of the Astronomical Society of Japan*, 70(SP1):S4, 2018. 85

[59] Dark Energy Survey Collaboration et al. The Dark Energy Survey. arXiv e-prints, pages astro–ph/0510346, 2005. 85

[60] Donald G York, J Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579, 2000. 86

[61] Kyle S Dawson, Jean-Paul Kneib, Will J Percival, Shadab Alam, Franco D Albareti, Scott F Anderson, Eric Armengaud, Éric Aubourg, Stephen Bailey, Julian E Bautista, et al. The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: overview and early data. *The Astronomical Journal*, 151(2):44, 2016. 86

[62] Matthew Colless, Gavin Dalton, Steve Maddox, Will Sutherland, Peder Norberg, Shaun Cole, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Chris Collins, et al. The 2df galaxy redshift survey: spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328(4):1039–1063, 2001. 86

[63] Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, Otger

Ballester, et al. The desi experiment part ii: Instrument design. *arXiv preprint arXiv:1611.00037*, 2016. 86

[64] Brenna Flaugher, HT Diehl, K Honscheid, TMC Abbott, O Alvarez, R Angstadt, JT Annis, M Antonik, O Ballester, L Beaufore, et al. The dark energy camera. *The Astronomical Journal*, 150(5):150, 2015. 86, 87

[65] Matthew Smith, Chris B D'Andrea, M Sullivan, A Möller, RC Nichol, RC Thomas, AG Kim, M Sako, FJ Castander, AV Filippenko, et al. First cosmology results using supernovae ia from the dark energy survey: Survey overview, performance, and supernova spectroscopy. *The Astronomical Journal*, 160(6):267, 2020. 87

[66] Hiroaki Aihara, Yusra AlSayyad, Makoto Ando, Robert Armstrong, James Bosch, Eiichi Egami, Hisanori Furusawa, Junko Furusawa, Andy Goulding, Yuichi Harikane, et al. Second data release of the Hyper Suprime-Cam Subaru strategic program. *Publications of the Astronomical Society of Japan*, 71(6):114, 2019. 87

[67] Richard G McMahon, Manda Banerji, Eduardo Gonzalez, Sergey E Koposov, Victor J Bejar, Nicolas Lodieu, Rafael Rebolo, VHS collaboration, et al. First scientific results from the VISTA hemisphere survey (VHS). *The Messenger*, 154(35-37):188, 2013. 87

[68] Ignacio Sevilla-Noarbe, K Bechtol, M Carrasco Kind, A Carnero Rosell, MR Becker, A Drlica-Wagner, RA Gruendl, ES Rykoff, E Sheldon, B Yanny, et al. Dark energy survey year 3 results: Photometric data set for cosmology. *The Astrophysical Journal Supplement Series*, 254(2):24, 2021. 87, 106, 108

[69] Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998. 90

[70] Shantanu Desai, TMC Abbott, FB Abdalla, et al. The Dark Energy Survey: Data Release 1. *The American Astronomical Society*, 239(2), 2018. 91

[71] Emmanuel Bertin and Stephane Arnouts. SExtractor: Software for source extraction. *Astronomy and astrophysics supplement series*, 117(2):393–404, 1996. 92

[72] M Rodríguez-Monroy, N Weaverdyck, J Elvin-Poole, M Crocce, A Carnero Rosell, F Andrade-Oliveira, S Avila, K Bechtol, G M Bernstein, J Blazek, H Camacho, R Cawthon, J De Vicente, J DeRose, S Dodelson, S Everett, X Fang, I Ferrero, A Ferté, O Friedrich, E Gaztanaga, G Giannini,

R A Gruendl, W G Hartley, K Herner, E M Huff, M Jarvis, et al. Dark Energy Survey Year 3 results: galaxy clustering and systematics treatment for lens galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 511(2):2665–2687, Jan 2022. 92, 110, 112, 136, 137, 141, 177

[73] Eduardo Rozo, ES Rykoff, Alexandra Abate, C Bonnett, Martin Crocce, C Davis, Ben Hoyle, Boris Leistedt, Hiranya V Peiris, Risa H Wechsler, et al. redMaGiC: selecting luminous red galaxies from the DES Science Verification data. *Monthly Notices of the Royal Astronomical Society*, 461(2):1431–1450, 2016. 92, 93, 106

[74] Anna Porredon, M Crocce, J Elvin-Poole, R Cawthon, G Giannini, J De Vicente, A Carnero Rosell, Ismael Ferrero, E Krause, X Fang, et al. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and galaxy-galaxy lensing using the MagLim lens sample. *Physical Review D*, 106(10):103530, 2022. 93, 106

[75] Erin S. Sheldon and Eric M. Huff. Practical Weak-lensing Shear Measurement with Metacalibration. *The Astrophysical Journal*, 841(1):24, may 2017. 93

[76] Eric Huff and Rachel Mandelbaum. Metacalibration: direct self-calibration of biases in shear measurement. *arXiv preprint arXiv:1702.02600*, 2017. 93

[77] Daniel J Eisenstein, James Annis, James E Gunn, Alexander S Szalay, Andrew J Connolly, RC Nichol, Neta A Bahcall, Mariangela Bernardi, Scott Burles, Francisco J Castander, et al. Spectroscopic target selection for the Sloan Digital Sky Survey: The Luminous red galaxy sample. *The Astronomical Journal*, 122(5):2267, 2001. 105

[78] Mohammadjavad Vakili, Maciej Bilicki, Henk Hoekstra, Nora Elisa Chisari, Michael JI Brown, Christos Georgiou, Arun Kannawadi, Konrad Kuijken, and Angus H Wright. Luminous red galaxies in the Kilo-Degree Survey: selection with broad-band photometry and weak lensing measurements. *Monthly Notices of the Royal Astronomical Society*, 487(3):3715–3733, 2019. 105

[79] Rongpu Zhou, Jeffrey A Newman, Kyle S Dawson, Daniel J Eisenstein, David D Brooks, Arjun Dey, Biprateep Dey, Yutong Duan, Sarah Eftekharzadeh, Enrique Gaztañaga, et al. Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample. *Research Notes of the AAS*, 4(10):181, 2020. 105

[80] Martin Crocce, AJ Ross, Ignacio Sevilla-Noarbe, Enrique Gaztanaga, Jack Elvin-Poole, Santiago Avila, Alex Alarcon, Kwan Chuen Chan, N Banik, Jorge Carretero, et al. Dark Energy Survey year 1 results: galaxy sample for BAO measurement. *Monthly Notices of the Royal Astronomical Society*, 482(2):2807–2822, 2019. 106, 107, 108

[81] Alex Drlica-Wagner, Ignacio Sevilla-Noarbe, Eli S Rykoff, RA Gruendl, B Yanny, DL Tucker, B Hoyle, A Carnero Rosell, GM Bernstein, K Bechtol, et al. Dark energy survey year 1 results: The photometric data set for cosmology. *The Astrophysical Journal Supplement Series*, 235(2):33, 2018. 108

[82] Erin S Sheldon. An implementation of Bayesian lensing shear measurement. *Monthly Notices of the Royal Astronomical Society: Letters*, 444(1):L25–L29, 2014. 108

[83] A Carnero Rosell, M Rodriguez-Monroy, M Crocce, J Elvin-Poole, A Porredon, I Ferrero, J Mena-Fernández, R Cawthon, J De Vicente, E Gaztanaga, A J Ross, E Sanchez, I Sevilla-Noarbe, O Alves, F Andrade-Oliveira, J Asorey, S Avila, A Brandao-Souza, H Camacho, K C Chan, A Ferté, J Muir, W Riquelme, et al. Dark Energy Survey Year 3 results: galaxy sample for BAO measurement. *Monthly Notices of the Royal Astronomical Society*, 509(1):778–799, 10 2021. 108, 110, 112, 141

[84] Juan De Vicente, Eusebio Sánchez, and Ignacio Sevilla-Noarbe. DNF–Galaxy photometric redshift by Directional Neighbourhood Fitting. *Monthly Notices of the Royal Astronomical Society*, 459(3):3078–3088, 2016. 108, 109, 167

[85] MARCO Scodeggio, L Guzzo, BIANCA Garilli, BR Granett, M Bolzonella, S De La Torre, U Abbas, C Adami, S Arnouts, D Bottini, et al. The VIMOS Public Extragalactic Redshift Survey (VIPERS)-Full spectroscopic data and auxiliary information release (PDR-2). *Astronomy & Astrophysics*, 609:A84, 2018. 109, 155, 167, 169

[86] Jack Elvin-Poole, Martin Crocce, AJ Ross, T Giannantonio, E Rozo, ES Rykoff, S Avila, N Banik, J Blazek, SL Bridle, et al. Dark Energy Survey year 1 results: Galaxy clustering for combined probes. *Physical Review D*, 98(4):042006, 2018. 110

[87] I. Ferrero, M. Crocce, I. Tutusaus, A. Porredon, L. Blot, P. Fosalba, A. Carnero Rosell, S. Avila, A. Izard, J. Elvin-Poole, K. C. Chan, H. Camacho, R. Rosenfeld, E. Sanchez, P. Tallada-Crespí, J. Carretero, I. Sevilla-Noarbe, E. Gaztanaga, F. Andrade-Oliveira, J. De Vicente, J. Mena-Fernández, A. J. Ross, et al. Dark Energy Survey Year 3 Results: Galaxy mock catalogs for BAO analysis. *A&A*, 656:A106, December 2021. 112, 113

[88] Albert Izard, Martin Crocce, and Pablo Fosalba. ICE-COLA: towards fast and accurate synthetic galaxy catalogues optimizing a quasi-N-body method. *Monthly Notices of the Royal Astronomical Society*, 459(3):2327–2341, 2016. 112

[89] Svetlin Tassev, Matias Zaldarriaga, and Daniel J Eisenstein. Solving large scale structure in ten easy steps with COLA. *Journal of Cosmology and Astroparticle Physics*, 2013(06):036, 2013. 112

[90] Albert Izard, Pablo Fosalba, and Martin Crocce. ICE-COLA: fast simulations for weak lensing observables. *Monthly Notices of the Royal Astronomical Society*, 473(3):3051–3061, 2018. 112

[91] P Fosalba, M Crocce, E Gaztañaga, and FJ Castander. The MICE grand challenge lightcone simulation–I. Dark matter clustering. *Monthly Notices of the Royal Astronomical Society*, 448(4):2987–3000, 2015. 113

[92] P Fosalba, E Gaztañaga, FJ Castander, and M Crocce. The MICE Grand Challenge light-cone simulation–III. Galaxy lensing mocks from all-sky lensing maps. *Monthly Notices of the Royal Astronomical Society*, 447(2):1319–1332, 2015. 113

[93] Eiichiro Komatsu, J Dunkley, MR Nolta, CL Bennett, B Gold, G Hinshaw, N Jarosik, D Larson, M Limon, L Page, et al. Five-year wilkinson microwave anisotropy probe* observations: cosmological interpretation. *The Astrophysical Journal Supplement Series*, 180(2):330, 2009. 113, 119

[94] Santiago Avila, Martin Crocce, AJ Ross, J García-Bellido, WJ Percival, N Banik, H Camacho, N Kokron, KC Chan, F Andrade-Oliveira, et al. Dark Energy Survey Year-1 results: galaxy mock catalogues for BAO. *Monthly Notices of the Royal Astronomical Society*, 479(1):94–110, 2018. 113

[95] TMC Abbott, FB Abdalla, A Alarcon, S Allam, F Andrade-Oliveira, J Annis, S Avila, Mandakranta Banerji, N Banik, K Bechtol, et al. Dark Energy Survey Year 1 Results: Measurement of the Baryon Acoustic Oscillation scale in the distribution of galaxies to redshift 1. *Monthly Notices of the Royal Astronomical Society*, 483(4):4866–4883, 2019. 114

[96] Kuei Cheng Chan, Martín Crocce, AJ Ross, S Avila, J Elvin-Poole, M Manera, Will J Percival, R Rosenfeld, TMC Abbott, FB Abdalla, et al. BAO from angular clustering: optimization and mitigation of theoretical systematics. *Monthly Notices of the Royal Astronomical Society*, 480(3):3031–3051, 2018. 116, 117

[97] Glen Cowan. *Statistical data analysis*. Oxford university press, 1998. 116

[98] MICHAEL Lampton, BRUCE Margon, and STUART Bowyer. Parameter estimation in X-ray astronomy. *The Astrophysical Journal*, 208:177–190, 1976. 117

[99] Alan Heavens. Statistical techniques in cosmology. *arXiv preprint arXiv:0906.0664*, 2009. 117

[100] Antony Lewis and Anthony Challinor. Code for anisotropies in the microwave background. *Documentation, January. http://camb. info*, 2017. 117

[101] Daniel J Eisenstein and Wayne Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605, 1998. 118

[102] Zvonimir Vlah, Uroš Seljak, Man Yat Chu, and Yu Feng. Perturbation theory, effective field theory, and oscillations in the power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2016(03):057, 2016. 118

[103] Mikhail M Ivanov and Sergey Sibiryakov. Infrared resummation for biased tracers in redshift space. *Journal of Cosmology and Astroparticle Physics*, 2018(07):053, 2018. 118

[104] Mikhail M Ivanov, Marko Simonović, and Matias Zaldarriaga. Cosmological parameters from the BOSS galaxy power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2020(05):042, 2020. 118

[105] Benjamin Wallisch. *Cosmological probes of light relics*. Springer Nature, 2019. 119

[106] Elisabeth Krause and Tim Eifler. COSMOLIKE–cosmological likelihood analyses for photometric galaxy surveys. *Monthly Notices of the Royal Astronomical Society*, 470(2):2100–2112, 2017. 122

[107] Xiao Fang, Tim Eifler, and Elisabeth Krause. 2D-FFTLog: efficient computation of real-space covariance matrices for galaxy clustering and weak lensing. *Monthly Notices of the Royal Astronomical Society*, 497(3):2699–2714, 2020. 122

[108] Xiao Fang, Elisabeth Krause, Tim Eifler, and Niall MacCrann. Beyond Limber: Efficient computation of angular power spectra for galaxy clustering and weak lensing. *Journal of Cosmology and Astroparticle Physics*, 2020(05):010, 2020. 122

[109] Martín Crocce and Román Scoccimarro. Nonlinear evolution of baryon acoustic oscillations. *Physical Review D*, 77(2):023533, 2008. 130

[110] Mike Jarvis. TreeCorr: Two-point correlation functions. *Astrophysics Source Code Library*, pages ascl–1508, 2015. 136

[111] Nick Hand, Yu Feng, Florian Beutler, Yin Li, Chirag Modi, Uroš Seljak, and Zachary Slepian. nbodykit: An open-source, massively parallel toolkit for large-scale structure. *The Astronomical Journal*, 156(4):160, 2018. 138

[112] Shadab Alam, Metin Ata, Stephen Bailey, Florian Beutler, Dmitry Bizyaev, Jonathan A Blazek, Adam S Bolton, Joel R Brownstein, Angela Burden, Chia-Hsun Chuang, et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 470(3):2617–2652, 2017. 141

[113] Amélie Tamone, Anand Raichoor, Cheng Zhao, Arnaud de Mattia, Claudio Gorgoni, Etienne Burtin, Vanina Ruhlmann-Kleider, Ashley J Ross, Shadab Alam, Will J Percival, et al. The completed SDSS-IV extended baryon oscillation spectroscopic survey: growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the emission-line galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 499(4):5527–5546, 2020. 141

[114] Daniel Foreman-Mackey, David W Hogg, Dustin Lang, and Jonathan Goodman. emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013. 143

[115] Antony Lewis. GetDist: a Python package for analysing Monte Carlo samples. *arXiv preprint arXiv:1910.13970*, 2019. 143

[116] Adam G Riess, Stefano Casertano, Wenlong Yuan, Lucas M Macri, and Dan Scolnic. Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM. *The Astrophysical Journal*, 876(1):85, 2019. 144

[117] Ryan J Cooke, Max Pettini, and Charles C Steidel. One percent determination of the primordial deuterium abundance. *The Astrophysical Journal*, 855(2):102, 2018. 144

[118] Eric G Adelberger, A García, RG Hamish Robertson, KA Snover, AB Balantekin, K Heeger, MJ Ramsey-Musolf, D Bemmerer, A Junghans, CA Bertulani, et al. Solar fusion cross sections. II. The p p chain and CNO cycles. *Reviews of Modern Physics*, 83(1):195, 2011. 144

[119] TMC Abbott, S Allam, P Andersen, Charlotte Angus, J Asorey, A Avelino, S Avila, BA Bassett, K Bechtol, GM Bernstein, et al. First cosmology results using type Ia supernovae from the dark energy survey: constraints on cosmological parameters. *The Astrophysical Journal Letters*, 872(2):L30, 2019. 149

[120] Joe Zuntz, Marc Paterno, Elise Jennings, Douglas Rudd, Alessandro Manzotti, Scott Dodelson, Sarah Bridle, Saba Sehrish, and James Kowalkowski. CosmoSIS: modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59, 2015. 149

[121] Hee-Jong Seo and Daniel J Eisenstein. Improved forecasts for the baryon acoustic oscillations and cosmological distance scale. *The Astrophysical Journal*, 665(1):14, 2007. 159, 160

[122] Max Tegmark. Measuring cosmological parameters with galaxy surveys. *Physical Review Letters*, 79(20):3806, 1997. 159

[123] Ross Cawthon, J Elvin-Poole, A Porredon, M Crocce, G Giannini, M Gatti, AJ Ross, ES Rykoff, A Carnero Rosell, J DeRose, et al. Dark Energy Survey Year 3 results: calibration of lens sample redshift distributions using clustering redshifts with BOSS/eBOSS. *Monthly Notices of the Royal Astronomical Society*, 513(4):5517–5539, 2022. 171

[124] Peter Coles and Bernard Jones. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society*, 248(1):1–13, 1991. 177

[125] Lucinda Clerkin, Donnacha Kirk, M Manera, O Lahav, F Abdalla, Adam Amara, D Bacon, C Chang, E Gaztañaga, A Hawken, et al. Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps. *Monthly Notices of the Royal Astronomical Society*, 466(2):1444–1461, 2017. 177

[126] Risa H Wechsler and Jeremy L Tinker. The connection between galaxies and their dark matter halos. *Annual Review of Astronomy and Astrophysics*, 56:435–487, 2018. 189, 190, 191, 192

[127] Asantha Cooray and Ravi Sheth. Halo models of large scale structure. *Physics reports*, 372(1):1–129, 2002. 190

[128] Andreas A Berlind and David H Weinberg. The halo occupation distribution: Toward an empirical determination of the relation between galaxies and mass. *The Astrophysical Journal*, 575(2):587, 2002. 190

[129] Andreas A Berlind, David H Weinberg, Andrew J Benson, Carlton M Baugh, Shaun Cole, Romeel Davé, Carlos S Frenk, Adrian Jenkins, Neal Katz, and Cedric G Lacey. The halo occupation distribution and the physics of galaxy formation. *The Astrophysical Journal*, 593(1):1, 2003. 190, 191, 192

[130] Antonio Vale and JP Ostriker. Linking halo mass to galaxy luminosity. *Monthly Notices of the Royal Astronomical Society*, 353(1):189–200, 2004. 190

[131] F Shankar, A Lapi, P Salucci, G De Zotti, and L Danese. New relationships between galaxy properties and host halo mass, and the role of feedbacks in galaxy formation. *The Astrophysical Journal*, 643(1):14, 2006. 190

[132] Ryan Hausen, Brant E Robertson, Hanjue Zhu, Nickolay Y Gnedin, Piero Madau, Evan E Schneider, Bruno Villasenor, and Nicole E Drakos. Revealing the Galaxy–Halo Connection through Machine Learning. *The Astrophysical Journal*, 945(2):122, 2023. 190

[133] Charlie Conroy and Risa H Wechsler. Connecting galaxies, halos, and star formation rates across cosmic time. *The Astrophysical Journal*, 696(1):620, 2009. 190

[134] Alex Smith, Etienne Burtin, Jiamin Hou, Richard Neveux, Ashley J Ross, Shadab Alam, Jonathan Brinkmann, Kyle S Dawson, Salman Habib, Katrin Heitmann, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body mock challenge for the quasar sample. *Monthly Notices of the Royal Astronomical Society*, 499(1):269–291, 2020. 191

[135] Graziano Rossi, Peter D Choi, Jeongin Moon, Julian E Bautista, Hector Gil-Marín, Romain Paviot, Mariana Vargas-Magana, Sylvain de La Torre, Sebastien Fromenteau, Ashley J Ross, et al. The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: N-body mock challenge for galaxy clustering measurements. *Monthly Notices of the Royal Astronomical Society*, 505(1):377–407, 2021. 191

[136] Idit Zehavi, Zheng Zheng, David H Weinberg, Michael R Blanton, Neta A Bahcall, Andreas A Berlind, Jon Brinkmann, Joshua A Frieman, James E Gunn, Robert H Lupton, et al. Galaxy clustering in the completed SDSS redshift survey: the dependence on color and luminosity. *The Astrophysical Journal*, 736(1):59, 2011. 191

[137] John K Parejko, Tomomi Sunayama, Nikhil Padmanabhan, David A Wake, Andreas A Berlind, Dmitry Bizyaev, Michael Blanton, Adam S Bolton, Frank van den Bosch, Jon Brinkmann, et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: the low-redshift sample. *Monthly Notices of the Royal Astronomical Society*, 429(1):98–112, 2013. 191

[138] Johannes U Lange, Andrew P Hearin, Alexie Leauthaud, Frank C van den Bosch, Hong Guo, and Joseph DeRose. Five per cent measurements of the growth rate from simulation-based modelling of redshift-space clustering in BOSS LOWZ. *Monthly Notices of the Royal Astronomical Society*, 509(2):1779–1804, 2022. 191

[139] Yosuke Kobayashi, Takahiro Nishimichi, Masahiro Takada, and Hironao Miyatake. Full-shape cosmology analysis of the SDSS-III BOSS galaxy power spectrum using an emulator-based halo model: A 5% determination of σ_8. *Physical Review D*, 105(8):083517, 2022. 191

[140] Zheng Zheng, Alison L Coil, and Idit Zehavi. Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering. *The Astrophysical Journal*, 667(2):760, 2007. 192

[141] Sihan Yuan, Lehman H Garrison, Boryana Hadzhiyska, Sownak Bose, and Daniel J Eisenstein. ABACUSHOD: a highly efficient extended multitracer HOD framework and its application to BOSS and eBOSS data. *Monthly Notices of the Royal Astronomical Society*, 510(3):3301–3320, 2022. 192

[142] Lehman H Garrison, Daniel J Eisenstein, Douglas Ferrer, Nina A Maksimova, and Philip A Pinto. The ABACUS cosmological N-body code. *Monthly Notices of the Royal Astronomical Society*, 508(1):575–596, 2021. 193

[143] Nina A Maksimova, Lehman H Garrison, Daniel J Eisenstein, Boryana Hadzhiyska, Sownak Bose, and Thomas P Satterthwaite. ABACUSSUMMIT: a massive set of high-accuracy, high-resolution N-body simulations. *Monthly Notices of the Royal Astronomical Society*, 508(3):4017–4037, 2021. 193

[144] Martin White, Jeremy L Tinker, and Cameron K McBride. Mock galaxy catalogues using the quick particle mesh method. *Monthly Notices of the Royal Astronomical Society*, 437(3):2594–2606, 2014. 193

[145] Angela Burden, Will J Percival, and Cullan Howlett. Reconstruction in Fourier space. *Monthly Notices of the Royal Astronomical Society*, 453(1):456–468, 2015. 195

[146] Daniel J Eisenstein, Hee-Jong Seo, Edwin Sirko, and David N Spergel. Improving cosmological distance measurements by reconstruction of the baryon acoustic peak. *The Astrophysical Journal*, 664(2):675, 2007. 195

[147] Samuel R Hinton, Cullan Howlett, and Tamara M Davis. BARRY and the BAO model comparison. *Monthly Notices of the Royal Astronomical Society*, 493(3):4078–4093, 2020. 199

[148] Oliver HE Philcox and Daniel J Eisenstein. Estimating covariance matrices for two-and three-point correlation function moments in Arbitrary Survey Geometries. *Monthly Notices of the Royal Astronomical Society*, 490(4):5931–5951, 2019. 200

[149] Oliver HE Philcox, Daniel J Eisenstein, Ross O'Connell, and Alexander Wiegand. RascalC: a jackknife approach to estimating single-and multitracer galaxy covariance matrices. *Monthly Notices of the Royal Astronomical Society*, 491(3):3290–3317, 2020. 200

[150] Digvijay Wadekar and Roman Scoccimarro. Galaxy power spectrum multipoles covariance in perturbation theory. *Physical Review D*, 102(12):123517, 2020. 200

[151] DESI Collaboration, Amir Aghamousa, Jessica Aguilar, Steve Ahlen, Shadab Alam, Lori E Allen, Carlos Allende Prieto, James Annis, Stephen Bailey, Christophe Balland, et al. The DESI Experiment part I: Science, targeting, and survey design. *arXiv preprint arXiv:1611.00036*, 2016. 207