Cosmological Implications of Galaxy Clustering in BOSS and eBOSS

Agnė Semėnaitė

München 2023

Cosmological Implications of Galaxy Clustering in BOSS and eBOSS

Agnė Semėnaitė

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Agnė Semėnaitė
aus Vilnius

München, den 24.08.2023

Erstgutachter: Priv. Doz. Dr. Ariel G. Sánchez
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 17.10.2023

Seneliui
(for my grandpa)

"*Nevermind the distorted light, gravity wins!*"
Jiamin Hou

Contents

List of Figures

List of Tables

Zusammenfassung

Die Kosmologie befindet sich derzeit im Wandel von Experimenten, die das kosmologische Standardmodell etabliert haben, zu einer neuen Generation von Experimenten, die etwaige Abweichungen von diesem Modell aufzudecken versuchen, sofern diese existieren. Die aktuellen Daten bestätigen weitgehend, dass wir in einem flachen Universum leben, dominiert von dunkler Energie, die die beschleunigte Ausdehnung antreibt, und von dunkler Materie, der wichtigsten Gravitationskomponente, die eine netzartige großräumige Struktur bildet. Dennoch gibt es Hinweise darauf, dass die von verschiedenen kosmologischen Sonden präzise gemessenen Werte der Modellparameter dieses ΛCDM-Modells Diskrepanzen aufweisen. Um die Ursache dieser Diskrepanzen zu untersuchen, müssen größere Datensätze gesammelt werden und Daten von verschiedenen kosmologischen Sonden in einheitlichen Analysen kombiniert werden. Beide Wege erfordern neue Werkzeuge, die es ermöglichen größere und komplexere Datensätze zu verarbeiten, wobei beachtet werden muss, dass die Verwendung von Methoden und Annahmen, die ursprünglich für viel einfachere Analysen entwickelt wurden, zu Einschränkungen und systematischen Fehlern führen können.

In dieser Dissertation wird eine Verbesserung der Vollformanalyse von anisotropen Zweipunkt-Galaxien-Clusterstatistiken vorgestellt. Wir stützen uns dabei stark auf das Konzept des Evolution Mappings und konzentrieren uns auf zwei Aspekte der Analyse: die Modellierungswerkzeuge und die Interpretation der Ergebnisse. Das Ziel der Analysen ist eine Einschränkung des Raumes der physikalischen Parameter, anhand deren die Ergebnisse interpretiert werden. Unter anderem sind wir an kosmologischen Parametern interessiert, die nicht durch den dimensionslosen Hubble-Parameter h definiert sind: wir definieren den Parameter für die Cluster-Amplitude, der die lineare Dichtefeldvarianz auf der Skala $8\,h^{-1}$Mpc misst, σ_8, neu und verwenden stattdessen sein Äquivalent, das auf der Skala von 12 Mpc gemessen wird, σ_{12}. Außerdem verwenden wir die physikalischen Dichten ω_i der kosmologischen Energiespezies i, die sich auf die relativen Dichten als $\omega_i = \Omega_i h^2$ beziehen. Im Hinblick auf die Entwicklung von Modellierungswerkzeugen bieten wir einen schnellen Weg, um Vorhersagen über die Vollform-Galaxienhäufung zu erhalten, indem wir die Beziehungen zwischen den kosmologischen Parametern in physikalischen Parameterraum nutzen, die einen Einfluss auf das Galaxien-Leistungsspektrum haben. Die hier vorgestellten Analysen nutzen die Messungen der BOSS- und eBOSS-Galaxiendurchmusterungen, die beide Teil des größten öffentlich zugänglichen spektroskopischen Datensatzes sind - des Sloan Digital Sky Survey (SDSS).

Im ersten Teil der Arbeit betrachten wir das ΛCDM-Modell und führen die erste Vollformanalyse der eBOSS-Quasar-Clustermessungen durch, welche wir mit neu analysierten Galaxien-Clusterkeilen aus BOSS kombinieren. Wir vergleichen und kombinieren die resultierenden Ergebnisse mit denen aus 3×2pt Messungen von DES Y1 und Temperatur- und Polarisationsleistungsspektren des kosmischen Mikrowellenhintergrundes (CMB) von *Planck*. Wir zeigen, dass alle betrachteten Datensätze eine konsistente Clustering-Amplitude σ_{12} ergeben. Ferner zeigen wir, dass der physikalische Parameterraum es ermöglicht, die enge Beziehung zwischen den anfänglichen und endgültigen Amplituden des Leistungsspektrums wiederherzustellen, und stellen mithilfe dieses Ergebnisses fest, dass die DES Y1-Messungen allein oder in Kombination mit der Galaxienhäufung eine Präferenz für ein geringeres Wachstum der Gesamtstruktur zeigen als von *Planck* vorhergesagt. Wir weisen nach, dass dieser Unterschied mit dem Diskrepanzniveau des Lensing-Amplitudenparameters übereinstimmt und mit den unterschiedlichen Werten der dunklen Energiedichte von Messungen bei niedriger und hoher Rotverschiebung zusammenhängt.

Im zweiten Teil der Arbeit dehnen wir unsere Analyse auf Modelle jenseits des ΛCDM-Modells aus, mit besonderem Fokus auf Modellen mit variierenden Zustandsgleichungsparametern der dunklen Energie w, und zeigen, dass alle durch Galaxienhäufung bestimmte Parameterintervalle mit den Werten des Standardmodells konsistent sind. Darüber hinaus demonstrieren wir, dass der CMB in der Lage ist, die Amplitude der Galaxienhäufung auch in Modellen mit variierendem w einzuschränken, wenn sie auf einer physikalischen Skala, definiert in Mpc, gemessen wird. Wir liefern das erste CMB-Konfidenzintervall für σ_{12} in solchen Kosmologien, welches mit dem äquivalenten Intervall im ΛCDM-Modell übereinstimmt, wenn auch etwas höher als dieses liegt. Weiter bestätigen wir *Planck*s Präferenz für ein geschlossenes Universum mit der bisher größten Signifikanz von 4σ.

Der letzte Teil der Arbeit ist der Konstruktion eines auf Evolution Mapping basierenden Zweipunkt-Korrelationsfunktionsemulators gewidmet. Dieser wird als Erweiterung des bestehenden Fourierraum-Emulators `COMET` implementiert und basiert auf dem EFT-Modell für das nichtlineare Leistungsspektrum. Wir demonstrieren, dass die Idee des Evolution Mappings auch auf Vorhersagen im Konfigurationsraum angewandt werden kann, wobei der resultierende Emulator eine große Flexibilität bezüglich der Vielfalt der kosmologischen Modelle und der gewählten Referenzkosmologie für einen Bereich von Rotverschiebungen im realen und Rotverschiebungsraum aufweist. Wir bestätigen, dass der Emulator theoretische Vorhersagen für Galaxienhäufungsmultipole auf Skalen von $20\,\mathrm{Mpc} < s < 250\,\mathrm{Mpc}$ frei von systematischen Fehlern liefert. Unsere `COMET`-Erweiterung bietet eine große Beschleunigung für aktuelle und zukünftige kosmologische Analysen und wird es ermöglichen, verschiedene Modellannahmen in Vorbereitung auf die kommenden Galaxien-Rotverschiebungsdurchmusterungen einfacher zu testen.

Abstract

Cosmology is currently in the transitional period between the observations that established the standard cosmological model and the new generation of experiments that aim to uncover any deviations from it, should they exist. Current data largely confirms that we live in a flat Universe, dominated by dark energy, which fuels the accelerated expansion, and dark matter, which is the major gravitating component that makes up the web-like large-scale structure. Nevertheless, the exact values of the model parameters measured by different cosmological probes appear to be in tension. The most immediate way forward for investigating the origins of the emerging tensions relies on obtaining larger datasets as well as combining the cosmological probes in unified analyses. Both of these routes require that we develop new tools capable of processing larger and more complex data while being careful about the biases and the limitations that may be introduced by using the methods and the assumptions first developed for much simpler analyses.

This thesis presents work done towards updating the full-shape analysis of anisotropic two-point galaxy clustering statistics. We rely heavily on the evolution mapping concept and focus on two aspects of the analysis: the modelling tools and the result interpretation. On the side of the result interpretation, all of our analyses focus on obtaining constraints on the physical parameter space - i.e., we are interested in cosmological parameters that are not defined through the dimensionless Hubble parameter h: we redefine the clustering amplitude parameter that measures linear density field variance on the scale of $8\,h^{-1}$Mpc, σ_8, and use its equivalent measured on the scale of 12 Mpc instead, σ_{12}. Additionally, we use physical densities ω_i of cosmological energy species i, which relate to relative densities as $\omega_i = \Omega_i h^2$. In terms of developing modelling tools, we provide a quick way to obtain full-shape galaxy clustering predictions by making use of the degeneracies that are followed by the cosmological parameters that affect the amplitude of the power spectrum, as long as the physical parameter space is used. The analyses presented here make use of the measurements obtained from BOSS and eBOSS galaxy surveys which are both part of the largest publicly available spectroscopic dataset - the Sloan Digital Sky Survey (SDSS).

In the first part of the thesis, we consider ΛCDM and perform the first full shape analysis of eBOSS quasar clustering measurements, combining them with re-analysed galaxy clustering wedges from BOSS. We compare and combine the resulting constraints with those from 3×2pt measurements of DES Y1 and CMB temperature and polarisation power spectra measurements by *Planck*. We show that all datasets considered recover consistent clustering amplitude σ_{12}. We also demonstrate that physical parameter space allows

us to recover the tight degeneracy between the initial and final power spectrum amplitudes and use this result to determine that DES Y1 measurements alone or in combination with galaxy clustering show a preference for less total structure growth than predicted by *Planck*. We show that this difference is consistent with the level of tension in the lensing amplitude parameter and is related to different values of dark energy density preferred by the low- and high-redshift measurements.

In the second part of the thesis, we extend our analysis to beyond ΛCDM models, with a particular interest in models with varying dark energy equation of state parameter w, and show that all extended parameter space constraints obtained by galaxy clustering are consistent with the standard model values. We additionally demonstrate that CMB is able to constrain the clustering amplitude today even in models with varying w, as long as it is measured on a physical scale, defined in Mpc. We provide the first CMB constraint on σ_{12} in such cosmologies and show that it is consistent with, albeit slightly higher than, the equivalent constraint within ΛCDM. We additionally confirm *Planck*'s preference for a closed Universe at the greatest significance yet - 4σ.

The last part of the thesis is dedicated to the construction of an evolution-mapping-based two-point correlation function emulator. This is built as an extension to an existing Fourier space emulator `COMET` and is based on the EFT model for the nonlinear power spectrum. We show that the evolution mapping ideas can also be applied to obtain predictions in configuration space and that the resulting emulator demonstrates great flexibility in terms of the variety of cosmological models covered as well as its ability to obtain predictions for an arbitrary fiducial cosmology for a continuous range of redshifts in real and redshift spaces. We validate that the emulator is able to provide unbiased theory predictions for galaxy clustering multipoles at scales of $20\,\mathrm{Mpc} < s < 250\,\mathrm{Mpc}$. Our `COMET` extension provides great speedup for current and future cosmological analyses and will subsequently allow us to more easily test different model assumptions in preparation for the upcoming galaxy redshift surveys.

Chapter 1

Introduction

During the last hundred or so years, cosmologists have managed to build the theories and design the observations required to conduct scientific, quantitative experiments on the most ambitious of targets - the Universe itself. It is truly remarkable that in such a short period of time, the community has progressed from disputing whether the Universe extends beyond the Milky Way, to being able to probe the constituent energy species through the analyses of three-dimensional galaxy maps, the observations of the earliest light in the Universe or the measurements of the slight distortion of galaxy shapes, as their light passes through the gravitational potential of all the structure in its way. By detecting excess galaxy clustering, we have adopted as a standard ruler the signatures of physics of the Universe in its infancy, long before the existence of any galaxy or a star. Cosmology today, nonetheless, finds itself in a rather curious situation: the standard cosmological model, the crowning result of the increasingly ambitious and precise observational campaigns, can fit a number of vastly different data sets incredibly well. However, the physical nature of some of its key components remains unexplained. Moreover, modern observations have become so powerful that the small mismatches in the best-fit parameters of the model yielded by different analyses are emerging as increasingly more significant, pointing towards our ignorance of the systematics associated with the data or, potentially more intriguingly, the incompleteness of the standard model itself.

The standard cosmological model, ΛCDM, has come to be closely associated with the cosmic microwave background (CMB) experiments. The Cosmic Background Explorer (COBE) was the first one to confirm that the CMB spectrum matched that of a blackbody, confirming Big Bang as the origin theory of the Universe (Mather et al., 1994). This early observation already contained evidence of many of the main ingredients of the ΛCDM model. According to ΛCDM, shortly after the Big Bang takes place, the Universe starts out with a phase of rapid expansion, called inflation, during which the initial quantum fluctuations get stretched out to produce small-scale inhomogeneities in the density field. These inhomogeneities will continue to grow as the Universe expands and cools. Once inflation ends and particles form, baryons and photons are coupled to form a baryon-photon plasma that undergoes pressure oscillations (baryon acoustic oscillations, BAO). CMB itself originates when the Universe cools down enough that photons can decouple

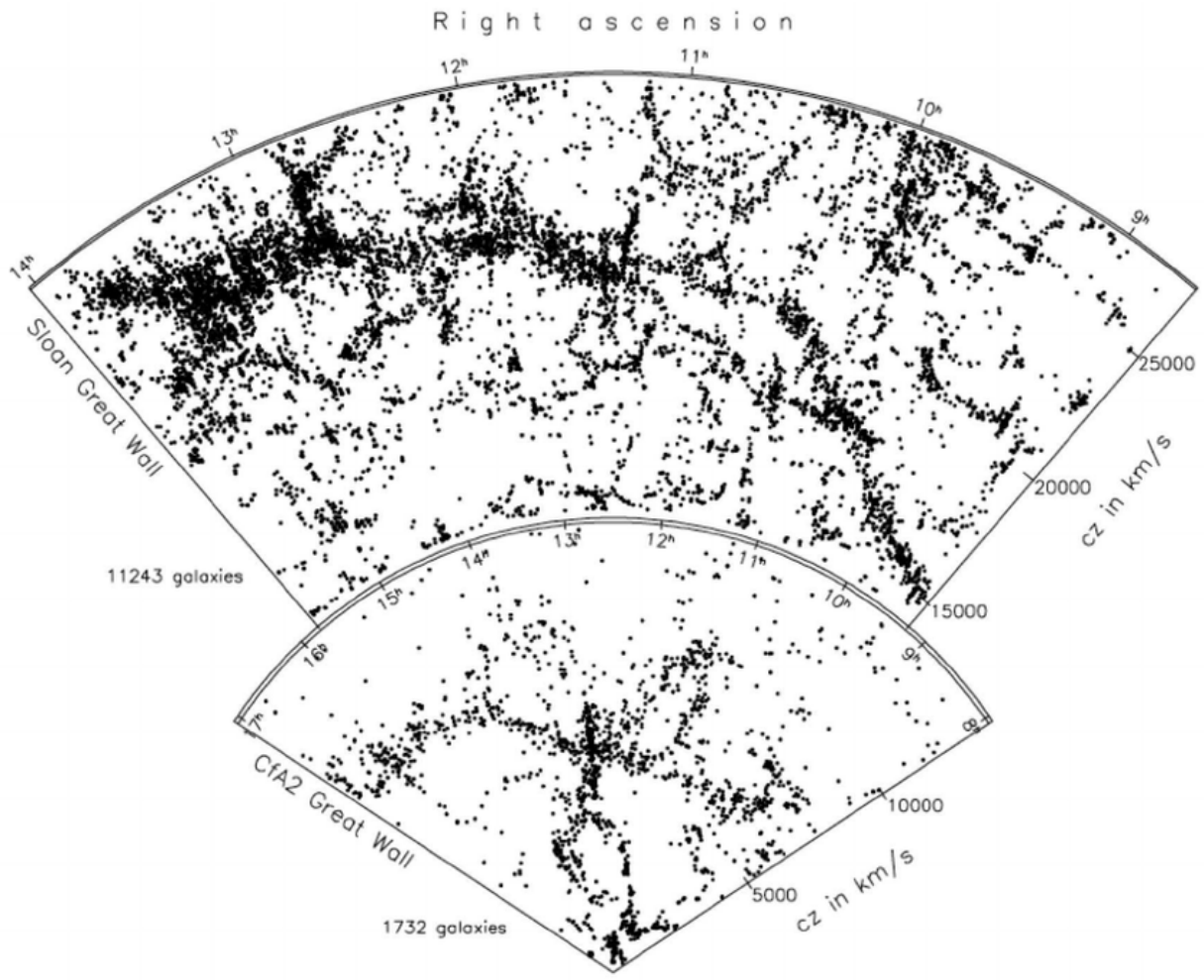

Figure 1.1: Sloan Great Wall compared to CfA2 Great Wall at the same scale in comoving coordinates. Both slices are approximately the same physical width at the two walls. The figure illustrates some of the first superstructures observed by galaxy redshift surveys. Image credit: Gott et al. (2005).

from baryons and travel freely to reach us today. However, the signature of their shared past with baryons remains imprinted in CMB as an oscillatory feature in the temperature power spectrum. The amplitude of this feature furthermore provides confirmation for a component that interacts through gravitational force only - dark matter. While COBE's resolution was not high enough to resolve the acoustic peaks, it was the first experiment to detect that the CMB temperature spectrum did display the small-scale anisotropies expected. COBE's successors were able to not only provide high-resolution measurements of the temperature power spectrum but also observe CMB polarization (first detected by the Degree Angular Scale Interferometer, DASI, Kovac et al., 2002) with one of the latest CMB experiments by the *Planck* satellite performing the state-of-the-art observations that resulted in the most precise complete set of cosmological parameter constraints that are available today (Planck Collaboration et al., 2020).

Despite the great success of CMB observations, cosmology has always relied on multiple probes to complete and confirm the full picture. According to ΛCDM, the cold dark matter and baryons are believed to make up around a third of the total energy budget with the rest of the contribution provided by another mysterious component which is responsible for the accelerated expansion of the Universe - dark energy (represented by the cosmological constant Λ). Here, the crucial observations were those of individual distant stars and galaxies: the first indication that distant objects were receding from us was obtained by Slipher in 1917 who measured the spectra of a sample of galaxies to find that the majority of them were redshifted (Slipher, 1917). It took until the end of the century until the Type Ia supernovae observations showed that the Universe was not only expanding, but was, in fact, doing so in an accelerated manner (Perlmutter et al., 1999; Riess et al., 1998).

When Slipher first observed the redshifted galaxy spectra, his sample consisted of just fifteen galaxies. Starting in 1977, the first systematic galaxy redshift survey - The Center for Astrophysics Redshift Survey (CfA, Huchra et al., 1983), obtained the redshifts of around 2,200 galaxies in its first phase and 18,000 galaxies in the second phase. These observations unravelled a complex pattern of galaxy clustering, demonstrating that galaxies are distributed inhomogeneously (an artefact of quantum fluctuations stretched out during inflation), forming such massive structures as, for example, 'The Great Wall' - one of the galaxy filaments observed by CfA2 which stretches for hundreds of millions of lightyears, making it one of the largest known superstructures (see Figure 1.1). Subsequent galaxy redshift surveys have been providing key evidence in support of the ΛCDM model ever since. In Tadros et al. (1999) one of the first measurements of the anisotropic distortion in galaxy clustering pattern due to galaxy peculiar velocities (an effect known as the redshift space distortions or RSD) was carried out on the Point Source Catalogue (PSCz, Saunders et al., 1999). This work already includes an attempt to distinguish between different cosmological models. While this early analysis remained inconclusive, it was already noted that the recovered value of the distortion parameter (which is proportional to the relative matter density) was low, in line with the supernovae observations, supporting the idea of the presence of an additional energy species. The spectroscopic galaxy surveys that followed allowed for cosmological analyses that more closely resemble the ones carried out today: The Two-degree-Field-Galaxy Redshift Survey (2dF-GRS Colless et al., 2001)

already contained redshifts for $\sim 250,000$ galaxies and provided many important results. By fitting the power spectra of the data Cole et al. (2005) obtained a measurement of the matter content of the Universe, while Efstathiou et al. (2002) carried out a joint analysis with the CMB observations to confirm the non-zero cosmological constant, independently of the Type Ia supernovae. Lahav et al. (2002) combined galaxy clustering measurements with CMB data to provide a measurement of galaxy bias and further cosmological constraints were obtained by Sánchez et al. (2006).

The Sloan Digital Sky Survey (SDSS, York et al., 2000), began at the start of the millenium to become the largest galaxy redshift survey at that time. During its first two phases in the first eight years, SDSS imaged over a million galaxies and 100,000 quasars, of which 260,490 were spectroscopically confirmed. Figure 1.1 shows a slice from the early SDSS data, displaying a massive filament: building on the developments by the early galaxy redshift surveys, SDSS was able to produce bigger and more detailed maps of the large-scale structure. These observations led to the first detection of the BAO signal in the configuration space by Eisenstein et al. (2005). Almost simultaneously, the BAO signal was also confirmed in the power spectrum of galaxy clustering in 2dF-GRS (Cole et al., 2005). The final two stages of SDSS included the carrying out and completion of Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al., 2013) and extended Baryon Oscillation Spectroscopic Survey (eBOSS, Dawson et al., 2016). In addition to providing spectroscopy for more than 1.5 million galaxies up to the redshift of $z = 0.6$, BOSS also measured the distribution of quasar absorption lines for more than 150,000 quasars. Impressively, BOSS was able to obtain 1% measurements of angular diameter distance from its galaxy sample. eBOSS was designed as an extension of BOSS. It consists of spectroscopic measurements of just under half a million galaxies spanning the redshift range of $0.6 < z < 1.1$ and almost 350,000 quasars in the redshift range of $0.8 < z < 2.2$. The latter sample led to the first detection of BAO using quasars as tracers (Ata et al., 2018). The complete SDSS reveals an impressive three-dimensional galaxy map, such that the full spectroscopic sample provides precise measurements of the expansion history of the Universe throughout eighty percent of cosmic history.

The improvement in the constraining power of all cosmological probes has directed cosmologists' attention towards verifying the consistency in the recovered cosmological parameter constraints. It was soon revealed that, while no probe showed a preference for an alternative to the ΛCDM model, some tensions emerged between the model parameters as measured by different probes. The most significant discrepancy concerns the measurement of the expansion rate of the Universe today and is referred to as the 'H_0 tension'. The tension can be summarised as a mismatch between the local direct and indirect measurements of H_0, with direct probes preferring higher values (although the full picture is somewhat more complex than this, see Figure 1.2), with the greatest discrepancy reaching 5σ between the H_0 value predicted from CMB measurements by *Planck* and the one measured from Cepheid calibrated Supernovae Ia observations (Riess et al., 2021).

In addition to the H_0 tension, the advent of another large-scale structure probe - weak lensing, has brought another tension, this time in the parameter describing the amplitude of the weak lensing signal, S_8 (a combination of the relative matter density parameter

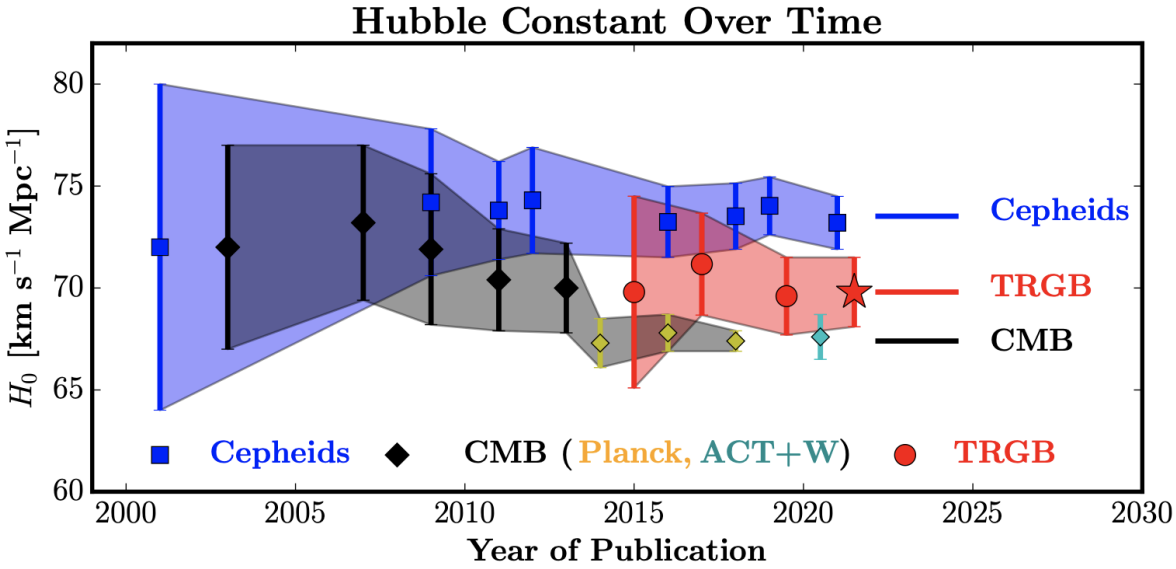

Figure 1.2: Hubble constant values over time, based on Cepheid variables (blue), the tip of the red giant branch (TRGB, red) and CMB (black). The CMB constraints represent a model prediction of H_0 assuming flat ΛCDM cosmology. Cepheids and TRGB refer to two different methods used to calibrate the Supernovae Ia observations to produce a direct measurement of H_0. The TRGB-based result shows agreement with both CMB and Cepheids values of H_0, indicating that the level of tension between Supernovae Ia and CMB constraints can depend on the chosen calibration method. The figure also demonstrates the significant improvement in the constraining power achieved by all probes. Image credit: Freedman (2021).

$\Omega_{\rm m}$ and a parameter describing linear density field variance at the scale of $8h^{-1}$Mpc, σ_8). Weak lensing analyses measure the distortion in the observed galaxy shape induced by the gravitational potential of the structure in between the source galaxies and the observer. The resulting effect is so small that weak lensing surveys require an order of magnitude greater number of galaxies than available in typical spectroscopic samples in order to obtain a statistically significant signal. As a result, weak lensing surveys rely on cheaper photometric redshifts. The great advantage of weak lensing measurements is that the observed effect is sensitive to the total matter distribution: i.e., weak lensing probes the full matter density field in an unbiased manner. Already the early weak lensing surveys displayed inconsistencies with *Planck* predictions with the first evidence of tension in the lensing signal amplitude observed in Canada–France–Hawaii Telescope Lensing Survey (CFHTLenS, Heymans et al., 2012), as described in MacCrann et al. (2015). While a number of subsequent weak lensing surveys did not report significant discrepancies (see, for example, the most recent result by The Dark Energy Survey Collaboration et al., 2021), all of the lensing amplitude constraints so far have yielded lower measurements than predicted by *Planck*. Meanwhile, the most significant tension was reported by Heymans

et al. (2021) who found a mismatch at the level of 3σ.

In light of these tensions, galaxy clustering stands out as a probe distinctly placed to provide more insight: in addition to displaying complimentary constraining properties to both weak lensing and CMB data, it is also unique among the low redshift probes in terms of its constraining power on a number of cosmological parameters. Galaxy clustering and weak lensing complement each other in a number of ways. Thanks to their accurate redshifts, spectroscopic galaxy samples can be used for redshift calibration of the weak lensing photometry or even as lenses themselves in the measurements of shear, induced in the source galaxies behind them (as done in, for example, Heymans et al., 2021). Moreover, weak lensing, being a probe of the total density field, can help break the degeneracy between the matter clustering amplitude and galaxy bias. Galaxy clustering measurements have, furthermore, traditionally been analysed in combination with CMB data due to the complimentary constraining properties, with CMB providing a constraint on the spectral index as well as the sound horizon distance, necessary to calibrate the size of the BAO feature, and galaxy clustering angular diameter distance measurements providing information on the late time background evolution (see, for example, Planck Collaboration et al., 2020; Alam et al., 2021). The majority of work involving galaxy clustering has, nonetheless, focused solely on the constraints obtained from BAO and RSD imprints on the two-point correlation function, rather than fitting the full measurement. While analyses based solely on RSD and BAO summary statistics allow excellent internal consistency tests and may help constrain beyond - ΛCDM scenarios, it has been shown that they do not preserve all the information of the full measurement, in particular, losing the additional constraining power available from its shape (Brieden et al., 2021).

Other analyses, therefore, make use of the information recovered from fitting the full shape of two-point clustering measurements, either in Fourier or configuration space, directly comparing models against data (Tröster et al., 2020; d'Amico et al., 2020; Ivanov et al., 2020b; Chen et al., 2021; Sánchez et al., 2017; Grieb et al., 2017). These analyses tend to lose some of the immediate interpretability of the summary statistics but instead, allow us to directly obtain constraints of cosmological parameters independently of external data sets. This type of analyses have, therefore, recently received attention as a way to test the consistency between large-scale structure (LSS) and CMB measurements.

Tröster et al. (2020) showed that the full shape analysis of galaxy clustering on its own produces cosmological constraints that are comparable to those of other low-redshift probes. This work followed the analysis of correlation function wedges of Baryon Oscillation Spectroscopic Survey (BOSS) galaxies by Sánchez et al. (2017) in order to derive constraints on flat ΛCDM cosmologies from galaxy clustering alone (i.e. without combining it with CMB measurements, as was done in the BOSS Data Release 12 consensus analysis, Alam et al., 2017). Furthermore, the work also presented joint low-redshift constraints by combining galaxy clustering with weak lensing measurements from the Kilo-Degree Survey (KV450). The σ_8 value recovered from the full shape analysis of the correlation function wedges by Tröster et al. (2020) is 2.1σ low compared to *Planck*'s prediction, with the difference increasing to 3.4σ when weak lensing measurements from KV450, are added, indicating that there may be some consistent discrepancy between CMB predictions and

low redshift observations. This is also consistent with the already mentioned more recent analysis by Heymans et al. (2021) where BOSS galaxies are used as lenses in the so-called '3×2pt' analysis (a set of three correlation functions consisting of autocorrelation of the lens galaxy positions, source galaxy shapes and the cross-correlation of the two), which finds a $\sim 3\sigma$ discrepancy with *Planck*'s value of S_8.

A crucial point not addressed in the discussion above is that the parameters used to assess the tension between the different probes might not be defined appropriately for this purpose. The issue with σ_8 and, therefore, also S_8, is the fact that both of them are defined through the Hubble parameter proxy $h = H_0/100$ km s^{-1}Mpc^{-1}, which introduces additional dependence of non-cosmological origin on the particular value of h recovered by the specific probe (as first pointed out by Sánchez, 2020). The use of h^{-1}Mpc units dates back to the first redshift surveys we discussed at the start of this introduction: at low redshifts expressing distances in terms of h^{-1}Mpc could yield a measurement of the comoving distance that did not depend on the cosmology assumed to perform the conversion from redshifts. As pointed out by (Sánchez, 2020), due to the fact that σ_8 is measured on a scale defined in h^{-1}Mpc, the resulting value is averaged over all of the scales, in Mpc, that correspond to the posterior of h recovered by that particular data set. As a result, the measurements of σ_8, as obtained by different probes will not describe the clustering amplitude consistently. A similar issue is furthermore experienced by any other parameter defined through h as well. The confusion associated with the use of h^{-1}Mpc units can be avoided by simply redefining the parameter space in such a way that none of the cosmological parameters is defined through h. This change has implications both on the power spectrum modelling and the cosmological parameter constraints obtained in the updated parameter space (or, the *physical parameter space*, which is the term we will use to refer to the set of cosmological parameters not defined through h).

While great progress has been made in a number of aspects of the redshift space survey science both on the side of instrumentation, which allows us to obtain spectroscopy of increasingly larger galaxy samples, and the side of the modelling, where sophisticated theory methods have proved to be able to describe clustering to greater accuracy and at smaller scales than ever before, many analyses still often focus on the summary statistics only and express their measurements in the h^{-1}Mpc units, following the tradition of the early galaxy clustering analyses. The aim of the work presented in this thesis is then to update the analysis framework used for the two-point galaxy clustering and bring it closer to the demands and the power offered by the modern data sets. In this work, we present the full shape analysis of galaxy clustering in configuration space with an updated power spectrum model which offers an alternative to the currently more commonly used Effective Field Theory based prescriptions. Furthermore, we present the first analysis carried out in the physical parameter space: we discuss the resulting constraints and compare them to the results obtained in the usual parametrisation, demonstrating the advantages of our approach. Finally, we introduce a two-point correlation function theory emulator, which provides a significant speedup to the galaxy clustering analyses and which, owing to the physical parameter space-based modelling approach, is characterised by great flexibility and an outstanding validity range in terms of cosmologies and redshifts covered. Throughout

this thesis, we both demonstrate the power of galaxy clustering as a solo probe as well as explore its compatibility with weak lensing and CMB data. The analyses presented here are performed on galaxy clustering measurements from BOSS galaxy and eBOSS quasar samples (Dawson et al., 2013, 2016), which constitute the most powerful publicly available galaxy clustering data set at the moment.

We start the thesis by recapping the main theory concepts related to two-point galaxy clustering modelling in Chapter 2, where we also describe, in greater detail, the physical parameter space and the motivation behind its use. Chapter 3 presents the full shape cosmological analysis of the anisotropic galaxy two-point correlation function from BOSS and eBOSS within ΛCDM and is based on the work published in Semenaite et al. (2022). In this chapter, we demonstrate the observational implications of the physical parameter space constraints and discuss the consequences it has on assessing consistency and the σ_8 tension. Chapter 4 presents the material first published as Semenaite et al. (2023) and deals with constraints on the beyond-ΛCDM models and further demonstrates the particular issues associated with using h^{-1}Mpc units in cosmologies with evolving dark energy density. Finally, Chapter 5 presents the design and validation of the configuration space emulator, describing the theoretical model that it is based on, as well as the particular concepts that allow us to extend the number of cosmological parameters and redshifts supported. We provide a summary of the thesis and discuss the outlook for future work in Chapter 6.

Chapter 2

Theory of structure growth

This chapter provides a summary of the theoretical background relevant to the work presented in this thesis. As such, the focus of the material presented here is on structure formation and the statistical analysis of galaxy clustering. Crucially, the analysis presented in this thesis relies heavily on the insights allowed by the concepts collectively described as the 'evolution mapping' approach, which is described in a separate section with some relevant parts of the cosmology background (in particular, the discussion on dynamics of the Universe) already introduced in a compatible form.

We open the chapter with a description of the homogeneous Universe, starting with the introduction of the metric in Section 2.1 and then proceeding to discuss the dynamics of the Universe in Section 2.2. The concepts of redshifts and distances are furthermore defined in Section 2.3. Section 2.4 describes the seeding of the cosmic structure and the evolution of the density perturbations. This section also covers some of the main approaches for matter power spectrum modelling. The additional effects that need to be considered to model the *observed* galaxy clustering are presented in Section 2.5, which summarises the modelling of galaxy bias, redshift-space distortions and Alcock-Paczynski distortions. Finally, we end with Section 2.6 which describes the issues related to the usage of h^{-1}Mpc units from the perspective of power spectrum modelling and introduces a way to obtain accurate power spectrum predictions in a reduced parameter space.

The material presented in this chapter follows a similar discussion presented in the textbook by Dodelson & Schmidt (2020) as well as the lecture notes on "The Formation and Evolution of Cosmic Structures" by Sánchez.

2.1 The Cosmological Principle

The cosmological principle states that the Universe we live in is spatially homogeneous (uniform) and isotropic (it has no preferred direction or orientation). Importantly, the principle only holds on average and on large scales, indeed, the very existence of structure relies on the presence of density fluctuations. Nonetheless, the assumptions of homogeneity and anisotropy allow us to set up a framework for describing the very basics of the Universe

upon which we can build the required deviations.

In order to be able to accurately characterise the Universe it is, first of all, useful to define a coordinate system. We may use the equivalence principle to establish that in any free-falling frame of reference, the laws of physics will follow special relativity, which allows us to write down a metric - a quantity that naturally incorporates gravity through space-time distortions. The most general metric that obeys the cosmological principle is the *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric, which can be expressed in spherical coordinates as:

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-Kr^2} + r^2d\Omega^2\right]. \tag{2.1}$$

Here, the left-hand side defines the spacetime interval ds in terms of the speed of light c, the time coordinate t and the comoving space interval, expressed in terms of the comoving radius r and the angle Ω. The spatial part of the metric allows for a constant spatial curvature, characterised by K, which may take values $K < 0$ (open Universe) as well as $K > 0$ (closed Universe), with $K = 0$ corresponding to the flat (Euclidean) case. Finally, the factor $a(t)$ in front of the spatial part of the metric allows for a uniform spatial expansion or contraction, rescaling the distances between any two points in the Universe.

We can also rewrite the FLRW metric by defining the radial coordinate:

$$d\chi = \frac{dr}{\sqrt{1-Kr^2}}, \tag{2.2}$$

such that equation 2.1 is now expressed (after also expanding $d\Omega$) as:

$$ds^2 = c^2dt^2 - a^2(t)[d\chi^2 + S_K^2(\chi)(d\theta^2 + \sin^2\theta d\phi^2)], \tag{2.3}$$

where

$$S_K(\chi) = \begin{cases} \frac{1}{\sqrt{K}}\sin(\sqrt{K}\chi) & \text{if } K > 0 \\ \chi & \text{if } K = 0 \\ \frac{1}{\sqrt{K}}\sinh(\sqrt{|K|}\chi) & \text{if } K > 0. \end{cases} \tag{2.4}$$

This expression is equivalent to the one in equation (2.1) and will be useful in Section 2.3 when discussing distances.

2.2 Dynamics of the Universe

In the FLRW metric, the factor $a(t)$ is a relative quantity and, for convenience, it is normalised such that it is equal to unity today, i.e., $a_0 = 1.0$ (here and in the rest of this thesis the subscript '0' indicates quantities evaluated at present time). The time evolution of $a(t)$ depends on the contents of the Universe and may be derived by proceeding with our picture of an idealised Universe and modelling these contents as a *perfect fluid* (i.e. assuming that they have no shear, no viscosity and no heat conduction).

Einstein's field equations allow us to explicitly relate the metric to the constituents of the Universe:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \tag{2.5}$$

Here $g_{\mu\nu}$ represents the metric tensor which contains the information required to define the distance between two points on the manifold with its intrinsic curvature described by the Ricci tensor $R_{\mu\nu}$ and Ricci scalar R (which is a contraction of the Ricci tensor $R = g^{\mu\nu}R_{\mu\nu}$) with Λ as *the cosmological constant*. The curvature is sourced by the matter and energy distribution described by the stress-energy tensor $T_{\mu\nu}$. Finally, G is Newton's gravitational constant.

For an ideal fluid, described by its pressure p, density ρ and four-velocity u^μ, the stress-energy tensor is:

$$T^{\mu\nu} = (\rho + \frac{p}{c^2})u^\mu u^\nu + pg^{\mu\nu}. \tag{2.6}$$

In order to obtain the time evolution of $a(t)$, we can take this expression for the energy-momentum tensor (equation (2.6)) and the FLRW metric (equation (2.1)) and plug them into the Einstein's field equations (equation (2.5)), which results in the two *Friedmann equations*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\sum_i \rho_i + \frac{\Lambda}{3} - \frac{Kc^2}{a^2} \tag{2.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\left(\rho + \frac{3p}{c^2}\right), \tag{2.8}$$

which describe the dynamics of cosmic expansion. Here the sum $\sum_i \rho_i$ is over the densities of all the cosmic species i that make up the total energy budget of the Universe (except for the contribution from Λ, which is written out explicitly). Equation 2.8 illustrates that the acceleration is sourced by the *active mass density* $\rho + 3p/c^2$, i.e., in addition to the mass density ρ, the pressure also provides a contribution to the resulting dynamics. We can furthermore simplify our notation by defining the Hubble parameter:

$$H = \frac{\dot{a}}{a}. \tag{2.9}$$

Finally, in order to be able to solve the Friedman equations, we need to be able to describe how densities of different constituents evolve with time - i.e. we need an expression for $\rho(t)$. Making use of the conservation of energy $\nabla^\nu T_{\mu\nu} = 0$, the continuity equation can be derived as:

$$\dot{\rho} = -3\left(\rho + \frac{p}{c^2}\right)\frac{\dot{a}}{a}. \tag{2.10}$$

If we furthermore assume that the fluid has an equation of state $p = w\rho c^2$ and substitute this in the equation (2.10), we find the solution:

$$\rho(t) = \rho_0 a^{-3(1+w)}(t). \tag{2.11}$$

We now have a description of how the density of different constituents (characterised by their equation of state parameter w) will change with evolving $a(t)$. For pressureless non-relativistic matter, $w_{\rm m} = 0$ and the density changes with volume as a^{-3}, whereas radiation, which does have significant pressure, has $w_\gamma = 1/3$ and, therefore, responds to changes in $a(t)$ more strongly, as a^{-4}.

We may, on the other hand, also consider how components with different values of w influence the dynamics of the Universe. In order to do so, we need to plug in the expression for pressure from the equation of state into the Friedmann equation for the acceleration of the expansion (equation (2.8)):

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\rho\,(1+3w)\,. \tag{2.12}$$

It is then clear that both matter and radiation produce $\ddot{a} < 0$, which makes it necessary to introduce an additional component to achieve solutions where $\dot{a}$ is constant or increases. This can be done through a contribution from a constituent with $w < -1/3$ and, in the simplest case, we may take $w = -1$, which results in constant energy density $\rho(t) = \rho_0$. From the equation of state, this will also exhibit negative pressure and, therefore, corresponds to the cosmological constant term Λ in Einstein's field equations.

While a component with negative pressure does not correspond to any classical physical system, it is a key ingredient needed to describe the dynamics of the Universe that we actually observe. The first indication that the Universe may not be static was obtained by Slipher (Slipher, 1917) who measured the spectra of dozens of galaxies to find that they were all redshifted (we will describe what this means in greater detail shortly). In 1929 Hubble's measurement of Cepheid distances to 24 galaxies confirmed that they were receding with velocities linearly proportional to their distances which indicated that the Universe is expanding. Half a century later, the Supernovae Ia distance measurements further determined that this expansion was accelerating (Perlmutter et al., 1999; Riess et al., 1998). Nonetheless, little is known about the physical nature of the component responsible for this acceleration, accordingly named *dark energy*. While quantum systems are able to produce predictions consistent with $w = -1$, so far no such solution has been confirmed observationally (and the calculation attributing dark energy to the zero-point energy in quantum field theory produces an estimate for the amount of dark energy today that is, famously, tens of orders of magnitude too large).

We have now seen that each energy component will contribute to the expansion rate of the Universe. This connection can be written down straightforwardly by expressing the dimensionless Hubble parameter as a sum of redefined energy density parameters. In order to do so, we start from the Friedmann equation (2.7), which we evaluate today (so $a = 1$). We, furthermore, rewrite the contribution from the cosmological constant Λ as an additional energy component ρ_i and replace the left-hand side with the Hubble parameter to obtain:

$$H_0^2 = \frac{8\pi G}{3}\sum_i \rho_{i,0} - Kc^2. \tag{2.13}$$

The Hubble parameter has units of km $\mathrm{s}^{-1}\mathrm{Mpc}^{-1}$, but often a dimensionless definition of H is used:

$$h = H_0/(100 \text{ km s}^{-1}\mathrm{Mpc}^{-1}). \tag{2.14}$$

For convenience, we will define the constant $H_{100} = 100$ km $\mathrm{s}^{-1}\mathrm{Mpc}^{-1}$, which we will use to rewrite equation (2.7) in dimensionless quantities:

$$h^2 = \frac{8\pi G}{3H_{100}^2}\sum_i \rho_{i,0} - \frac{Kc^2}{H_{100}^2}. \tag{2.15}$$

The pre-factor in the first term has units of volume over mass, which allows us to define a reference density:

$$\rho_{100} \equiv \frac{3H_{100}^2}{8\pi G}. \tag{2.16}$$

We may then rewrite the first term as a sum of dimensionless *physical densities*, defined as:

$$\omega_i \equiv \frac{\rho_{i,0}}{\rho_{100}} = \frac{8\pi G}{3H_{100}^2}\rho_{i,0}. \tag{2.17}$$

Equivalently, we can define the physical curvature density as:

$$\omega_\mathrm{K} = -\frac{Kc^2}{H_{100}^2}. \tag{2.18}$$

Equation (2.15) then finally looks like:

$$h^2 = \sum_i \omega_i + \omega_\mathrm{K}, \tag{2.19}$$

making it explicit that the value of the expansion rate at any given time is determined by the sum of the physical densities of the relevant constituents of the Universe at that time. We may also rewrite the Friedmann equation (2.7) in terms of the sum of the physical densities of all energy species i, such that:

$$\dot{a}^2 = H_{100}^2\Big[\sum_{i=0}^{N}\omega_i a^{-(1+3w_i)} + \omega_\mathrm{K}\Big]. \tag{2.20}$$

Here equation (2.11) combined with the appropriate value of w allows us to obtain the time evolution of each density contribution.

Alternatively to the physical densities, it is common to define another dimensionless density quantity usually referred to as the *fractional* or *relative density*, Ω_i, which relates to the physical densities as

$$\Omega_i = \frac{\omega_i}{h^2} \tag{2.21}$$

and, therefore, signifies the relative contribution of a constituent to the total energy budget (and $\sum_i \Omega_i = 1$). Relative densities are usually formally defined as

$$\Omega_i \equiv \frac{\rho_{i,0}}{\rho_\mathrm{c}}. \tag{2.22}$$

Here ρ_c is the *critical density* - the solution to equation (2.7) for a flat Universe ($K = 0$):

$$\rho_c = \frac{3H_0^2}{8\pi G} = \rho_{100}h^2. \tag{2.23}$$

We note that, since Ω_i is defined through the critical density, it does depend on the value of H_0, unlike the physical density ω_i (this can be seen more clearly in equation (2.21)). This point will become crucial in the following chapters where we will illustrate the associated issues when attempting to measure these parameters. Finally, we may also use this notation to (yet again) rewrite the Friedmann equation (2.7) as:

$$H^2 = H_0^2 \left[\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_K a^{-2} + \Omega_{DE}\right]. \tag{2.24}$$

This is equivalent to the equation (2.20) and we can use this to consider the dominant constituents at different times: currently, we live in a dark energy-dominated epoch with $\Omega_{DE} = 0.68$ and matter contributing the remaining part of energy budget: $\Omega_m = 0.32$ (Planck Collaboration et al., 2020). The contribution from radiation is negligible at low redshifts and, Ω_K is consistent with 0. This composition leads to the expansion rate of $H_0 = 67$ km s^{-1}Mpc^{-1}.

We now have a recipe for describing the time evolution of the expansion rate $H(t)$, which is determined entirely by the relative contributions of different constituents at the corresponding time. With this in mind, let us now focus on one of the h-dependent quantities and discuss distance.

2.3 Redshifts and Distances

In the previous section, we already noted that the expansion of the Universe was first detected by measuring the *redshifts* of distant galaxies - this refers to the effect where the observed galaxy spectral lines are shifted in frequency with respect to what we would expect in the emitted spectrum. This is a direct consequence of the expanding Universe and is a key tool for determining distances in a galaxy survey.

In order to see these two points and obtain a definition for redshift, we first make use of the fact that light follows null geodesics, i.e. $ds = 0$ in FLRW metric in equation (2.1). We may orient the path of the photon in question such that $d\Omega = 0$ and we are left with:

$$\int_{t_e}^{t_o} \frac{c\, dt}{a(t)} = \int_{r_e}^{r_o} \frac{dr}{\sqrt{1 - Kr^2}}, \tag{2.25}$$

where the integral limits are from the time and position of emission, t_e and r_e, to the time and position of observation, t_o and r_o. We can consider another such photon emitted a small time interval dt_e later, which will be observed at time $t_o + dt_o$. The right-hand side of the equation corresponds to an integral over the radial metric coordinate (as defined in equation (2.2)) and, therefore, is the same for both photons (r_e and r_o do not change).

This means that the left-hand sides of the equation that correspond to each photon will be equal and we can subtract them to obtain:

$$\int_{t_\mathrm{e}}^{t_\mathrm{e}+dt_\mathrm{e}} \frac{c\ dt}{a(t)} = \int_{t_\mathrm{o}}^{t_\mathrm{o}+dt_\mathrm{o}} \frac{c\ dt}{a(t)}. \tag{2.26}$$

For a small time interval, we are left with

$$dt_\mathrm{e}/dt_\mathrm{o} = a(t_\mathrm{e})/a(t_\mathrm{o}). \tag{2.27}$$

Due to the expansion of the Universe, we have $a(t_\mathrm{o}) > a(t_\mathrm{e})$ and, therefore, $dt_\mathrm{e} < dt_\mathrm{o}$. This will also result in an equivalent change in frequencies and wavelengths of the light, such that: $\lambda_\mathrm{e}/\lambda_\mathrm{o} = a(t_\mathrm{e})/a(t_\mathrm{o})$. Therefore, $\lambda(t_\mathrm{o}) > \lambda(t_\mathrm{e})$, i.e., the wavelength increases as the photon is travelling towards the observer - the light is redshifted. If we define this difference as:

$$z \equiv \frac{\lambda_\mathrm{o} - \lambda_\mathrm{e}}{\lambda_\mathrm{e}}, \tag{2.28}$$

where λ_o describes the photon that is observed today, we get:

$$1 + z = \frac{1}{a(t)}. \tag{2.29}$$

The redshift experienced by the light also explains why the energy density of radiation evolves as a^{-4}: as the Universe expands, the number density (and, therefore, also rest-mass energy density) of photons gets diluted with the increase in volume, however, unlike matter, photons also suffer additional energy loss as their wavelength is stretched as a^{-1}.

As stated in equation (2.29), redshift allows us to assign a relative distance to observed galaxies by relating the shift of their spectra lines to the scale factor ratio, but, on its own, it still does not provide us with an actual measure of a physical distance. As we are about to see, in order to directly relate these two quantities, we need to assume a cosmology.

In order to obtain the distance to an object, let us go back to the path of the photon emitted at time t_e and travelling radially along the null geodesic to be observed by us today ($t_\mathrm{o} = t_0$). Once again, we can get an expression for the radial coordinate $d\chi$ from the FLRW metric (equation (2.3)):

$$\chi(t_\mathrm{e}) = \int_{t_\mathrm{e}}^{t_0} \frac{cdt}{a(t)}. \tag{2.30}$$

In order to re-express this integral in terms of redshift, we need to obtain an expression for dz/dt, which makes use of definitions in equation (2.9) and equation (2.29):

$$\frac{dz}{dt} = \frac{dz}{da}\frac{da}{dt} = -(1+z)^2 H(z) a = -(1+z)H(z). \tag{2.31}$$

Plugging this into the expression for $d\chi$ we get:

$$\chi = \int_0^z \frac{cdz'}{H(z')}, \tag{2.32}$$

where the integral is from today ($z = 0$) to the redshift of emission (z). The quantity χ is referred to as *comoving* distance, χ, constant with expansion. In order to obtain the physical (or *proper*) distance to the point at the time of the emission, taking the expansion into account, we simply need to multiply the comoving distance by the scale factor:

$$s(z) = a(z)\chi(z). \tag{2.33}$$

When defining distances, we may rewrite these expressions in terms of the value of Hubble constant today and the sum of the relative densities $E(z) = \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_{\rm K} a^{-2} + \Omega_\Lambda$, so that for χ we now have:

$$\chi = \frac{1}{H_0}\int_0^z \frac{cdz'}{E(z')}. \tag{2.34}$$

This expression is helpful to illustrate the reasoning behind the fact that distance in cosmology is commonly presented in h^{-1}Mpc units. At low redshift, we may approximate equation (2.34) as $\chi \approx (cz)/H_0$, so distances expressed in the units of h^{-1}Mpc should yield a measurement that is approximately independent of the h value used.

The relation above is useful for converting redshifts into distances, however, in order to more directly relate to observational scenarios we should furthermore consider the notion of distance in terms of the more readily available angular position of an object or feature of interest. A useful quantity to define in this case is the *angular diameter distance*, $D_{\rm A}$, which describes the relationship between the physical size of an object δl and its angular extent $\delta\phi$. $D_{\rm A}$ is defined to make this mapping Euclidean-like through the ratio $\delta l/\delta\phi$. Using our second expression for FLRW metric 2.3 and setting $d\chi = d\theta = 0$ we can write down

$$D_{\rm A}(z) \equiv \frac{\delta l}{\delta\phi} = (1+z)^{-1} S_K(\chi(z)). \tag{2.35}$$

Equivalently to the radial distance, we can also define a comoving angular diameter distance:

$$D_{\rm M}(z) = D_{\rm A}(z)/a(z) = D_{\rm A}(z)(1+z) = S_K(\chi(z)), \tag{2.36}$$

such that for a flat Universe $D_{\rm M}(z) = \chi(z)$.

Generally, we can measure the angular extent, $\delta\phi$ of an object, however, its physical size δl is not necessarily known. A type of object whose δl can be determined is referred to as a *standard ruler* and it allows us to probe the redshift evolution of $D_{\rm A}(z)$ by taking the measurements of the standard ruler at different redshifts. Similarly, if this object is instead situated along the line of sight, we can probe the redshift evolution of $H(z)$, as, following the definition of the comoving distance in equation (2.32) and multiplying by a to convert this into the proper distance, we find:

$$\delta l = \frac{c\delta z}{(1+z)H(z)}. \tag{2.37}$$

As will be further discussed in Section 2.4.4, the standard ruler in galaxy clustering measurements is the baryon acoustic oscillation imprint, which can be measured both along and perpendicularly to the line of sight to probe $H(z)$ and $D_{\rm A}(z)$.

Looking at the expressions for distances derived in this section, it is clear that the most relevant quantity when creating a physical map from the observed redshifts and angles is the Hubble parameter $H(z)$, which, in turn, is defined by the sum of the physical densities of the constituents (and in a flat Universe with negligible radiation density, will be determined by $\omega_{\rm m}$ and $\omega_{\rm DE}$). Measuring distances, therefore, allows us to explore *the background evolution* of the Universe. However, this does not fully define a cosmological model - in particular, as we are about to see in the next section, in order to describe the observed distribution of galaxies, we need a recipe for the structure growth, which will require additional parameters.

2.4 Structure Formation

2.4.1 Two point functions

The Universe we observe is clearly not homogeneous and isotropic on small scales, the very existence of galaxies and galaxy clusters is a result of fluctuations in the matter density field, i.e., the cosmological principle only holds on average and on large scales. However, in order to further discuss the inhomogeneities, we need to first define some quantities that will allow us to describe them.

Density fluctuations are related to fluctuations in the gravitational potential Φ via the Poisson-like weak-field form of Einstein's gravitational field equations (in comoving units):

$$\nabla^2\Phi/a^2 = 4\pi G(1+3w)\bar{\rho}\delta, \tag{2.38}$$

where we defined the dimensionless fluctuation amplitude in terms of mean density $\bar{\rho}$:

$$\delta \equiv \frac{\rho-\bar{\rho}}{\bar{\rho}}, \tag{2.39}$$

and the factor $(1+3w)$ accounts for the relativistic active mass density. It is often instructive to inspect the scale dependence of $\delta(\boldsymbol{x})$, in which case it is convenient to consider the fluctuation amplitude in Fourier space $\delta(\boldsymbol{k})$, where the different modes k evolve independently for small values of $|\delta| << 1$. The Fourier space fluctuation amplitude $\delta(\boldsymbol{k})$ is related to its configuration space equivalent $\delta(\boldsymbol{x})$ through the Fourier transform:

$$\delta(\boldsymbol{x}) = \frac{1}{(2\pi)^3}\int \delta(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{x}}d^3k. \tag{2.40}$$

The mean of δ is zero by construction, whereas the variance is obtained from the volume average of δ^2, which is $\langle\delta^2\rangle = \langle\delta\delta^*\rangle$ for a real field, with the cross terms integrating to zero. This variance is also what defines the *power spectrum*, $P(k)$:

$$\langle\delta(\boldsymbol{k})\delta(\boldsymbol{k'})\rangle = (2\pi)^3\delta_{\rm D}(\boldsymbol{k}+\boldsymbol{k'})P(k), \tag{2.41}$$

where $\delta_{\rm D}$ is the Dirac delta function.

The configuration space equivalent of the power spectrum is the *two-point correlation function,* $\xi(r)$:

$$\xi(\boldsymbol{r}) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r})\rangle = \frac{1}{(2\pi^3)}\int P(\boldsymbol{k})e^{(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r})}d^3k. \tag{2.42}$$

For an isotropic density field, we may replace all the vector quantities with their scalar amplitudes. We furthermore impose the homogeneity requirement, so that the two-point correlation function does not depend on a particular location $\boldsymbol{x}$, but rather only the separation between two points $r = |\boldsymbol{r}|$. For a discrete set of points, the two-point correlation function can be interpreted as the excess probability of finding two tracers separated by some distance r compared to a homogeneous distribution.

We can also re-express the relation between $\xi(r)$ and $P(k)$ as:

$$\xi(r) = \frac{1}{2\pi^2}\int P(k)j_0(kr)k^2dk, \tag{2.43}$$

where $j_0 = \sin(x)/x$ is the spherical Bessel function of the first kind. If the density field in consideration is Gaussian, all even higher-order moments of the density fluctuations can be obtained from the variance, $\langle\delta^{2n}\rangle \propto \langle\delta^2\rangle^n$, and all odd higher-order moments for a field with zero mean will also be zero, $\langle\delta^{2n+1}\rangle \propto \langle\delta\rangle$. Therefore, in the Gaussian density field case, the power spectrum or, equivalently, the two-point correlation function will carry the full description of the field.

2.4.2 Seeding the structure: some words on inflation

The fact that any density fluctuations should exist in the first place is not obvious. It turns out that, in order to uncover the early Universe origins of the structures that we observe today, we need to look at scales so small that quantum effects apply and invoke a process that allows us to grow these effects to observable scales.

Inflation, in addition to solving a number of other cosmological problems, provides us with exactly such a recipe through a brief period of rapid, accelerated growth of the very early Universe. In the simplest inflationary scenario, this growth is driven by a single scalar field - the *inflaton*, which naturally undergoes quantum fluctuations. However, as the Universe expands, the wavelength of these fluctuations eventually exceeds the distance of the causal connection.

The rate of expansion during inflation is nearly exponential and we can define the *comoving Hubble radius* $(aH)^{-1}$ - the approximate distance that the light can travel during the time in which the scale factor increases by a factor of e. This also determines the scale for causal contact and any Fourier modes of the inflaton with wavelengths exceeding $(aH)^{-1}$ can no longer evolve. The time evolution of $(aH)^{-1}$ is determined by:

$$\frac{d}{dt}(aH^{-1}) = \frac{d}{dt}\left(\frac{1}{\dot{a}}\right) = -\frac{\ddot{a}}{\dot{a}^2}. \tag{2.44}$$

This implies that in a Universe undergoing accelerated growth, as is the case during the inflationary epoch, the comoving Hubble radius decreases (as per equation (2.12), this

requires a dark-energy like component with $w < -1/3$). As a result, more and more modes exceed this scale and are "frozen out" at a nearly constant amplitude. Inflation ends once the inflaton reaches a determined field value, which will happen at slightly different times in different regions due to the fluctuations in the inflaton field. These fluctuations will, in turn, be reflected in the resulting density field, whose value at any position will depend on the amount of growth that patch of space experienced.

In the simplest scenario, the seeded fluctuations are predicted to be Gaussian (due to the quantum nature of the initial fluctuations the final field will be a superposition of Fourier modes with independent random phases leading to Gaussian distribution) and adiabatic (i.e. the perturbations in the energy species are such that the ratios in their number densities remain unchanged) with a nearly scale-invariant power spectrum. The latter property means that each decade of potential fluctuations of each wavelength provides equal levels of distortion to the spacetime, resulting in a fractal structure: each level of spacetime resolution exhibits the same level of deviation from the unperturbed metric. This subsequently results in an approximately scale-invariant density field power spectrum, when considered per log-length scale, characterised by a power law with the *spectral index* $n_{\rm s}$:

$$n_{\rm s} \equiv \frac{d\ln P(k)}{d\ln k}, \tag{2.45}$$

where $n_{\rm s}$ is expected to be close to unity.

2.4.3 Evolution in linear regime

In order to predict how the inflation-seeded density fluctuations will evolve over time, we need to obtain the equations of motion for the density field. Here, again, the scale of the perturbation plays an important role in determining the subsequent evolution. At the end of inflation, the density perturbations are characterised by modes with $k \ll aH$: i.e., they are outside of the Hubble scale and the potential Φ is almost constant. Nonetheless, during the radiation and matter domination epochs that follow, the comoving Hubble radius grows and density fluctuations with $k \gg aH$ can re-enter the horizon. The subsequent growth of the modes, however, depends on the dominant background component and we can distinguish between evolution in the radiation-dominated and the matter-dominated era.

In order to derive the equations of motion of Φ for the sub-horizon scales during the matter-dominated era it is sufficient to consider the Newtonian gravity case for the pressureless non-relativistic cold dark matter, taking into account the background expansion. We can work in comoving units and define the conformal time τ as $dt = a(\tau)d\tau$. The Poisson equation can then be written down in the following form (this is just the Newtonian case of equation (2.38) with explicit dependencies on τ and $\boldsymbol{x}$):

$$\nabla^2\Phi(\boldsymbol{x},\tau) = a^2(\tau)4\pi G\rho(\tau)\delta_{\rm m}(\boldsymbol{x},\tau). \tag{2.46}$$

Furthermore, mass conservation implies the *continuity equation*:

$$\frac{\partial\delta(\boldsymbol{x},\tau)}{\partial\tau} + \nabla\cdot\Big[\Big(1+\delta(\boldsymbol{x},\tau)\Big)\boldsymbol{v}(\boldsymbol{x},\tau)\Big] = 0, \tag{2.47}$$

where $\boldsymbol{v}(\boldsymbol{x},\tau)$ is the comoving peculiar velocity field. Finally, from the conservation of momentum, we obtain the *Euler equation*:

$$\frac{\partial \boldsymbol{v}(\boldsymbol{x},\tau)}{\partial \tau} + a(\tau)H(\tau)\boldsymbol{v}(\boldsymbol{x},\tau) + \boldsymbol{v}(\boldsymbol{x},\tau)\cdot\nabla\boldsymbol{v}(\boldsymbol{x},\tau) = -\nabla\Phi(\boldsymbol{x},\tau). \tag{2.48}$$

In the linear regime $|\delta| \ll 1$ and $|\boldsymbol{v}| \ll 1$, so we can rewrite equations (2.47) and (2.48) as:

$$\frac{\partial \delta(\boldsymbol{x},\tau)}{\partial \tau} + \theta(\boldsymbol{x},\tau) = 0, \tag{2.49}$$

$$\frac{\partial \boldsymbol{v}(\boldsymbol{x},\tau)}{\partial \tau} + a(\tau)H(\tau)\boldsymbol{v}(\boldsymbol{x},\tau) = -\nabla\Phi(\boldsymbol{x},\tau), \tag{2.50}$$

where we defined the velocity divergence $\theta(\boldsymbol{x},\tau) \equiv \nabla\cdot\boldsymbol{v}(\boldsymbol{x},\tau)$. Combining the linearised Euler and continuity equations and inserting the expression for $\nabla\Phi$ from Poisson equation, we can obtain the following second-order differential equation for $\delta(\boldsymbol{x},\tau)$:

$$\frac{\partial^2 \delta(\boldsymbol{x},\tau)}{\partial \tau} + a(\tau)H(\tau)\frac{\partial \delta(\boldsymbol{x},\tau)}{\partial \tau} - 4\pi G a^2(\tau)\rho(\tau)\delta_{\mathrm{m}}(\boldsymbol{x},\tau) = 0. \tag{2.51}$$

For a matter-dominated Universe, we can take $a \sim t^{2/3}$ [1], which gives the growing power law solution $\delta(t) \propto t^{2/3}$ and so the density contrast grows proportionally to the scale factor $\delta(t) \propto a$. Subsequently, as the dark energy contribution grows and is eventually no longer negligible, deviations from the matter domination behaviour appear. In general, the growing mode solution to equation (2.51) can be simply written down as:

$$\delta(\boldsymbol{x},\tau) = D_1(\tau)\delta(\boldsymbol{x}), \tag{2.52}$$

where $D_1(\tau)$ is the *linear growth factor* and its full form is given by:

$$D_1(a) = \frac{5\Omega_{\mathrm{m}}}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3}. \tag{2.53}$$

The value of the linear growth factor is, therefore, determined by $H(a)$ and Ω_{m}, i.e., the composition of the Universe. It is customary to normalise D_1 such that $D_1 = 1$ today.

Assuming Newtonian physics is, however, no longer valid for the radiation-dominated epoch or for perturbations outside of the Hubble radius, which both require a full general relativity perturbation theory treatment. The full derivation for this case is beyond the scope of this thesis and we will simply quote the results. During radiation domination the potential Φ decays inside the horizon while the matter perturbation δ experiences logarithmic growth as $\delta \propto \ln t \propto \ln a$. For perturbations with wavelengths exceeding $(aH)^{-1}$ the constant potential results in $\delta \propto a^2$ during the radiation era and $\delta \propto a$ during matter domination [2].

[1] This can be obtained from equation (2.24) by integrating $d\ln(a)/dt \approx H_0\sqrt{\Omega_i}a^{-\frac{3}{2}(1+w_i)}$ with $\Omega_{\mathrm{m}} = 1$.

[2] This is a simplified picture and the quoted results additionally depend on the choice of coordinates (gauge). The full description goes beyond the scope of this chapter.

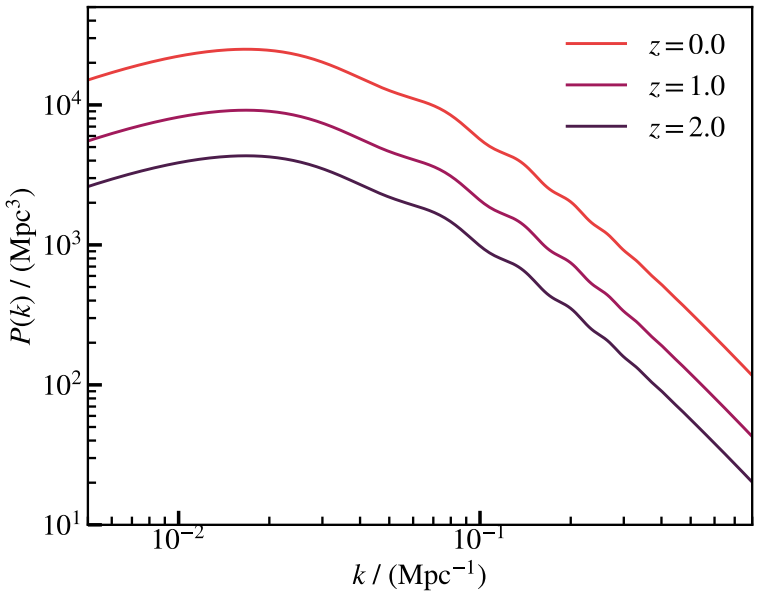

Figure 2.1: Comparison of linear matter power spectra for redshifts $z = 0, 1, 2$. As stated in equation (2.54), time evolution corresponds to simple rescaling in amplitude.

The latter result means that the evolution of the density contrast is scale-independent during the matter domination - the growth of δ is proportional to the scale factor both inside and outside the Hubble radius. On the other hand, during the radiation domination, the growth is suppressed inside the Hubble radius with respect to the evolution outside of it ($\delta \propto \ln a$ versus $\delta \propto a^2$). These differences lead to overall scale-dependent growth and are reflected in the resulting matter power spectrum.

2.4.4 The Linear Power Spectrum

As discussed in the previous subsection, at the end of the inflation the resulting power spectrum is predicted to have the form of $P(k) \propto k^{n_s}$. During the radiation domination epoch, the Hubble radius increases, which means that the fluctuations can gradually re-enter, starting with the ones with the smallest wavelengths (largest k). As soon as a mode enters the Hubble radius, its growth will be suppressed - the smallest scale (largest k) modes will spend the longest time in this regime and will experience the most suppression, whereas the large scale (small k) modes will be unaffected and continue to grow as $\delta \propto a^2$. After the matter-radiation equality, all modes, regardless of their wavelength, will grow at the same rate as $\delta \propto a$.

The suppressed small-scale growth during radiation domination introduces a turnover scale in the power spectrum which corresponds to the size of the Hubble radius at matter-radiation equality. In general, the resulting k-dependent growth is described by the transfer

function $T(k)$. The transfer function, together with the linear growth factor, allows us to obtain an expression for the linear power spectrum in terms of Fourier mode k and scale factor a. The initial shape of the power spectrum is set by the primordial power spectrum and is characterised by the spectral index $n_{\rm s}$. The shape is subsequently altered during radiation domination, as encoded in $T(k)$ while the growth of the amplitude is fully described by the squared growth factor $D_1(a)^2$. Putting it all together the final expression for the linear power spectrum is:

$$P_{\rm lin}(k,a) = A\Big[\frac{D_1(a)}{D_1(a_0)}\Big]^2 T^2(k)\Big(\frac{k}{k_0}\Big)^{n_{\rm s}}. \tag{2.54}$$

Here the time evolution of the power spectrum is obtained by scaling the global amplitude of the fluctuations A as measured at $z = 0$ (or $a_0 = 1.0$) at an arbitrary pivot scale k_0, which is conventionally chosen to be $k_0 = 0.05{\rm Mpc}^{-1}$. This effect is illustrated in Figure 2.1, which displays linear power spectra at different redshifts: we can see that all of the predictions differ in the amplitude only, with the power spectrum with the largest amplitude corresponding to the lowest redshift, where density fluctuations will have experienced the most growth. Equation (2.54) also highlights how the amplitude evolution of the power spectrum is expressed relative to some reference value at a set scale and redshift or scale factor. The choice of appropriate normalisation is, however, not as straightforward as it might seem and will be further discussed in Section 2.6.

Baryon Acoustic Oscillations

In addition to the imprint of the horizon length at matter-radiation equality, the linear power spectrum is characterised by the *sound horizon* - another important scale, which is an imprint of the *baryon acoustic oscillations* (BAO) in the early Universe. During the radiation domination epoch, before the atoms are formed, baryons and photons are coupled together through Compton scattering. During this epoch, the matter perturbations inside the Hubble radius can grow at a logarithmic rate, however, unlike dark matter, baryons are subject to the radiation pressure from photons, which prevents the growth. The interplay between gravity and pressure instead produces sound waves in the photon-baryon fluid - it undergoes acoustic oscillations. This process lasts until the Universe expands enough to cool down and the atoms are formed during *recombination*. During this epoch, at the redshift of $z_* \approx 1090$, the mean free path of photons increases and they can decouple from baryons forming the cosmic microwave background radiation (CMB). The baryons decouple from photons somewhat later, at the redshift of $z_{\rm d} \approx 1060$, which marks the so-called *drag epoch*. After this point, baryons, no longer supported by radiation pressure, fall into the dark matter potential wells and the two species grow together as a single matter component.

The acoustic oscillations leave an imprint both in the power spectrum of the CMB and the matter power spectrum. This is because baryons constitute a significant enough fraction of the total matter that the dark matter component is affected by the gravitational force exerted by the baryons as well. This means that the BAO signal is present in the total

matter distribution with an amplitude determined by the fraction of baryons, $f_{\rm b} = \Omega_{\rm b}/\Omega_{\rm m}$. In Fourier space the oscillations show up as a series of wiggles, modulating the amplitude of the power spectrum with the extrema at $k_n = n\pi/r_{\rm s}$, where $r_{\rm s}$ is the sound horizon which marks the maximum distance that the acoustic waves could travel before decoupling. In configuration space, the two-point correlation function exhibits a broad peak at a scale of ~ 155 Mpc. This is an important feature that can be used as a *standard ruler*, a concept that we already discussed in Section 2.3: by fitting the angular and redshift extent of the BAO peak in galaxy clustering measurements and comparing it with a known $r_{\rm d} = r_{\rm s}(z_{\rm d})$ (determined accurately from the CMB observations), we can probe the background expansion of the Universe, i.e., we can constrain $H(z)r_{\rm d}$ and $D_{\rm M}(z)/r_{\rm d}$.

2.4.5 Non-linear evolution

The linear power spectrum, as described above, neglects the non-linear contributions to density and velocity fields and can, therefore, only provide reliable predictions for large scales and the very early Universe where $|\delta| \ll 1$. As structure continues to form and overdensities grow, it is crucial to model the resulting non-linearities in order to obtain accurate fits of the galaxy clustering measurements. In Figure 2.2 we show the linear theory prediction (black dotted line) for the matter power spectrum at redshifts $z = 0$, 1, 2 and compare this result with two different non-linear matter power spectrum predictions (solid and dot-dashed lines). It is clear that, while all models agree on large scales, where the linear theory is accurate, the three predictions diverge on small scales with linear theory underpredicting the power, as it does not take into account the contribution from mode coupling, which is significant and large k. The efforts to model the non-linear regime can be broadly categorised into three different types of approaches: predictions based on N-body simulations, perturbative methods, and the halo model-based theories.

The N-body simulations, while the most computationally expensive of the three, also provide us with the most accurate theory predictions (as also demonstrated in Figure 2.2, where at $z = 0$ the solid line marking the power spectrum measured on simulations has more power at small scales than the perturbative prediction shown in dot-dashed line). This is because, instead of trying to obtain an analytical or phenomenological description of the matter distribution, N-body simulations allow us to directly model the evolution of the structure by following the evolution of particles in a cubic volume with periodic boundary conditions, updating their positions and velocities at a set of timesteps. The initial conditions at the start of the simulation are determined by the cosmological model and the equations of motion for each particle are calculated based on the resulting gravitational potential, which is also updated, as the simulation evolves. The power spectrum predictions can then be obtained by simply performing measurements based on the particle positions at the simulation snapshots at the desired redshift. While extremely accurate, the simulations are limited by the computational power available, which sets the number of particles (which defines the resolution of the simulation) and the size of the box (which determines the number of modes simulated). In addition to this, one must also run separate simulations for each cosmology of interest, which further adds to computational costs.

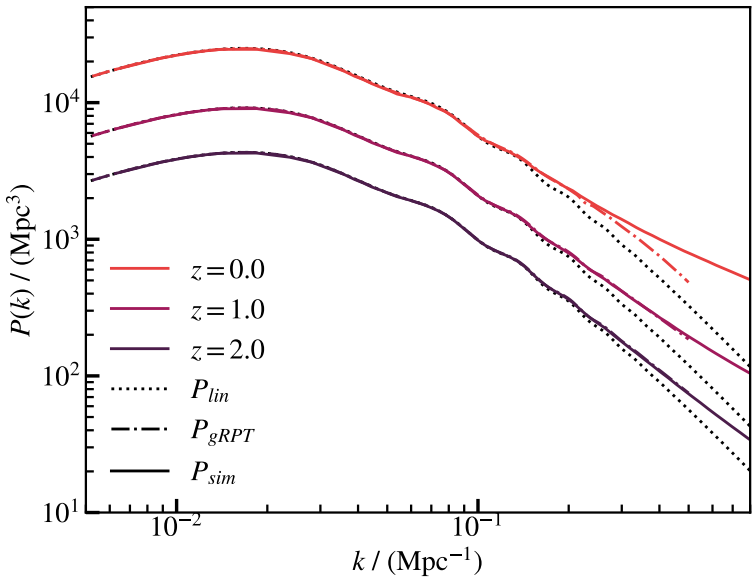

Figure 2.2: Comparison of the non-linear matter power spectra for redshifts $z = 0, 1, 2$, as obtained from N-body simulation (solid line) and the gRPT prediction (dot-dashed line). The black dotted line marks the linear power spectrum displayed for reference.

Finally, the final power spectrum prediction, while accurate, does not contain insights into the physical processes that contributed to its different features.

Both perturbative and halo-model-based methods incorporate these physical processes more explicitly, however, the two approaches are valid at different scales. Perturbative theory predictions are based on modelling the evolution of a fluid by considering small displacements around the linear case and performing an expansion where each higher-order term includes more non-linear physics. The halo model, on the other hand, assumes that all matter in the Universe is contained within dark matter haloes and the power spectrum is obtained by considering the contributions from halos of different masses. The former method is valid in the mildly non-linear regime where $|\delta| \sim 1$, but it fails to describe much smaller scales. The halo model, on the other hand, provides a good description of the power spectrum on fully non-linear scales but it is difficult to build for general cosmologies and requires the knowledge of a number of halo properties, such as the halo mass function, halo density profile, and the mass-concentration relation.

The methods used in the analysis performed in this thesis rely on perturbative and N-body simulation approaches. We will, therefore, proceed to introduce the basics of the perturbation theory while leaving out further details of the halo model.

Standard Perturbation Theory

Standard perturbation theory (SPT) refers to Eulerian perturbation theory, where the fluid is described in terms of density and velocity field on a fixed comoving coordinate system of time and position. In this approach, the perturbative variable is the density contrast, δ. We start back from the non-linear form of the continuity and Euler equations (equation (2.47) and (2.48)). We already noted that the continuity equation is expressed in terms of the velocity divergence rather than the velocity field itself and it is, therefore, common to also take the divergence of the Euler equation to re-express it in terms of θ. The Euler and continuity equations can then be Fourier transformed to obtain the coupled equations:

$$\frac{\partial\delta(\boldsymbol{k},\tau)}{\partial\tau}+\theta(\boldsymbol{k},\tau)=-\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3}(2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}_{12})\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)\theta(\boldsymbol{k}_1,\tau)\delta(\boldsymbol{k}_2,\tau), \quad (2.55)$$

$$\frac{\partial\theta(\boldsymbol{k},\tau)}{\partial\tau}+a(\tau)H(\tau)\theta(\boldsymbol{k},\tau)+\frac{3}{2}\Omega_{\mathrm{m}}(\tau)(a(\tau)H(\tau))^2\delta(\boldsymbol{k},\tau)=-\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3} \quad (2.56)$$
$$(2\pi)^3\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{k}_{12})\beta(\boldsymbol{k}_1,\boldsymbol{k}_2)\theta(\boldsymbol{k}_1,\tau)\theta(\boldsymbol{k}_2,\tau).$$

Here $\boldsymbol{k}_{12}=\boldsymbol{k}_1+\boldsymbol{k}_2$ and α and β are kernels describing the coupling between all pairs of modes $\boldsymbol{k_1}$ and $\boldsymbol{k_2}$:

$$\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)=\frac{\boldsymbol{k}_{12}\cdot\boldsymbol{k}_1}{k_1^2}, \qquad \beta(\boldsymbol{k}_1,\boldsymbol{k}_2)=\frac{k_{12}^2(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)}{2k_1^2k_2^2}. \quad (2.57)$$

The resulting non-linear evolution equations (2.55) and (2.56) can be contrasted with the linear case where the terms on the right-hand side are zero and each mode $\boldsymbol{k}$ evolves independently. In equation (2.56) we also made use of the Poisson equation (2.46) to eliminate $\nabla^2\Phi$.

The non-linear evolution equations do not have a general analytical solution, so we proceed to expand the non-linear density and velocity fields in terms of higher order terms $\delta^{(n)}\propto(\delta_{\mathrm{lin}})^n$:

$$\delta(\boldsymbol{k},\tau)=\sum_{n=1}^{\infty}\delta^{(n)}(\boldsymbol{k},\tau), \qquad \theta(\boldsymbol{k},\tau)=\sum_{n=1}^{\infty}\theta^{(n)}(\boldsymbol{k},\tau). \quad (2.58)$$

A separable solution for density and velocity fields of a given order, $\delta^{(n)}(\boldsymbol{k})$ and $\theta^{(n)}(\boldsymbol{k})$, is only available in cosmologies where $\Omega_{\mathrm{m}}=1$. In this case, a general solution can be written down in terms of wavevectors $\{\boldsymbol{q}_1,...\boldsymbol{q}_n\}$, such that $\boldsymbol{k}_1\equiv\boldsymbol{q}_1+...+\boldsymbol{q}_m$, $\boldsymbol{k}_2\equiv\boldsymbol{q}_{m+1}+...+\boldsymbol{q}_n$ and $\boldsymbol{k}\equiv\boldsymbol{k}_1+\boldsymbol{k}_2$:

$$\delta^n(\boldsymbol{k})=\int d^3\boldsymbol{q}_1...\int d^3\boldsymbol{q}_n\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{q}_{1...n})F_n(\boldsymbol{q_1},...\boldsymbol{q_n})\delta^{(1)}(\boldsymbol{q}_1)...\delta^{(1)}(\boldsymbol{q}_n), \quad (2.59)$$

$$\theta^n(\boldsymbol{k})=\int d^3\boldsymbol{q}_1...\int d^3\boldsymbol{q}_n\delta_{\mathrm{D}}(\boldsymbol{k}-\boldsymbol{q}_{1...n})G_n(\boldsymbol{q_1},...\boldsymbol{q_n})\delta^{(1)}(\boldsymbol{q}_1)...\delta^{(1)}(\boldsymbol{q}_n). \quad (2.60)$$

The perturbation theory kernels F_n and G_n are homogeneous functions of $\{\boldsymbol{q}_1, ...\boldsymbol{q}_n\}$ and obey recursion relations, allowing us to recover the linear theory solution without mode-coupling at first order: $F_1 = G_1 = 1$. At quadratic order, the kernels have the form of:

$$F_2(\boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{5}{7} + \frac{1}{2}\frac{\boldsymbol{q}_1 \cdot \boldsymbol{q}_2}{q_1 q_2}\Big(\frac{q_1}{q_2} + \frac{q_2}{q_1}\Big) + \frac{2}{7}\frac{(\boldsymbol{q}_1 \cdot \boldsymbol{q}_2)^2}{q_1^2 q_2^2}, \tag{2.61}$$

$$G_2(\boldsymbol{q}_1, \boldsymbol{q}_2) = \frac{3}{7} + \frac{1}{2}\frac{\boldsymbol{q}_1 \cdot \boldsymbol{q}_2}{q_1 q_2}\Big(\frac{q_1}{q_2} + \frac{q_2}{q_1}\Big) + \frac{4}{7}\frac{(\boldsymbol{q}_1 \cdot \boldsymbol{q}_2)^2}{q_1^2 q_2^2}. \tag{2.62}$$

The first term in both cases is the monopole (isotropic) contribution, which is directly proportional to $\propto \delta^2$ and which describes spherical collapse (although, for the sake of brevity, in the equations above the monopole also absorbs the isotropic part of the quadrupole contribution). The second term - the dipole, describes the bulk flow and is generated by the $\boldsymbol{v} \cdot \nabla$ term in the equation of motion. Finally, the third - quadrupole term, represents the effect of the tidal forces. While the solution presented above is only exact in the Universe made completely of matter, it has been shown that the perturbation theory kernel form presented in equations (2.61) and (2.62) is a good approximation for the ΛCDM cosmological model as well as a number of its extensions.

Following our approach with the density field, we can similarly perform an expansion of the power spectrum itself:

$$\begin{aligned}\langle\delta(\boldsymbol{k},\tau)\delta(\boldsymbol{k'},\tau)\rangle = \langle\delta^{(1)}(\boldsymbol{k},\tau)\delta^{(1)}(\boldsymbol{k'},\tau)\rangle + \langle\delta^{(1)}(\boldsymbol{k},\tau)\delta^{(3)}(\boldsymbol{k'},\tau)\rangle + \\ \langle\delta^{(3)}(\boldsymbol{k},\tau)\delta^{(1)}(\boldsymbol{k'},\tau)\rangle + \langle\delta^{(1)}(\boldsymbol{k},\tau)\delta^{(3)}(\boldsymbol{k'},\tau)\rangle + \langle\delta^{(2)}(\boldsymbol{k},\tau)\delta^{(2)}(\boldsymbol{k'},\tau)\rangle + ...\end{aligned} \tag{2.63}$$

Here all the odd moments are zero due to the Gaussian initial conditions and the sum is over the even moments only. Alternatively, we can rewrite this equation as a sum of the linear power spectrum $P_{\rm lin} = P^{(11)} = \langle\delta^{(1)}\delta^{(1)}\rangle$ and non-linear corrections:

$$P(k,a) = P_{\rm lin}(k,a) + P^{(22)}(k,a) + 2P^{(13)}(k,a) + ... \tag{2.64}$$

The linear order contribution $P_{\rm lin}$ is referred to as 'tree-level', whereas the corrections $P^{(22)}$, $P^{(13)}, ...$ are called 'loop' corrections and the second-order (or one-loop, so called because only a single integral over the wavenumber is involved) contributions are given by:

$$P^{(22)}(k,a) = 2\int \frac{d^3q}{(2\pi)^3}[F_2^s(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q})]^2 P_{\rm lin}(q,a) P_{\rm lin}(|\boldsymbol{k}-\boldsymbol{q}|, a), \tag{2.65}$$

$$P^{(13)}(k,a) = 3P_{\rm lin}(k,a)\int \frac{d^3q}{(2\pi)^3}[F_3(\boldsymbol{q}, -\boldsymbol{q}, \boldsymbol{k})] P_{\rm lin}(q,a), \tag{2.66}$$

where F_2^s is the symmetrical form of the kernel given in equation (2.61) and F_3 is another perturbation theory kernel whose full expression will be omitted here. In order to obtain a more precise non-linear power spectrum one can proceed in a similar manner by including higher-order corrections through higher loop terms.

Renormalised Perturbation Theory

The limiting feature of the SPT approach is the difficulty in achieving convergence on small scales. This is due to the fact that the loop corrections can vary both in sign and magnitude, with higher loop terms potentially carrying comparable contributions to the lower loop terms, making it difficult to truncate the series. As a result, SPT cannot model nonlinearities with sufficient accuracy for modern galaxy surveys.

The convergence can, however, be improved by grouping the loop corrections into the propagator and mode-coupling terms, as first suggested by Crocce & Scoccimarro (2006). In the renormalised perturbation theory (RPT) approach the non-linear power spectrum can then be rewritten as the sum:

$$P(k,a) = P_{\rm lin}(k)G(k,a)^2 + P_{\rm MC}(k,a). \tag{2.67}$$

The renormalised propagator term, $G(k,a)$, is obtained by grouping the loop corrections proportional to $P_{\rm lin}$ (such as $P^{(13)}$). The propagator measures how much power is directly linked to the initial conditions as a function of scale and redshift and it can be well described by a Gaussian damping of the linear growth factor (D_1, equation (2.53)). As the propagator is more damped at high k, the memory of the initial conditions is lost and the mode coupling contributions, $P_{\rm MC}$, become more important. This second term has a loop expansion which resembles that of the SPT power spectrum with its terms ordered based on the number of initial modes coupled. The one-loop contribution to $P_{\rm MC}$ at low-k can be approximated as $P^{(22)}$ (Crocce et al., 2012), given in equation (2.65). Crucially, all of the loop terms in the expansion are positive and their amplitudes become successively smaller - each of the higher loop corrections, therefore, only dominates at increasingly smaller scales, which allows for a non-ambiguous determination of the appropriate point of truncation for a particular analysis.

The series convergence can be improved even further by finding a way to resum not only the propagator but also the mode-coupling term. The Galilean-invariant renormalised perturbation theory (gRPT) approach imposes Galilean-invariance on the equations of motion in order to obtain a resummed mode-coupling term that is consistent with the propagator. The resulting convergence improvement allows us to obtain accurate power spectrum predictions at the one-loop level (see the power spectra in the dot-dashed line in Figure 2.2). Indeed, the gRPT modelling has been used in a number of modern galaxy clustering analyses, such as Grieb et al. (2017), Sánchez et al. (2017), reaching scales as small as $k_{\rm max} \sim 0.25\ h^{-1}$Mpc in Fourier space and $s_{\rm min} = 20\ h^{-1}$Mpc in configuration space.

RESPRESSO

In addition to purely perturbative or purely simulated methods, it is possible to obtain non-linear power spectrum predictions using a hybrid approach. The theory model used for the full-shape galaxy clustering analysis presented in this thesis also falls within this category.

We compute model predictions of the non-linear matter power spectrum, $P_{\rm mm}(k)$, using the Rapid and Efficient SPectrum calculation based on RESponSe functiOn approach (RESPRESSO, Nishimichi et al., 2017). The key ingredient of RESPRESSO is the response function, $K(k,q)$, which quantifies the variation of the non-linear matter power spectrum at scale k induced by a change of the linear power at scale q as

$$K(k,q) \equiv q\frac{\partial P_{\rm mm}(k)}{\partial P_{\rm lin}(q)}. \tag{2.68}$$

Nishimichi et al. (2016) presented a phenomenological model for $K(k,q)$ based on regularised perturbation theory (Taruya et al., 2012), which gives a good agreement with simulation results over a wide range of scales for k and q. The response function allows to obtain $P_{\rm mm}(k)$ for arbitrary cosmological parameters $\boldsymbol{\theta}$ based on a measurement from N-body simulations of a fiducial cosmology $\boldsymbol{\theta}_{\rm fid}$ as

$$\begin{aligned} P_{\rm mm}(k|\boldsymbol{\theta}) = P_{\rm mm}(k|\boldsymbol{\theta}_{\rm fid}) \int {\rm d}\ln q\, K(k,q) \\ \times [P_{\rm lin}(q|\boldsymbol{\theta}) - P_{\rm lin}(q|\boldsymbol{\theta}_{\rm fid})]. \end{aligned} \tag{2.69}$$

The choice of $\boldsymbol{\theta}_{\rm fid}$ in RESPRESSO corresponds to the best-fitting ΛCDM model to the *Planck* 2015 data (Planck Collaboration et al., 2016). Equation (2.69) is most accurate for cosmologies that are close to $\boldsymbol{\theta}_{\rm fid}$. For cosmologies further away from the fiducial, its accuracy can be improved by performing a multi-step reconstruction.

As RESPRESSO is partially based on numerical simulation measurements, it automatically incorporates higher-order corrections beyond the one-loop gRPT expansion and is, therefore, able to provide more accurate results. This was also confirmed by Eggemeier et al. (2020), who showed that RESPRESSO outperforms other perturbation theory-based models in terms of the range of validity and accurate recovery of the mean cosmological parameter posterior values. This approach was also used in the final eBOSS quasar clustering multipoles analysis by Hou et al. (2021).

2.5 Modelling Observed Galaxy Clustering

So far we described the methods required for predicting the matter power spectrum, however, the dark matter field cannot be observed directly but only through the galaxies that trace it. In this section, we, therefore, discuss the theoretical ingredients required to predict the *observed* galaxy power spectrum. A number of relevant effects that need to be taken into account can be well described by more than one recipe (for example, there are several galaxy bias parameterizations and different descriptions of the small-scale redshift-space distortions used in modern galaxy clustering analyses), however, all of the modelling choices presented here largely follow the original BOSS and eBOSS clustering analyses of Sánchez et al. (2017), Grieb et al. (2017) and Hou et al. (2021). The majority of this section describes the modelling of the power spectrum, the prediction for the observed two-point correlation function can then be obtained by performing a Fourier transform.

2.5.1 Galaxy Bias

Bias describes the relation between the luminous tracers that we observe and the total matter field. Here, perturbative approaches are used to describe the galaxy overdensity $\delta_{\rm g}$ in terms of a series of operators, such that, in the most general form $\delta_{\rm g}$ can be written down as the sum:

$$\delta_{\rm g}(\boldsymbol{x},\tau) = \sum_O b_O(\tau) O(\boldsymbol{x},\tau). \tag{2.70}$$

Here b_O are the bias parameters and $O(\boldsymbol{x},\tau)$ are the operators, which include the relevant quantities that affect galaxy formation, such as matter fluctuation at a given point in space or spatial derivatives of the gravitational and velocity potentials.

Following the notation of Eggemeier et al. (2019), The one-loop bias expansion can be written down as:

$$\delta_{\rm g} = b_1\delta_m + \frac{b_2}{2}\delta_m^2 + \gamma_2\mathcal{G}_2(\Phi_v) + \gamma_{21}\mathcal{G}_2(\varphi_1,\varphi_2). \tag{2.71}$$

At the linear order, the relationship between galaxy and matter fields is simply $\delta_{\rm g} = b_1\delta$ and the bias parameter b_1 is referred to as the *linear bias*. Including the second order, or *quadratic*, bias, b_2, provides a further correction to the spherical collapse description of galaxy bias (which assumes that galaxies form from the gravitational collapse of spherically symmetric perturbations).

The first two terms of the bias expansion in equation (2.71) describe galaxy bias as a function of the local matter density contrast only, however, anisotropies in gravitational collapse lead to additional tidal effects that affect the local distribution of galaxies (Sheth et al., 2013; Baldauf et al., 2012). These non-local effects are described by the terms that involve the Galilean invariant operator $\mathcal{G}_2$. The first of these terms represents the tidal stress tensor generated by the velocity potential Φ_ν and is given by:

$$\mathcal{G}_2(\Phi_\nu) = (\nabla_{ij}\Phi_\nu)^2 - (\nabla^2\Phi_\nu)^2. \tag{2.72}$$

The second of the tidal bias terms is a higher-order correction:

$$\mathcal{G}_2(\varphi_1,\varphi_2) = \nabla_{ij}\varphi_2\nabla_{ij}\varphi_1 - \nabla^2\varphi_2\nabla^2\varphi_1, \tag{2.73}$$

where φ_1 and φ_2 are obtained by expressing the non-linear velocity potential up to the second order ($\Phi_\nu = \Phi_\nu^{(1)} + \Phi_\nu^{(2)}$ and $\varphi_1 = -\Phi_\nu^{(1)}$, $\varphi_2 = -\Phi_\nu^{(2)}$), such that $\nabla^2\varphi_1 = -\theta$ is the linear velocity divergence field and $\nabla^2\varphi_2 = -\mathcal{G}_2(\varphi_1)$ is the next-to-leading order correction.

The one-loop galaxy bias expansion, as presented in equation (2.71), therefore, includes four free bias parameters: b_1, b_2, γ_2 and γ_{21}. This is a large number of free parameters for real-world applications, especially if the focus of the analysis is the information contained in two-point clustering measurements alone (in particular, the tidal bias parameters γ_2 and γ_{21} are degenerate with each-other). It is, therefore, common to make use of relations which set the values of tidal bias parameters in terms of linear bias.

One class of such relations is the *local Lagrangian* relations, which are obtained under the assumption that galaxy formation is driven exclusively by the local matter density

field and that all tracers are formed instantaneously at the infinite past time and their number remains conserved from that point on. This approach was used in, for example, BOSS galaxy clustering analyses of Sánchez et al. (2017) and Grieb et al. (2017). One can, however, take a step further, relaxing the requirement that galaxy formation is determined by the local matter density field and only assuming the tracer number conservation (Fry, 1996; Catelan et al., 1998, 2000; Chan et al., 2012). We can then obtain the following relation for γ_{21}:

$$\gamma_{21} = -\frac{2}{21}(b_1 - 1) + \frac{6}{7}\gamma_2. \tag{2.74}$$

This relation was thoroughly tested against constraints derived from a combination of power spectrum and bispectrum data in Eggemeier et al. (2021), and found to be in excellent agreement with a synthetic sample of BOSS-like galaxies. A similar expression can also be derived for γ_2, however, in the analyses performed in this thesis we instead use the following quadratic relation:

$$\gamma_2(b_1) = 0.524 - 0.547 b_1 + 0.046 b_1^2, \tag{2.75}$$

which describes the results of Sheth et al. (2013) using excursion set theory. Eggemeier et al. (2020) showed that the relation in equation (2.75) is more accurate for tracers with $b_1 \gtrsim 1.3$ than the equivalent co-evolution relation and is, therefore, more appropriate for our set of tracers.

In Sec. 3.2, we confirm that the use of these relations gives an accurate description of the results of N-body simulations and we therefore implement them in our analysis of the BOSS and eBOSS data. In this way, the only required free bias parameters in our one-loop galaxy bias model are b_1 and b_2, while the non-local bias terms can be fully expressed in terms of the linear bias through equations (2.75) and (2.74).

Finally, the resulting galaxy power spectrum $P_{\rm gg}$ can be written down as:

$$\begin{aligned} P_{\rm gg}(k) = {} & b_1^2 P_{\rm mm}(k) + b_1 b_2 P_{b_1 b_2}(k) + b_1 \gamma_2 P_{b_1 \gamma_2}(k) \\ & + b_2^2 P_{b_2 b_2}(k) + b_2 \gamma_2 P_{b_2 \gamma_2}(k) + \gamma_2^2 P_{\gamma_2 \gamma_2}(k) \\ & + b_1 \gamma_{21} P_{b_1 \gamma_{21}}(k) \end{aligned} \tag{2.76}$$

and the one-loop bias corrections are:

$$P_{b_1 b_2}(k) = 2 \int_{\boldsymbol{q}} F_2(\boldsymbol{k} - \boldsymbol{q}, \boldsymbol{q}) P_{\rm lin}(|\boldsymbol{k} - \boldsymbol{q}|) P_{\rm lin}(q), \tag{2.77}$$

$$\begin{aligned} P_{b_1 \gamma_2} = {} & 4 \int_{\boldsymbol{q}} F_2(\boldsymbol{k} - \boldsymbol{q}, \boldsymbol{q}) S(\boldsymbol{k} - \boldsymbol{q}, \boldsymbol{q}) P_{\rm lin}(|\boldsymbol{k} - \boldsymbol{q}|) P_{\rm lin}(q) \\ & + 8 P_{\rm lin}(k) \int_{\boldsymbol{q}} G_2(\boldsymbol{k}, \boldsymbol{q}) S(\boldsymbol{k} - \boldsymbol{q}, \boldsymbol{q}) P_{\rm lin}(q), \end{aligned} \tag{2.78}$$

$$P_{b_2 b_2}(k) = \frac{1}{2} \int_{\boldsymbol{q}} [P_{\rm lin}(|\boldsymbol{k} - \boldsymbol{q}|) P_{\rm lin}(q) - P_{\rm lin}^2(q)], \tag{2.79}$$

$$P_{b_2\gamma_2} = 2\int_{\boldsymbol{q}} S(\boldsymbol{k}-\boldsymbol{q},\boldsymbol{q})P_{\rm lin}(|\boldsymbol{k}-\boldsymbol{q}|)P_{\rm lin}(q), \tag{2.80}$$

$$P_{b_2\gamma_2} = 2\int_{\boldsymbol{q}} S^2(\boldsymbol{k}-\boldsymbol{q},\boldsymbol{q})P_{\rm lin}(|\boldsymbol{k}-\boldsymbol{q}|)P_{\rm lin}(q), \tag{2.81}$$

$$P_{b_1\gamma_{21}} = 4P_{\rm lin}\int_{\boldsymbol{q}} S(\boldsymbol{k}-\boldsymbol{q},\boldsymbol{q})S(\boldsymbol{k},\boldsymbol{q})P_{\rm lin}(q). \tag{2.82}$$

Here, F_2 and G_2 are the symmetrised mode-coupling kernels and $S(\boldsymbol{k}_1,\boldsymbol{k}_2) = (\hat{\boldsymbol{k}}_1\cdot\hat{\boldsymbol{k}}_2)^2 - 1$ is the Fourier transform of the kernel that describes $\mathcal{G}(\Phi_\nu)$.

The galaxy bias expression in equation (2.71) does not include the so-called higher derivative bias, which involves operators higher than the second derivative of the gravitational and velocity potentials and whose leading order scales as $\nabla^2\delta$ (Desjacques, 2008; Desjacques et al., 2010). This is the contribution from the regions where galaxy formation occurs and which are, therefore, influenced by short-range gravitational effects as well as baryonic effects. For halos, such scales are determined by the Lagrangian radius (McDonald & Roy, 2009; Lazeyras & Schmidt, 2019), defined for a halo of mass M as $R(M) = (3M/4\pi\bar{\rho}_{\rm m})^{1/3}$ and below which galaxy formation can no longer be considered to be a local process. The minimum fitting scales $s = 20h^{-1}$Mpc used in the analyses presented in Chapters 3 and 4 are large enough that this contribution is unlikely to be important (as also confirmed in the model validation, where we are able to recover the input cosmology of galaxy mocks). Nonetheless, this contribution is accounted for in the theory emulator which we describe in Chapter 5, ensuring that its predictions can potentially be used down to small scales.

2.5.2 Redshift-Space Distortions

In addition to tracer bias, another effect to consider when modelling observed galaxy clustering is the redshift-space distortions (RSD), which arise due to the fact that the measured galaxy redshift is not due to the cosmic expansion alone.

Even in comoving coordinates galaxies are not static tracers but possess a velocity induced by the gravitational potential of their environment - a *peculiar velocity*. In general, on large scales galaxies tend to exhibit a coherent motion from underdense areas towards overdensities, as the large-scale structure continues to grow. In addition to this, on small scales, highly overdense virialized regions like galaxy clusters are characterised by large velocity dispersions.

The observed galaxy redshift, $z_{\rm obs}$, will then, in addition to the redshift due to the cosmic expansion, $z_{\rm cos}$, receive a contribution from the parallel component of the galaxy peculiar velocity, $v_\parallel$, such that:

$$1 + z_{\rm obs} = (1 + z_{\rm cos})\Big(1 + \frac{v_\parallel}{c}\Big). \tag{2.83}$$

Equivalently, the galaxy position inferred from the redshift (or "in redshift space"), s, will be modified as:

$$s = \chi(z_{\rm obs}) = \chi(z_{cos}) + \frac{d\chi}{dz}(z)\Delta z = \chi(z_{cos}) + \frac{(1+z)}{H(z)}v_{\parallel}. \tag{2.84}$$

The galaxy will, therefore, seem displaced along the line of sight, while its angular position in the sky remains the same. As a result, the overall observed clustering pattern will become direction-dependent (or anisotropic) and the amount of the anisotropy will depend on the peculiar velocities that sourced it. Measuring the distortions in the galaxy clustering pattern is, therefore, a powerful tool to probe the growth of structure.

RSD in linear regime

To see how the galaxy power spectrum is modified due to the RSD in linear regime (i.e. due to the galaxy bulk flow peculiar velocity), we can start by relating the real and redshift spaces through the fact that the total number of galaxies and, therefore also the total mass enclosed in a volume element is conserved:

$$(1 + \delta^{\rm s}(\boldsymbol{s}))d^3s = (1 + \delta(\boldsymbol{r}))d^3r, \tag{2.85}$$

where $\delta^{\rm s}(s)$ marks density fluctuations in redshift and $\delta(r)$ in real space (we have also assumed that the mean number density remains the same in real and redshift space). The transformation of the volume element from real to redshift space is encoded in the Jacobian $J = |\partial r_i/\partial s_j|$, which can be obtained from the transformation of equation (2.84):

$$J = \left(1 + \frac{1}{aH(a)}\frac{\partial v_{\parallel}}{\partial r_{\parallel}}\right)^{-1} \simeq \left(1 - \frac{1}{aH(a)}\frac{\partial v_{\parallel}}{\partial r_{\parallel}}\right), \tag{2.86}$$

Plugging this expression for J in the equation (2.85) and keeping only linear terms in perturbations we obtain:

$$\delta^{\rm s}(\boldsymbol{s}) = \delta(\boldsymbol{r}) - \frac{1}{aH(a)}\frac{\partial v_{\parallel}}{\partial r_{\parallel}}. \tag{2.87}$$

Furthermore, in order to be able to describe the changes that the RSD induce in the power spectrum, we need to Fourier transform the above relation to obtain:

$$\delta^{s}(\boldsymbol{k}) = \delta(\boldsymbol{k}) - \frac{ik_{\parallel}}{aH(a)}v_{\parallel}. \tag{2.88}$$

We can additionally make use of the Fourier transformed linearised continuity equation (2.49) to get the expression for velocity in terms of the density fluctuations:

$$\frac{\partial\delta(k,\tau)}{\partial\tau} + k^2 v(k,\tau) = 0. \tag{2.89}$$

Plugging in the growing mode solution $\delta \propto D_1(a)$ from equation (2.52) we get:

$$v = \frac{1}{k^2}\frac{\partial}{\partial\tau}\Big(\frac{\delta(k,\tau)}{D_1}D_1\Big) = \frac{\delta(k,\tau)}{k^2 D_1}\frac{dD_1}{d\tau}. \tag{2.90}$$

It is then convenient to define the dimensionless *linear growth factor* f and re-express everything in terms of the scale factor a [3]:

$$f \equiv \frac{a}{D_1}\frac{dD_1}{da} = \frac{d\ln D_1}{d\ln a}. \tag{2.91}$$

The relation between the density fluctuations and the velocity potential is then:

$$v(k,a) = \frac{f(a)aH(a)}{k^2}\delta(k,a), \tag{2.92}$$

which we can take the gradient of to obtain the relation for the velocity field:

$$v_j(k,a) = i\frac{k_j}{k^2}aHf(a)\delta(k,a). \tag{2.93}$$

This can now be plugged back into equation (2.88) to obtain:

$$\delta^{\mathrm{s}}(\boldsymbol{k}) = \delta(\boldsymbol{k})\Big(1 + f(a)\mu_k^2\Big) \tag{2.94}$$

with $\mu_k = k_\parallel/k$ as the cosine of the angle between the wavevector $\boldsymbol{k}$ and the line of sight direction. The relation above, therefore, tells us that density fluctuations when observed in the redshift space are enhanced by the angle-dependent factor $f\mu_k$, such that the maximum enhancement is present along the line of sight (where $\mu_k = 1$) and the real space density fluctuation amplitude is recovered in the direction perpendicular to the line of sight (where $\mu_k = 0$). The result of the RSD in the linear regime is then a clustering pattern which is squished along the line of sight.

The relation we just derived is applicable for the matter fluctuations - in order to obtain an equivalent expression for galaxies we need to take bias into account. In the simplest case, considering only linear bias for the density fluctuations $\delta_{\mathrm{g}} = b_1\delta_{\mathrm{m}}$ and assuming no velocity bias (so that $\boldsymbol{v}_{\mathrm{g}} = \boldsymbol{v}_{\mathrm{m}}$), equation (2.94) becomes:

$$\delta^{\mathrm{s}}_{\mathrm{g}}(\boldsymbol{k}) = \delta_{\mathrm{g}}(\boldsymbol{k})\Big(1 + \beta(z)\mu_k^2\Big) \tag{2.95}$$

with

$$\beta(z) = \frac{f(z)}{b}. \tag{2.96}$$

The resulting galaxy power spectrum will then simply be given by:

$$P^{\mathrm{s}}_{\mathrm{g}}(k,\mu_k) = b_1^2(1 + \beta\mu_k^2)^2P(k). \tag{2.97}$$

[3] By making use of $d/d\tau = a^2Hd/da$.

The prefactor $(1+\beta\mu_k^2)^2$ is also referred to as *the Kaiser factor*. It is furthermore common to rewrite the relation for the power spectrum as:

$$P_{\mathrm{g}}^{\mathrm{s}}(k,\mu_k) = \Big(b_1\sigma_8(z) + f\sigma_8(z)\mu_k^2\Big)^2\Big(\frac{P(k,z)}{\sigma_8^2(z)}\Big), \tag{2.98}$$

where σ_8 is the linear density field variance as measured on a scale of $8h^{-1}$Mpc. Traditionally, σ_8 is taken to represent the amplitude of the power spectrum, such that $P(k,z)/\sigma_8^2(z)$ is supposed to depend on the shape of $P(k,z)$ only, and the RSD-induced anisotropies are described by the parameter combinations $b\sigma_8(z)$ and $f\sigma_8(z)$. Nonetheless, as argued by Sánchez (2020), due to the fact that σ_8 is defined on a scale that depends on h, this parameter does not actually represent the power spectrum amplitude, as will be discussed in Section 2.6.1. Presently, we will simply note that for a correct version of equation (2.98), one should replace σ_8 with an equivalent parameter measured on a scale defined in Mpc.

RSD in the non-linear regime

The derivation presented in the previous sections is only valid in the linear regime, i.e., where densities and peculiar velocities can be assumed to be small. Nonetheless, we already hinted at the fact that galaxy peculiar velocities are expected to be large around high-density regions, such as galaxy clusters. As a result, the linear RSD description is not sufficient in order to accurately describe the results from modern galaxy surveys. In particular, the linear theory fails to capture not only the galaxy motion within virialized structures, but it also fails to take into account the non-linear coupling between density and velocity fields that is induced by non-linear evolution. In this section, we, therefore, present a description of non-linear redshift space power spectrum which follows Sánchez et al. (2017), Scoccimarro (2004) and Taruya et al. (2010).

One of the ways to describe the non-linear redshift space power spectrum follows a simple ansatz:

$$P_{\mathrm{g}}(k,\mu_k) = F_{\mathrm{FoG}}(ifk\mu_k)P_{\mathrm{novir}}(k,\mu_k). \tag{2.99}$$

Here, P_{novir} is the redshift-space power spectrum that includes the effect of the non-linear bulk flow of matter on larger scales and $F_{\mathrm{FoG}}(ifk\mu)$ is the *fingers-of-god* factor, which accounts for the small-scale RSD.

The non-linear anisotropic power spectrum P_{novir} can be written down as a sum:

$$P_{\mathrm{novir}}(k,\mu_k) = P_{\mathrm{novir}}^{(1)}(k,\mu_k) + P_{\mathrm{novir}}^{(2)}(k,\mu_k) + P_{\mathrm{novir}}^{(3)}(k,\mu_k). \tag{2.100}$$

The first term here is the non-linear version of the Kaiser formula presented in equation (2.97):

$$P_{\mathrm{novir}}^{(1)}(k,\mu_k) = P_{\mathrm{gg}}(k) + 2f\mu_k^2P_{\mathrm{g}\theta}(k) + f^2\mu_k^4P_{\theta\theta}(k). \tag{2.101}$$

Here, the term is made up of contributions from the galaxy-galaxy power spectrum $P_{\mathrm{gg}}(k) = \langle\delta_{\mathrm{g}}(\boldsymbol{k}),\delta_{\mathrm{g}}(\boldsymbol{k'})\rangle$, the galaxy-velocity power spectrum $P_{\mathrm{g}\theta}(k) = \langle\delta_{\mathrm{g}}(\boldsymbol{k}),\theta(\boldsymbol{k'})\rangle$ and the velocity-velocity power spectrum $P_{\theta\theta}(k) = \langle\theta(\boldsymbol{k}),\theta(\boldsymbol{k'})\rangle$. We have already discussed how we may

obtain a prediction for $P_{\rm gg}$ by taking a $P_{\rm mm}$ prediction and a recipe for galaxy bias, however, our chosen theory model RESPRESSO does not provide recipes for computing $P_{{\rm g}\theta}$ or $P_{\theta\theta}$. We, therefore, model these power spectra using empirical relations proposed by Bel et al. (2019) from the Dark Energy and Massive Neutrinos Universe (DEMNUni) set of N-body simulations (Carbone et al., 2016).

The second term of $P_{\rm novir}(k,\mu_k)$ is given as:

$$P^{(2)}_{\rm novir}(k,\mu_k) = \int d^3q \frac{q_\parallel}{q^2}[B_\sigma(\boldsymbol{q},\boldsymbol{k}-\boldsymbol{q},-\boldsymbol{k}) - B_\sigma(\boldsymbol{q},-\boldsymbol{k},\boldsymbol{k}-\boldsymbol{q})], \tag{2.102}$$

where B_σ is the cross bispectrum between the density and velocity field with $\sigma(\boldsymbol{k}) = \delta_{\rm g}(\boldsymbol{k}) + f(k_\parallel/k^2)\theta(\boldsymbol{k})$ such that:

$$\langle\theta(\boldsymbol{k}_1)\sigma(\boldsymbol{k}_2)\sigma(\boldsymbol{k}_3)\rangle = (2\pi)^3\delta_{\rm D}(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)B_\sigma(\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{k}_3) \tag{2.103}$$

and the cross bispectra are calculated from the tree-level PT for the density and velocity including bias up to γ_2. The final term required to obtain $P_{\rm novir}$ is given by:

$$P^{(3)}_{\rm novir}(k,\mu_k) = \int d^3q \frac{q_\parallel(k_\parallel - q_\parallel)}{q^2(\boldsymbol{k}-\boldsymbol{q})^2}(b_1+f\mu_q^2)(b_1+f\mu_{k-q}^2)P_{\delta\theta}(k-q)P_{\delta\theta}(q). \tag{2.104}$$

The non-virial power spectrum is additionally corrected by the factor $F_{\rm FoG}$, which accounts for the contribution from random motions on small scales (however, it enters the redshift-space power spectrum model as a multiplicative factor, since the small scale velocities also affect the observed galaxy distribution on large scales). The FoG effect shows up as an elongation of the clustering pattern along the line of sight and is often modelled using a Gaussian function. However, as pointed out in Scoccimarro (2004), the effect of the non-linear corrections is that the large-scale limit of the pairwise velocity distribution function is non-Gaussian. Therefore, following the analysis of Sánchez et al. (2017), we adopt the FoG correction of the form:

$$F_{\rm FoG}(ifk\mu_k) = \frac{1}{\sqrt{1-(ifk\mu_k)^2a_{\rm vir}^2}}\exp\left(\frac{(ifk\mu_k)^2\sigma_v^2}{1-(ifk\mu_k)^2a_{\rm vir}^2}\right), \tag{2.105}$$

where $a_{\rm vir}$ is a free parameter characterizing the kurtosis of the small-scale velocity distribution, and $\sigma_{\rm v}$ is the one-dimensional linear velocity dispersion defined in terms of the linear matter power spectrum as

$$\sigma_v^2 \equiv \frac{1}{6\pi^2}\int {\rm d}k\, P_{\rm L}(k). \tag{2.106}$$

2.5.3 Anisotropic two-point correlation function

In order to make use of the structure growth information imprinted by the RSD, we need to be able to measure how the two-point functions depend on μ (i.e., the angle averaged

measurements $\xi(s)$ and $P(k)$ do not carry the full constraining power available). Nonetheless, direct measurements of $\xi(s,\mu)$ and $P(s,\mu)$ do not have a high enough signal-to-noise ratio for cosmological analyses. An alternative method that preserves the anisotropic information present in the original two-point function is to express the measurements as a reduced set of one-dimensional projections. The two commonly used types of projections that we will briefly describe here are *Legendre multipoles* and *galaxy clustering wedges*.

Legendre multipoles, $P_\ell(k)$, of the power spectrum are given by:

$$P_\ell(k) = \frac{(2\ell+1)}{2}\int_{-1}^{1} P(k,\mu_k)L_\ell(\mu_k)d\mu_k, \tag{2.107}$$

where $L_\ell(\mu_k)$ are Legendre polynomials and the full anisotropic redshift space galaxy power spectrum $P_{\rm g}^s(k,\mu_k)$ can then be expressed as the series:

$$P_{\rm g}^s(k,\mu_k) = \sum_{\ell,\rm even} P_\ell(k)L_\ell(\mu_k). \tag{2.108}$$

The sum is performed over even ℓ multipoles only, because all of the odd ℓ multipoles are zero (since $P_{\rm g}^s(k,\mu_k)$ is an even function of μ_k). The terms with $\ell = 0, 2, 4$ are referred to as monopole, quadrupole and hexadecapole and l indicates the maximum power of μ_k appearing in $L_\ell(\mu_k)$, such that the monopole corresponds to the angle-averaged power spectrum $P(k)$ and the subsequent multipoles carry the information encoded in the anisotropies. Relations equivalent to those presented in equations (2.107) and (2.108) also hold for the two-point correlation function.

Alternatively, we can perform the measurements of the average of $\xi(s,\mu)$ over a given interval $\Delta\mu = \mu_{\rm max} - \mu_{\rm min}$, as proposed by Kazin et al. (2012). This statistic is referred to as galaxy clustering wedges and is given by:

$$\xi_{\Delta\mu}(s) \equiv \frac{1}{\Delta\mu}\int_{\mu_{\rm min}}^{\mu_{\rm max}} \xi(\mu,s)d\mu \tag{2.109}$$

Galaxy clustering wedges and Legendre multipoles carry equivalent information and the two statistics are related by:

$$\xi_{\Delta\mu}(s) = \sum_\ell \xi_\ell(s)\bar{L}_\ell, \tag{2.110}$$

where $\bar{L}_\ell$ is the average of the Legendre polynomial of order ℓ over the μ-bin of the clustering wedge. Regardless of the same information content, clustering wedges and multipoles have different noise properties and Hou et al. (2018) showed that for eBOSS quasar clustering analysis the multipoles provide tighter cosmological constraints. For this reason, the work presented in this thesis (Chapters 3 and 4) follows the original analyses for BOSS galaxies by Sánchez et al. (2017) and eBOSS quasars by Hou et al. (2021) and uses clustering wedges for the former and Legendre multipoles for the latter data set.

2.5.4 Alcock-Paczynski distortions

The final effect to take into account when modelling the observed two-point statistics arises due to the mismatch between the true cosmology and the one that is used to convert the observed redshifts into distances. This is referred to as the *Alcock-Paczynski* (AP) distortions (Alcock & Paczyński, 1979) and can be written down as a rescaling of the distance components parallel and perpendicular to the line of sight $s_\parallel$ and $s_\perp$:

$$s_\parallel = q_\parallel s'_\parallel, \qquad\qquad s_\perp = q_\perp s'_\perp, \tag{2.111}$$

where the primes denote the distances in the fiducial cosmology and the scaling factors $q_\parallel$ and $q_\perp$ are the *geometric distortion parameters* evaluated at the effective redshift of the sample $z_{\rm eff}$:

$$q_\parallel = \frac{H'(z_{\rm eff})}{H(z_{\rm eff})}, \qquad\qquad q_\perp = \frac{D_{\rm M}(z_{\rm eff})}{D'_{\rm M}(z_{\rm eff})}. \tag{2.112}$$

As a result, when modelling the anisotropic two-point correlation function $\xi(s,\mu)$, we need to rescale the total separation $s = \sqrt{s_\parallel^2 + s_\perp^2}$ and the cosine of the angle between the separation vector $\boldsymbol{s}$ and the line of sight μ as:

$$s = s' \left(q_\parallel^2 \mu'^2 + q_\perp^2 (1-\mu'^2) \right), \tag{2.113}$$

$$\mu = \mu' \frac{q_\parallel}{\sqrt{q_\parallel^2 \mu'^2 + q_\perp^2 (1-\mu'^2)}}. \tag{2.114}$$

2.6 Evolution Mapping

Throughout this chapter, we have shown how the matter power spectrum, the main ingredient required to model the large-scale-structure observations, encodes the information on the cosmological parameters which describe the physics of our Universe. Nevertheless, the choice of a set of parameters that should characterise the cosmology and, therefore, the resulting structure is not unique. Given a number of different possible parameterizations, one may, therefore, ask if there is a parameter space that is particularly suited to clearly describe the effect that each cosmological parameter has on the power spectrum. Sánchez (2020) and Sánchez et al. (2022) proposed exactly such a parametrization and showed that it not only allows for a clearer interpretation of cosmological observations results but also greatly facilitates the modelling of the matter power spectrum through a technique referred to as *evolution mapping*. The observational implications of these results are mainly discussed in Chapters 3 and 4, whereas the evolution mapping approach itself is a key ingredient in building the emulator described in Chapter 5. In this section, we summarise the main ideas behind evolution mapping.

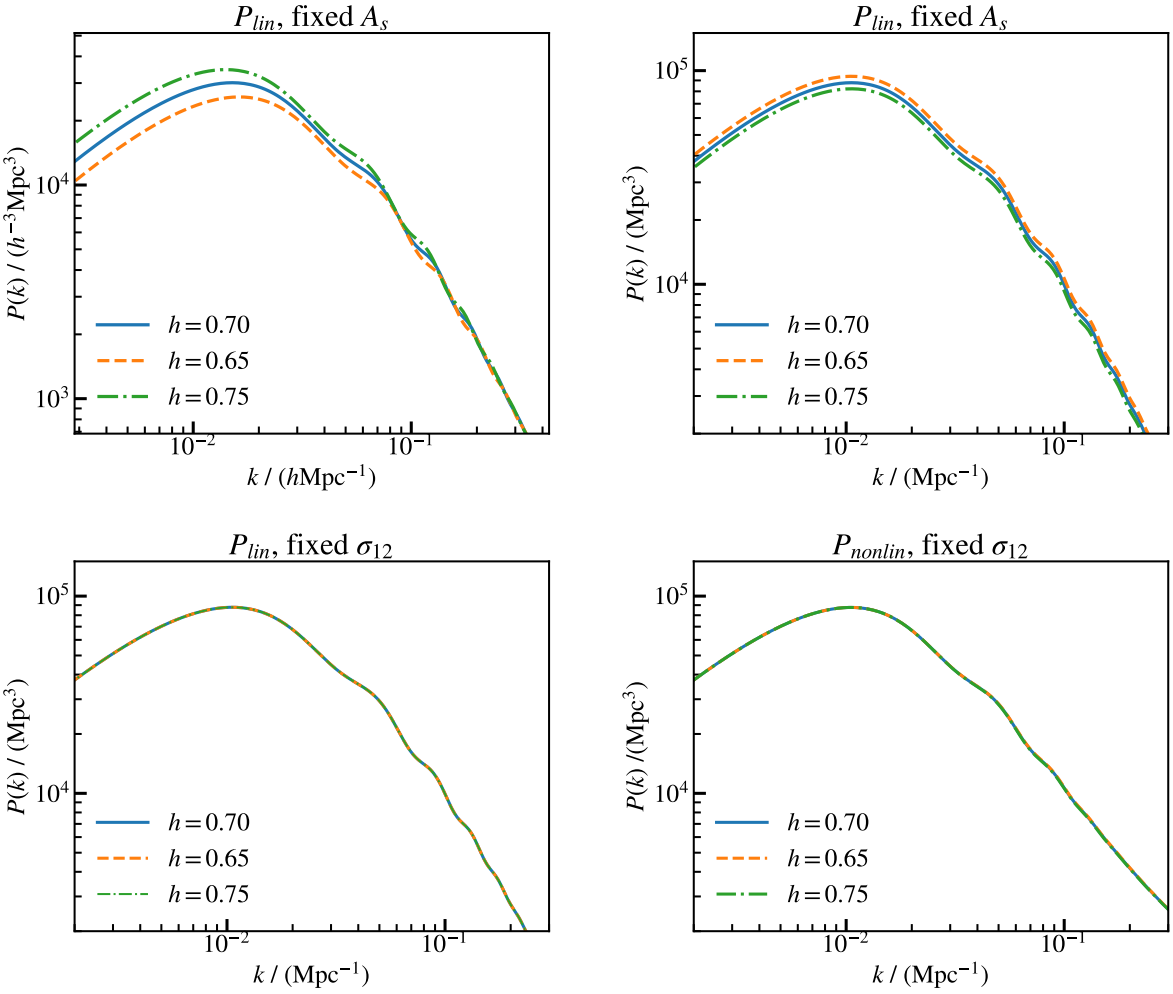

Figure 2.3: *Top:* Linear matter power spectra at $z = 0$ for three ΛCDM cosmologies with the same values of $\omega_b, \omega_c, \omega_\nu, A_s$ and n_s but different values of h, expressed in h^{-1}Mpc units (left) and in Mpc units (right). It is clear that h^{-1}Mpc units introduce unphysical shifts in the power spectrum features. *Bottom:* Matter power spectra at $z = 0$ for the same cosmologies as the ones shown in the top panels, except here A_s is chosen so that all three cosmologies have the same value of σ_{12}. The comparison is shown both for linear (left) and non-linear (right) matter power spectra.

2.6.1 Characterising the amplitude of the power spectrum

In Section 2.4 we described the power spectrum as a volume-averaged measurement, which means that it will inherit whatever distance units we use to define this volume. We furthermore already noted, in Section 2.3, that it is very common to express cosmological distances in h^{-1}Mpc units, which is also true when characterising the power spectrum: it is common to see $P(k)$ expressed in h^{-3}Mpc on k scales defined in hMpc^{-1}. This choice of the unit system, however, results in shifts of the features of the power spectrum, which are related not to any changes in physics, but simply a changing value of h, as first pointed out in Sánchez (2020).

This point is illustrated in the top panels of Figure 2.3, which shows the linear power spectrum computed for different cosmologies that share the same initial amplitude $A_{\rm s}$ as well as the same values of $\omega_{\rm b}$, $\omega_{\rm c}$, ω_ν and $n_{\rm s}$ expressed in Mpc (right) and h^{-1}Mpc (left) units. In these two cases, the only difference in the two cosmologies is the dark energy density, which has no effect on the BAO signal. This is reflected in the power spectra expressed in Mpc units: the three power spectra differ in amplitude only, with the highest h cosmology showing the smallest amplitude, as faster expansion results in a more significant clustering suppression. In contrast, the power spectra expressed in h^{-1}Mpc units not only show shifts in the BAO features, but additionally are in the reverse order in terms of their amplitudes: the power spectrum corresponding to the cosmology with the highest dark energy density and, therefore, the least clustering, has the greatest amplitude. The consequence of using h^{-1}Mpc units is then the obscurity of the true effect that h has on the power spectrum.

This issue translates more directly into the cosmological parameter space as well. The power spectrum amplitude is one of the parameters required to specify a cosmology. A possible choice is the initial amplitude of the fluctuation spectrum, $A_{\rm s}$, however, alternatively, one might want to use the amplitude of the power spectrum today. Traditionally, this is characterised by the parameter σ_8, which represents the linear power spectrum amplitude as measured on a scale of $R = 8\,h^{-1}$Mpc. In practice, σ_R represents the amplitude of a smoothed density field, or the average of density fluctuations within a sphere of radius R and can be defined as a convolution of the power spectrum and a window function W_k:

$$\sigma^2(R) = \int \Delta^2_{\rm lin}(k)|W_k|^2 d\ln k. \tag{2.115}$$

Here $\Delta^2_{\rm lin}(k)$ is the dimensionless linear power spectrum, such that:

$$\Delta^2_{\rm lin}(k) \equiv \frac{1}{(2\pi)^3} 4\pi k^3 P_{\rm lin}(k) \tag{2.116}$$

and the window function takes the form of:

$$W_k = \frac{3}{(kR)^3}\Big(\sin(kR) - kR\cos(kR)\Big). \tag{2.117}$$

The fact that σ_R is defined in terms of the dimensionless power spectrum, means that the issue of units is only introduced when defining the smoothing scale R. The problem with

σ_8 is then simply the fact that it is measured on a scale that is defined in terms of h^{-1}Mpc, which means that this parameter does not capture the amplitude of the power spectrum in a consistent way.

An alternative to σ_8 was proposed by Sánchez (2020) who suggested to replace this parameter with σ_{12} - an equivalent σ_R measurement with $R = 12$ Mpc. Here the value of R is defined in Mpc units and the value of 12 Mpc is chosen such that σ_{12} is of a similar magnitude as σ_8 for the standard value of $h = 0.67$. We can demonstrate that σ_{12} correctly describes the power spectrum amplitude, as shown in the bottom panels of Figure 2.3. Here the power spectra from the top panels of the same figure are computed for values of A_s that are adjusted to give the same σ_{12}. We can see that, indeed, the power spectra for all cosmologies show an excellent agreement both for linear and non-linear cases. Although not seen by eye, it is important to note that the agreement is perfect in the linear case and with only a few percent difference at high k for the non-linear case. We can, therefore, conclude that σ_{12} captures the degenerate effect that A_s and h have on the power spectrum when it is expressed in Mpc units.

The use of particular units does not alter the information content of the power spectrum, however, the cosmological constraints on σ_8 and σ_{12} differ in important ways, as discussed in Chapters 3 and 4. In addition to this, being able to correctly describe the power spectrum amplitude allows us to easily map between cosmologies that are characterised by the power spectra with identical shapes as we will discuss in the next section.

2.6.2 Evolution and shape parameters

In the previous section, we saw that, when Mpc units are used and the rest of the cosmological parameters are held fixed, A_s and h follow a degeneracy that accurately describes the amplitude of the matter power spectrum. In order to be able to perform evolution mapping we must proceed to similarly define a cosmological parameter space where each parameter's impact on the power spectrum can be clearly characterised.

Once again, in order to be able to do so, it is crucial that none of the quantities used are expressed through h. In addition to substituting σ_8 with σ_{12}, it is then also necessary to use the correct parameterization for the homogeneous background evolution. In Section 2.2 we introduced the concept of physical densities ω_i which in evolution mapping replace the relative densities Ω_i and are related to them through h: $\omega_i = \Omega_i h^2$.

Using this parametrisation and considering the dimensionless linear matter power spectrum (or, equivalently, the linear power spectrum expressed in Mpc units), we can proceed to classify the cosmological parameters into *shape* and *evolution*, based on the effect that they have on $\Delta^2_\mathrm{lin}(k)$, as proposed by Sánchez et al. (2022). The shape parameters (Θ_s) characterise the shape of the primordial power spectrum and the transfer function. They are:

$$\Theta_\mathrm{s} = (\omega_\gamma, \omega_\mathrm{b}, \omega_\mathrm{c}, n_\mathrm{s}, ...). \tag{2.118}$$

At the linear level, the rest of the parameters then only affect the amplitude of $\Delta^2_\mathrm{lin}(k)$.

These are referred to as evolution parameters ($\Theta_{\rm e}$) and include:

$$\Theta_{\rm e} = (A_{\rm s}, \omega_{\rm K}, \omega_{\rm DE}, w_{\rm DE}(a), ...). \tag{2.119}$$

Here, $w_{\rm DE}(a)$ includes all possible parametrizations of the dark energy equation of state parameter $w_{\rm DE}$. It is important to note that the physical neutrino density ω_ν cannot be characterised as either shape or evolution parameter because, already at the linear level, it introduces scale dependence of the growth factor. As a result, an additional correction needs to be applied in order to use evolution mapping to model cosmologies with massive neutrinos. For this reason, for the rest of this section, we will proceed to only consider cosmologies with $\omega_\nu = 0$. In Section 2.6.1, we showed how varying h only changed the amplitude of the power spectrum, however, this illustrated a special case where we fixed $\omega_{\rm m}$ and, therefore, effectively only varied the evolution parameter $\omega_{\rm DE}$. In general, h represents a mixture of shape and evolution parameters because it is obtained by summing the physical densities of all energy species. As a result, any parameter defined through h will represent a mixture of shape and evolution parameters as well.

The most precise constraints on shape parameters are provided by the CMB measurements. Nevertheless, the probe has little constraining power on the evolution parameters in general cosmologies. The large-scale structure data, on the other hand, exhibits complimentary constraining properties: it provides only weak constraints on the shape parameters but, when additional information on these parameters is included (either by adding a constraining data set or by imposing an informative prior), it can provide precise measurements of the evolution parameters.

This parameter classification can be furthermore used to perform an accurate mapping between cosmologies that share the same shape parameters. This procedure is based on the fact that, for a given set of shape parameters, any combination of evolution parameters that correspond to the same clustering amplitude, results in an identical linear power spectrum. As discussed in the previous section, a convenient way to characterise the power spectrum amplitude and, therefore, the degeneracy between the evolution parameters, is to use σ_{12}, so that:

$$\Delta^2_{\rm lin}(k|z, \Theta_{\rm s}, \Theta_{\rm e}) = \Delta^2_{\rm lin}(k|\Theta_{\rm s}, \sigma_{12}(z, \Theta_{\rm s}, \Theta_{\rm e})), \tag{2.120}$$

which we will refer to as the evolution mapping relation. One can, therefore, obtain a prediction for the linear power spectrum for any combination of evolution parameters at any redshift by simply taking a $\Delta^2_{\rm lin}$ prediction for a cosmology with matching shape parameters and the appropriate value of σ_{12}.

The evolution mapping relation also applies for the non-linear matter power spectrum obtained from perturbation theory calculations, assuming that the perturbation kernels are independent of cosmology, which has been shown to be a good approximation for a wide variety of cosmologies (Garny & Taule, 2021; Takahashi, 2008; Taruya, 2016). This is because, following the SPT approach, the non-linear $P(k)$ is a function of the linear power spectrum (Scoccimarro et al., 1998). As an example, we can consider one of the perturbative recipes, RPT, which we introduced in Section 2.4.5. As stated in equation (2.67), following this approach, the non-linear matter power spectrum is written as a sum

of the mode coupling term $P_{\rm MC}(k,a)$ and the product of the linear power spectrum and the propagator term $P_{\rm lin}G^2(k,a)$. The product term simply represents the contribution from the linearly-evolved power from each individual k, while the mode-coupling term accounts for the contribution from all other scales in the linear power. Both of these terms can be expressed in terms of contributions that are functions of $P_{\rm lin}$, which means that any model with identical $P_{\rm lin}$ will also lead to the same RPT prediction for the non-linear matter power spectrum. The same is true for any other recipe, in general, where $P(k)$ is defined as a function of the linear power spectrum only. The evolution mapping relation can, furthermore, be adapted to obtain predictions in redshift space and for tracer power spectra, as we will explicitly describe in Section 5.1.

Given a power spectrum for a single cosmological model, evolution mapping, therefore, provides a simple recipe to explore a range of different cosmologies that share the same $\Theta_{\rm s}$. This is extremely convenient when building an emulator - a sophisticated interpolation scheme that, once trained on theory models calculated for sample points in a parameter space, allows the user to obtain a theory prediction for any point within it at a fraction of the time it would take to evaluate the full model. The accuracy of the emulation, however, generally decreases when increasing the parameter space, which is why the evolution mapping approach is so useful when designing an emulator: it reduces the number of parameters required to fully specify the power spectrum as well as providing a recipe to easily map a power spectrum prediction to any desired redshift. In Chapter 5 we discuss the use of evolution mapping for emulation in greater detail and present the work on building an evolution mapping-based emulator of the two-point correlation function.

Chapter 3

Full shape analysis of BOSS and eBOSS QSO

The material in this chapter concerns the full shape analysis of the anisotropic clustering in BOSS galaxy and eBOSS quasar samples and was first presented in Semenaite et al. (2022).

In the work presented in this chapter, we are interested in building upon Tröster et al. (2020), who performed a full-shape analysis of BOSS galaxy clustering wedges on its own and in combination with weak lensing measurements from KiDS-450 (see also Chapter 1 for a summary of the main results). We wish to explore, whether the discrepancy between the low-redshift probes and Planck within the ΛCDM model holds when extending the redshift range probed by the clustering measurements with the addition of eBOSS quasar clustering. We provide the joint constraints from the full shape analysis of BOSS galaxy and eBOSS quasar clustering on their own, as well as in combination with weak lensing information. For our weak lensing data set we use the 3×2pt measurements from the Dark Energy Survey Year 1 (DES Y1, Abbott et al., 2018) release, which both cover a larger area than KV450 and include galaxy clustering and galaxy-galaxy lensing as well as the shear-only measurements. If the tension seen between the low-redshift probes and Planck is purely statistical, adding more data should not only tighten the posterior contours but be able to bring the constraints to a better agreement. The results from an equivalent analysis with KiDS-450 shear measurements are available in the Appendix A.

In addition to expanding our data sets we also aim to re-define the parameter space following Sanchez et al. (2021), who distinguish 'shape' and 'evolution' cosmological parameters. This classification is introduced to describe the degenerate way in which evolution parameters affect the linear matter power spectrum when expressed in Mpc units, as discussed in Section 2.6. While in Sanchez et al. (2021) the h-independent parameter space is presented to create a framework which allows them to reduce the number of parameters required to model the cosmology dependence of the matter power spectrum, the advantage of such parameter choice for this work is two-fold. First, the derived constraints do not depend on the posterior of h of the particular analysis and can, therefore, be directly compared with constraints from other data sets and, second, the effect that each of the

parameters has on the power spectrum is clear, with evolution parameters affecting its amplitude and shape parameters determining the shape.

This work largely follows the same structure and methods as Tröster et al. (2020) - we assume flat ΛCDM cosmology and obtain the joint low-redshift parameter constraints by combining the likelihoods for each data set considered independently. Our model for anisotropic galaxy and quasar clustering measurements follows that described in Sánchez et al. (2017) for the so-called 'full shape' analysis with some exceptions. First, the non-linear matter power spectrum is obtained using predictions by RESPRESSO (see Section 2.4.5), instead of the gRPT approach used by Sánchez et al. (2017). Second, our description of galaxy bias (provided in Section 2.5.1) includes a slightly different treatment of the tidal bias parameters. Our modelling of RSD effect and the AP distortions matches that of Sánchez et al. (2017) and is as described in Sections 2.5.2 and 2.5.4. For the '3 × 2pt' analysis (galaxy shear, galaxy-galaxy lensing, and galaxy clustering) we use the model described in Abbott et al. (2018).

The structure of this chapter is as follows. We provide a more detailed description of the parameter space we use (including the prior choices) in Section 3.1, with the validation of the model used in this work presented in Section 3.2. A summary of our data is given in Section 3.3.1. We illustrate how our parameter space compares with its h-dependent equivalent in Section 3.4.1, where we also present our cosmological constraints from BOSS and eBOSS. The results obtained when adding DES are further displayed in Section 3.4.3 We finish with a discussion of our results in Section 3.5 and a summary in Section 3.6.

3.1 Parameter spaces and prior ranges

Our goal is to obtain constraints on the parameters of the standard ΛCDM model, which corresponds to a flat universe, where dark energy is characterized by a constant equation of state parameter $w_{\rm DE} = -1$. Following Sanchez et al. (2021), we focus on cosmological parameters that can be classified as either "shape" or "evolution". The former are parameters that control the shape of the linear-theory power spectrum expressed in Mpc units. The latter only affect the amplitude of $P_{\rm L}(k)$ at any given redshift. Assuming a fixed total neutrino mass of $\sum m_\nu = 0.06\,\mathrm{eV}$, the ΛCDM model can be described by the parameters

$$\boldsymbol{\theta} = (\omega_{\rm b}, \omega_{\rm c}, \omega_{\rm DE}, A_{\rm s}, n_{\rm s})\,. \tag{3.1}$$

These are the present-day physical energy densities of baryons, cold dark matter, and dark energy, and the amplitude and spectral index of the primordial power spectrum of scalar perturbations at the pivot wavenumber of $k_0 = 0.05\,\mathrm{Mpc}^{-1}$.

Additional parameters can be derived from the set of equation (3.1). The dimensionless Hubble parameter, h, is defined by the sum of all energy contributions. For a ΛCDM model, this is

$$h^2 = \omega_{\rm b} + \omega_{\rm c} + \omega_\nu + \omega_{\rm DE}. \tag{3.2}$$

It is also common to express the contributions of the various energy components in terms

Table 3.1: Priors used in our analysis. U indicates a flat uniform prior within the specified range. The priors on the cosmological and clustering nuisance parameters match those of Tröster et al. (2020) with the exception of n_s, for which the allowed range is widened. The priors on the nuisance parameters of weak lensing data sets match those of Abbott et al. (2018).

Parameter	Prior
Cosmological parameters	
$\Omega_\mathrm{b}h^2$	$U(0.019, 0.026)$
$\Omega_\mathrm{c}h^2$	$U(0.01, 0.2)$
$100\theta_\mathrm{MC}$	$U(0.5, 10.0)$
τ	$U(0.01, 0.8)$
$\ln(10^{10}\mathrm{A_s})$	$U(1.5, 4.0)$
$\mathrm{n_s}$	$U(0.5, 1.5)$
Clustering nuisance parameters	
b_1	$U(0.5, 9.0)$
b_2	$U(-4.0, 8.0)$
a_vir	$U(0.0, 12.0)$
σ_err (eBOSS only)	$U(0.01, 6.0)$

of the density parameters

$$\Omega_i = \omega_i/h^2, \tag{3.3}$$

which represent the fraction of the total energy density of the Universe corresponding to a given component i. The overall amplitude of matter density fluctuations is often characterized in terms of σ_8, the linear-theory RMS mass fluctuations in spheres of radius $R = 8\,h^{-1}$Mpc. A common property of these parameters is their dependence on the value of h. The issues associated with this dependence are discussed in detail by Sánchez (2020) and can be summarised as follows.

The main consequence of using quantities that depend on h in cosmological analyses is that this complicates the comparison of constraints derived from probes that lead to different posterior distributions on h. This can be illustrated the most straightforwardly when considering σ_8, which is defined in terms of a scale in h^{-1}Mpc units. As done by Sánchez (2020), the one-dimensional marginalised posterior distribution for h can be used to obtain the corresponding posterior for $(8/h)$ Mpc to explore what physical distances this radius corresponds to. Fig. 3.1 repeats this simple exercise for the data sets considered in this work - as expected, the range of scales recovered in each case heavily depends on the type of probe considered (Planck displaying an extremely narrow posterior at the physical scale of approximately 12 Mpc, while the remaining probes cover varying ranges), especially in the case where the posterior of h is simply limited by the prior imposed, as is the case for weak lensing data sets.

The solid line in Fig. 3.2 shows the density field variance σ_R as a function of the scale R in a Planck ΛCDM Universe. The shaded areas indicate the range of physical scales

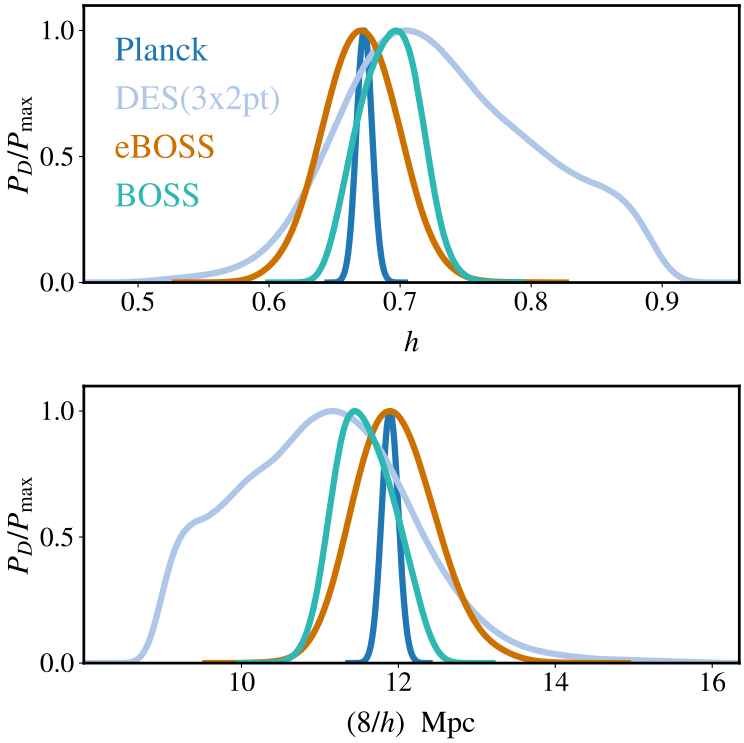

Figure 3.1: Upper panel one-dimensional marginalised posteriors for h for the different data sets considered in this work (with the priors used in this analysis). Lower panel: the corresponding posteriors of the physical value of $(8/h)$ Mpc - the scale used to define σ_8. Any distance defined in h^{-1}Mpc units will correspond to a range of physical scales, as determined by the posterior of h. If the posterior is prior limited, as is the case with weak lensing, the choice of prior will also influence the range of physical scales that contribute to σ_8. On the other hand, the effect is much smaller for the case of narrow Gaussian h-posterior - for Planck σ_8 will correspond to a scale of 12 Mpc.

covered by the the posterior distributions of $(8/h)$ Mpc for DES, BOSS, and Planck shown in Fig. 3.1. The issue with σ_8 is then that its marginalized value corresponds to a weighted average of σ_R on a range of scales that is different for each data set. A further complication is that the value of h also has an impact on the amplitude of σ_R. As discussed in Sánchez (2020), these issues can be avoided by considering the variance of the density field on a reference scale in Mpc such as σ_{12}, which is equivalent to σ_8 but is defined on a physical scale of 12 Mpc. We, therefore, opt to focus on σ_{12} and quantities that carry no explicit dependence on the Hubble constant h in order to enable us to appropriately combine and compare the constraints from our data sets.

We obtain the posterior distribution of all these parameters by performing Monte Carlo Markov chain (MCMC) sampling with CosmoMC (Lewis & Bridle, 2002), which uses CAMB to calculate the linear-theory matter power spectra (Lewis et al., 2000), adapted to compute the theoretical model of our anisotropic clustering measurements. CosmoMC

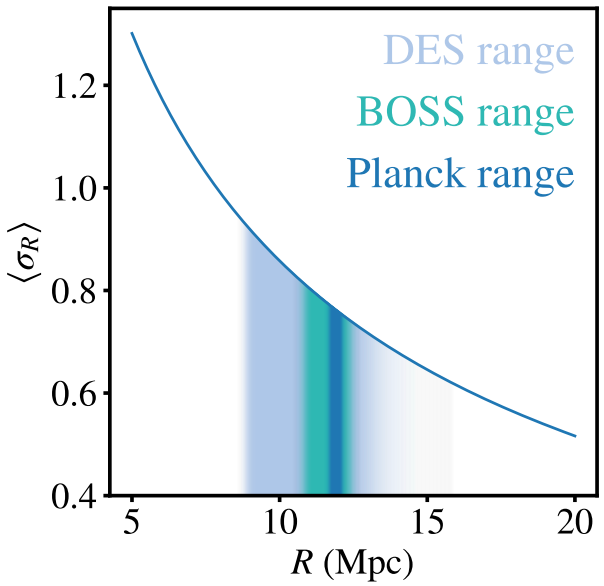

Figure 3.2: The change of the value of standard deviation of linear matter fluctuations σ_R measured in a sphere of physical radius R in Mpc in Planck ΛCDM Universe. The shaded areas indicate the ranges that $(8/h)$ Mpc correspond to for BOSS, DES and Planck based on the posteriors in Fig. 3.1. When R is defined in h^{-1}Mpc, as is the case for σ_8, the value measured is, in fact, a weighted average of σ_R over a range of R.

uses as basis parameters the set

$$\boldsymbol{\theta}_{\rm base} = (\omega_{\rm b}, \omega_{\rm c}, \Theta_{\rm MC}, A_{\rm s}, n_{\rm s})\,, \tag{3.4}$$

where $\Theta_{\rm MC}$ is defined by a factor 100 times the approximate angular size of the sound horizon at recombination. With the exception of the physical baryon density, we assign flat uninformative priors to all the parameters of equation (4.3) as was done in Tröster et al. (2020). Our prior for $\omega_{\rm b}$ has to be restrictive, as our clustering measurements cannot constrain this parameter by themselves. Nevertheless, it is chosen to be approximately 25 times wider than the constraints on this parameter derived from *Planck* data alone (Planck Collaboration et al., 2020). Even though we do not sample the Hubble parameter h, we still need to specify the values allowed - our chosen range $0.5 < h < 0.9$ is wider than that of the KiDS-450 analysis of Hildebrandt et al. (2016) and comparable to the one used in the DES-YR1 fiducial analysis of Abbott et al. (2018). Joudaki et al. (2016) showed that the prior on h has no impact on the significance of the σ_8 tension. We list all the priors used in this analysis in Table 3.1.

As discussed by Sanchez et al. (2021), the effect of all evolution parameters on the linear matter power spectrum is degenerate: for a given set of shape parameters, the linear power spectra of all possible combinations of evolution parameters that lead to the same value of $\sigma_{12}(z)$ are identical. This behaviour is inherited by the non-linear matter power spectrum predicted by RESPRESSO, which depends exclusively on $P_{\rm L}(k)$. However, the full

model of $P(k, \mu)$ does not follow this simple degeneracy due to the effect of bias, RSD and AP distortions. Of the parameters listed in equation (3.1), ω_b, ω_c, and n_s are shape parameters, while ω_DE and A_s are purely evolution parameters. Other quantities such as h or Ω_i represent a mixture of both shape and evolution parameters.

Present-day CMB measurements can constrain the values of most shape parameters with high accuracy, with posterior distributions that are well described by a multivariate Gaussian, independently of the evolution parameters being explored. On the other hand, clustering measurements on their own provide only weak constraints on the values of the shape parameters. However, if the shape parameters are fixed, clustering data can provide precise measurements of the evolution parameters. To test the impact of the additional information on the shape of the linear power spectrum, along with the priors described above, we use another set of priors to explore the constraints on the evolution parameters. For these runs, we impose Gaussian priors on the cosmological parameters that control the shape of the linear power spectrum - ω_b, ω_c, and n_s. We derived the covariance matrix and mean values for these priors from our Planck-only posterior distributions. We refer to these constraints as the 'Planck shape' case.

3.2 Model validation

As we are using an updated prescription for the modelling of both the non-linear matter power spectrum and galaxy bias compared to the previous work of Tröster et al. (2020), we want to assess if it can recover unbiased cosmological parameter estimates, using mock data based on numerical simulations as a testing ground. We do so by applying our model to the mocks that were used for model validation in the original analyses: the Minerva simulations (Grieb et al., 2016; Lippich et al., 2019) for a BOSS-like sample and OuterRim (Heitmann et al., 2019) for an eBOSS-like data set.

Minerva mocks are produced from a set of 300 N-body simulations with 1000^3 particles and a box size of $L = 1.5\,h^{-1}$Gpc. The snapshots at $z = 0.31$ and $z = 0.57$ were used to create halo catalogues with a minimum halo mass of $M_\mathrm{min} = 2.67 \times 10^{12} h^{-1}\mathrm{M}_\odot$, which were populated with synthetic galaxies using the halo occupation distribution (HOD) model by Zheng et al. (2007) with parameters designed to reproduce the clustering properties of the LOWZ and CMASS galaxy samples from BOSS.

The OuterRim (Heitmann et al., 2019) simulation uses $10\,240^3$ dark matter particles to trace the dark matter density field in a $L = 3\,h^{-1}$Gpc size box. We use a set of 100 mock catalogues constructed from the snapshot at $z = 1.433$, which was populated using an HOD model matching the clustering of the eBOSS QSO sample and tested extensively in the mock challenge (labeled as HOD0 in Smith et al., 2020). These realizations include catastrophic redshift failures at a rate of 1.5%, which corresponds to that of the eBOSS quasars.

We measured the mean clustering wedges of the samples from Minerva and the Legendre multipoles from OuterRim with the same binning and range of scales as those of the real data from BOSS and eBOSS and computed their corresponding theoretical covariance

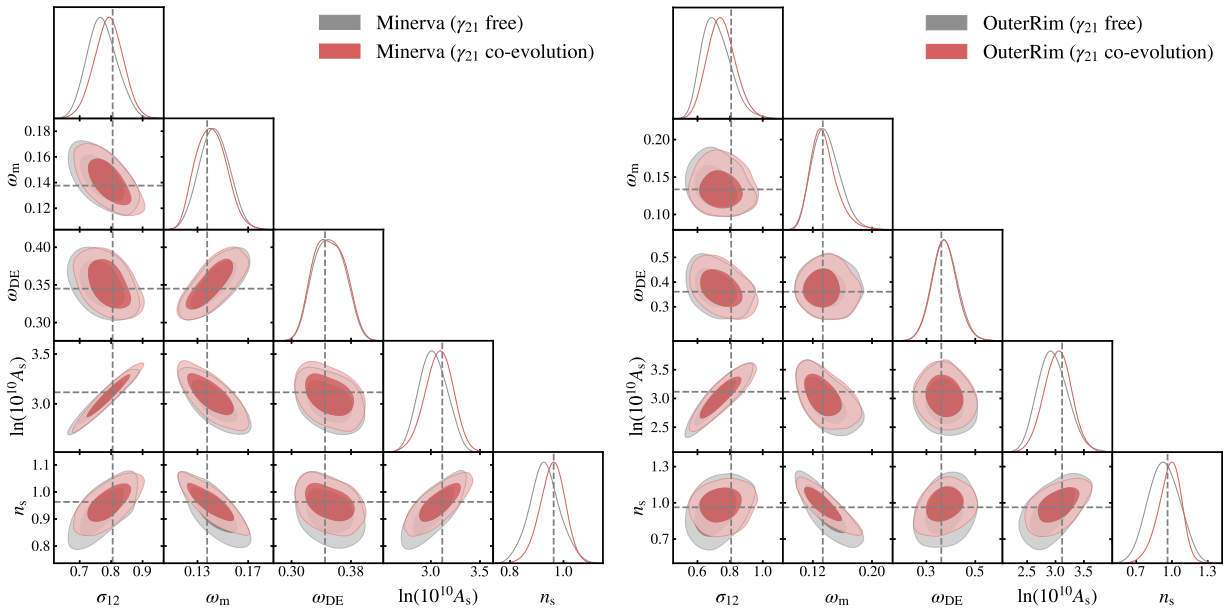

Figure 3.3: Flat ΛCDM constraints derived from mean measurements of Minerva (left) and OuterRim (right) HOD samples while freely varying the non-local bias parameter γ_{21} (grey contours) and when its value is fixed using the co-evolution relation of equation (2.74) (red). The dashed lines mark the true input parameter values. Both cases recover the input cosmology well but the co-evolution relations yield slightly more accurate and precise constraints.

matrices using the Gaussian recipe of Grieb et al. (2016). We analysed these measurements using identical nuisance and cosmological parameter priors as for our final results and tested the validity of our model with and without the assumption of the co-evolution relation for γ_{21} of equation (2.74). We performed a joint fit of the two BOSS-like samples from Minerva while the OuterRim measurements, which correspond to a different cosmology, were analysed separately. Fig. 3.3 shows the posterior distributions recovered from these measurements, which are in excellent agreement with the true input cosmology for all cases (shown by the dashed lines). Nevertheless, we find that setting the value of γ_{21} according to equation (2.74) recovers the true parameter values more accurately for both samples and results in tighter constraints than when it is freely varied. We, therefore, adopt this approach in the analysis of the clustering measurements from BOSS and eBOSS.

3.3 Data

3.3.1 Galaxy and QSO clustering measurements

The Sloan Digital Sky Survey (SDSS) has mapped the large-scale structure of the Universe thanks to the accurate measurements by the double-armed spectrographs (Smee et al.,

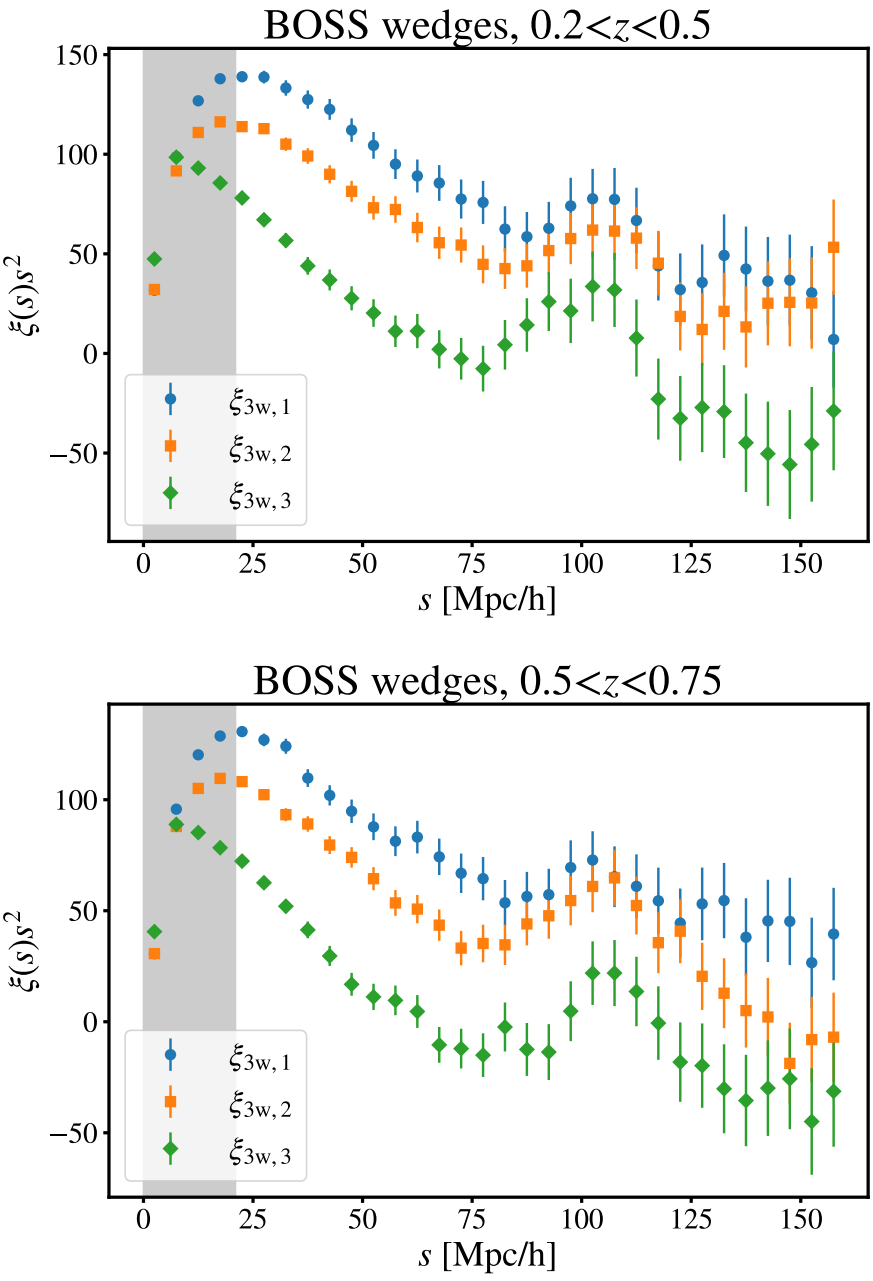

Figure 3.4: BOSS galaxy wedges measurements used in this analysis. The data set consists of two sets of three galaxy wedges measured by splitting the μ range from 0 to 1 into three equal intervals. The low redshift bin (top figure) corresponds to the effective redshift $z_{\rm eff} = 0.38$, whereas the high redshift bin (bottom figure), corresponds to $z_{\rm eff} = 0.61$. The shaded area represents the scales not used in this analysis.

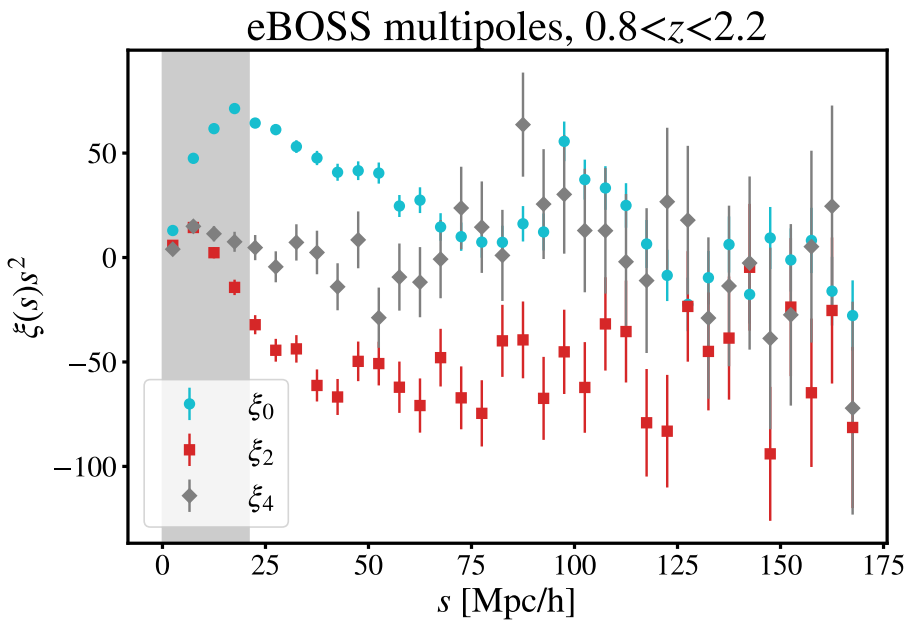

Figure 3.5: eBOSS quasar multipoles measurements used in this analysis. The data set-consists of the measurements of monopole (ξ_0), quadrupole (ξ_2) and hexadecapole (ξ_4) in a single redshift bin, corresponding to the effective redshift of $z_{\rm eff} = 1.48$. The shaded area represents the scales not used in this analysis.

2013) at the Sloan Foundation Telescope at Apache Point Observatory (Gunn et al., 2006). Throughout its different stages (York et al., 2000; Eisenstein et al., 2011; Blanton et al., 2017) the SDSS has provided redshift information on millions of galaxies and quasars.

We consider clustering measurements in configuration space from two data sets: the galaxy samples of BOSS (Dawson et al., 2013), corresponding to SDSS DR12 (Alam et al., 2015; Reid et al., 2016), and the QSO catalogue (Lyke et al., 2020) from eBOSS (Dawson et al., 2016), contained in SDSS DR16 (Ahumada et al., 2020; Ross et al., 2020). In each case, the information from the full anisotropic correlation function $\xi(s, \mu)$, where s denotes the comoving pair separation and μ represents the cosine of the angle between the separation vector and the line of sight, was compressed into different but closely related statistics.

We analyse the clustering properties of the combined BOSS galaxy sample using the measurements of Sánchez et al. (2017), who employs the clustering wedges statistic (Kazin et al., 2012), $\xi_{\Delta\mu}(s)$, which corresponds to the average of $\xi(s, \mu)$, over the interval $\Delta\mu = \mu_2 - \mu_1$ as given in equation (2.109). Sánchez et al. (2017) measured three wedges by splitting the μ range from 0 to 1 into three equal-width intervals. We consider their measurements in two redshift bins, with $0.2 < z < 0.5$ (the LOWZ sample) and $0.5 < z < 0.75$ (CMASS), corresponding to the effective redshifts $z_{\rm eff} = 0.38$ and 0.61, respectively (see Figure 3.4). The covariance matrices, $\mathbf{C}$, of these data were estimated using the set of 2045 MD-Patchy

mock catalogues described in Kitaura et al. (2016). These measurements were also used in the analysis of Tröster et al. (2020) and the recent studies of the cosmological implications of the KiDS 1000 data set (Heymans et al., 2021; Tröster et al., 2021a).

For the eBOSS QSO catalogue we use the measurements of Hou et al. (2021), who considered the Legendre multipoles given by

$$\xi_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{1}\xi(\mu,s)L_\ell(\mu)\,\mathrm{d}\mu, \tag{3.5}$$

where $L_\ell(\mu)$ denotes the ℓ-th order Legendre polynomial (this is equivalent to the equation we had for the power spectrum multipoles, equation (2.107)). We consider the multipoles $\ell = 0, 2, 4$ obtained using the redshift range $0.8 < z < 2.2$, with an effective redshift $z_{\rm eff} = 1.48$ (see Figure 3.5). The covariance matrix of these measurements were obtained using the set of 1000 mock catalogues described in Zhao et al. (2021). Besides the QSO sample used here, the full eBOSS data set contains two additional tracers, the luminous red galaxies (LRG) and emission line galaxies (ELG) samples (for the corresponding BAO and RSD analyses, see Bautista et al., 2021; Gil-Marín et al., 2020; de Mattia et al., 2021; Tamone et al., 2020). These samples overlap in redshift among them and with the galaxies from BOSS. We, therefore, restrict our analysis of eBOSS data to the QSO sample to, in combination with BOSS, cover the maximum possible redshift range while ensuring that the clustering measurements can be treated as independent in our likelihood analysis.

We treat the measurements from BOSS and eBOSS as in the original analyses of Sánchez et al. (2017) and Hou et al. (2021). We restrict our analysis to pair separations within the range $20\,h^{-1}$Mpc $< s < 160\,h^{-1}$Mpc. We assume a Gaussian likelihood for each set of measurements, in which the covariance matrices are kept fixed. We account for the impact of the finite number of mock catalogues used to derive $\mathbf{C}$ (Kaufman, 1967; Hartlap et al., 2007; Percival et al., 2014). The large number of mock catalogues used ensures that the effect of the noise in $\mathbf{C}$ on the obtained cosmological constraints corresponds to a modest correction factor of less than 2 per cent.

3.3.2 Additional data sets

We complement the information from our clustering measurements with the 3×2pt measurements from DES Y1 (Abbott et al., 2018). We also use the shear measurements from the Kilo-Degree Survey (KiDS-450, Hildebrandt et al., 2016) and present the results in Appendix A.

The source galaxy samples from DES are split into four redshift bins, spanning the redshift range of $0.2 < z \leq 1.3$. In addition to shear measurements from the source galaxies, the DES Y1 data set also includes galaxy clustering and galaxy-galaxy lensing two-point correlation function measurements, as well as the lens redshift distributions for five redshift bins in the range of $0.15 < z < 0.9$. Our scale cuts for these measurements match those of Abbott et al. (2018).

For our 3×2pt analysis, we use the DES likelihood as implemented in CosmoMC (Lewis & Bridle, 2002), which corresponds to the model described in Abbott et al. (2018).

The likelihood includes models for the two-point correlation functions describing galaxy-galaxy lensing, galaxy clustering, and cosmic shear. The correlation functions are modelled making use of Limber and flat-sky approximations (Limber, 1954; Kaiser, 1992; LoVerde & Afshordi, 2008; Kilbinger et al., 2017) with the non-linear power spectrum obtained using HMCode (Mead et al., 2015) as implemented in CAMB (Lewis et al., 2000). The smallest angular separations considered correspond to a comoving scale of $8\,h^{-1}$Mpc. The intrinsic alignment is modelled using a 'non-linear linear' alignment recipe (Hirata & Seljak, 2004; Bridle & King, 2007). The model also includes a treatment for multiplicative shear bias and photometric redshift uncertainty. The former is accounted for by introducing multiplicative bias terms of the form $(1 + m^i)$ for each bin i for shear and galaxy-galaxy lensing. The latter is modelled by the shift parameters δz^i assigned to each bin for both source and lens galaxies. Finally, baryonic effects are not included as they are expected to be below the measurement errors for the range of scales considered in the analysis. For all the weak lensing nuisance parameters we impose the same priors as the ones listed in Abbott et al. (2018).

Additionally, we test the consistency of the low-redshift LSS measurements with the latest CMB temperature and polarization power spectra from the *Planck* satellite (Planck Collaboration et al., 2020), to which we refer simply as 'Planck'. We do not include CMB lensing information. We use the public nuisance parameter-marginalised likelihood `plik_lite_TTTEEE+lowl+lowE` for all Planck constraints (Planck Collaboration et al., 2020).

3.4 Results

Our main results come from the combination of the full shape analyses of the BOSS galaxy clustering wedges and eBOSS QSO Legendre multipoles described in Sec. 3.3.1. We also present combined late-Universe constraints obtained from the joint analysis of these clustering measurements with the 3×2pt data setfrom DES Y1. For comparison, in Appendix A we present the constraints obtained using instead the cosmic shear measurements from KiDS-450, which lead to similar results. As we find a good agreement between BOSS + eBOSS + DES and Planck, we also present the parameter constraints obtained from the combination of all four data sets. These constraints are summarized in Table 4.2 and are discussed in Sects. 3.4.1 – 3.4.3.

3.4.1 Clustering constraints

Here we present the main result of our work - the combined flat ΛCDM constraints from the anisotropic clustering measurements from BOSS and eBOSS. Fig. 3.6 shows the posterior distributions for BOSS and eBOSS separately (light blue and orange contours, respectively) as well as their combined constraints (green contours) for two sub-sets of cosmological parameters. For comparison, we also show the Planck-only constraints in dark blue. The panels on the left show the results on the more traditional parameter set of σ_8, Ω_m, and

Table 3.2: Marginalised posterior constraints (mean values with 68 per-cent confidence interval) derived from the full shape analysis of BOSS + eBOSS clustering measurements on their own, as well as in combination with the 3×2pt measurements from DES Y1 and the CMB data from Planck. We present two sets of constraints: our main results derived with wide priors, as listed in Table 3.1, and the 'Planck shape' constraints obtained by imposing narrow Gaussian priors on the cosmological parameters controlling the shape of the linear power spectrum: the physical baryon density $\omega_{\rm b}$, the physical cold dark matter density $\omega_{\rm c}$ and the spectral index $n_{\rm s}$, as discussed in Section 3.1.

Wide priors			
Parameter	BOSS + eBOSS	BOSS + eBOSS + DES	BOSS + eBOSS + DES + Planck
σ_{12}	0.805 ± 0.049	$0.795^{+0.032}_{-0.037}$	0.7890 ± 0.0078
$\omega_{\rm m}$	0.134 ± 0.011	0.131 ± 0.011	0.14090 ± 0.00085
$\omega_{\rm DE}$	0.328 ± 0.020	0.327 ± 0.020	0.3268 ± 0.0064
$\ln 10^{10} {\rm A_s}$	3.13 ± 0.15	3.14 ± 0.13	3.041 ± 0.016
$n_{\rm s}$	1.009 ± 0.048	1.001 ± 0.047	0.9700 ± 0.0038
σ_8	0.815 ± 0.044	0.803 ± 0.028	0.8029 ± 0.0066
$\Omega_{\rm m}$	$0.290^{+0.012}_{-0.014}$	$0.286^{+0.011}_{-0.013}$	0.3014 ± 0.0053
h	0.679 ± 0.021	0.677 ± 0.021	0.6838 ± 0.0041
S_8	0.801 ± 0.043	0.783 ± 0.020	0.805 ± 0.011

Narrow priors		
Parameter	BOSS + eBOSS	BOSS + eBOSS + DES
σ_{12}	0.785 ± 0.039	0.766 ± 0.019
$\omega_{\rm m}$	0.1426 ± 0.0013	0.1423 ± 0.0012
$\omega_{\rm DE}$	$0.327^{+0.011}_{-0.013}$	0.335 ± 0.011
$\ln 10^{10} {\rm A_s}$	3.011 ± 0.099	2.976 ± 0.054
$n_{\rm s}$	0.9660 ± 0.0044	0.9665 ± 0.0043
σ_8	0.800 ± 0.039	0.785 ± 0.021
$\Omega_{\rm m}$	0.3037 ± 0.0081	0.2985 ± 0.0072
h	$0.6855^{+0.0084}_{-0.0094}$	0.6905 ± 0.0083
S_8	0.805 ± 0.042	0.783 ± 0.019

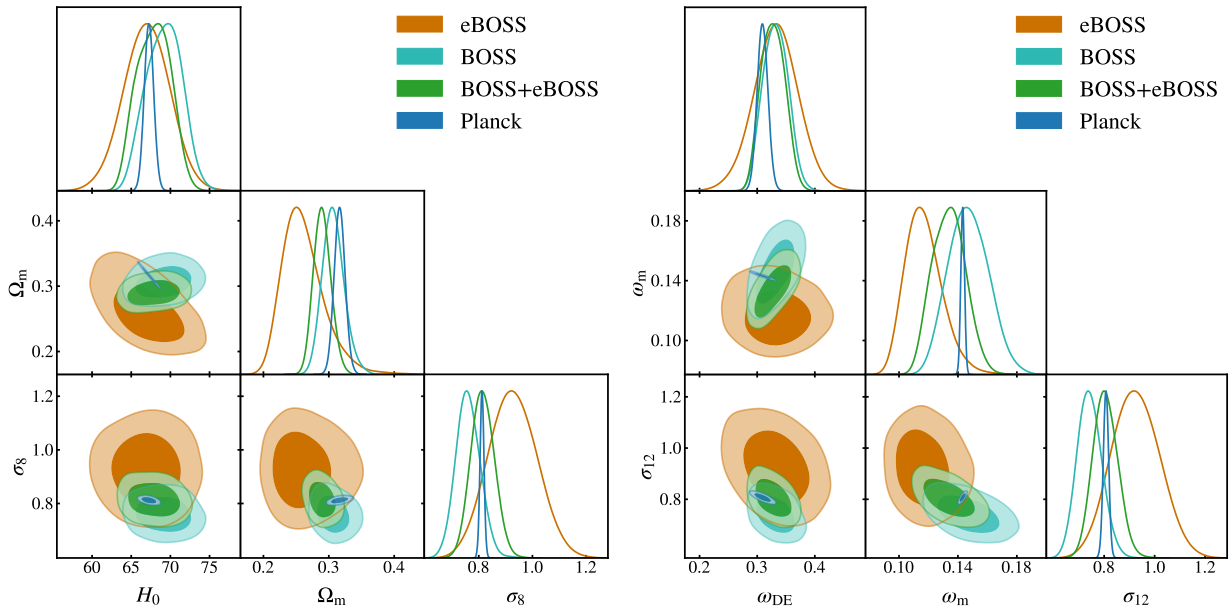

Figure 3.6: Marginalised posterior contours in the 'traditional' and h-independent parameter spaces from the Legendre multipoles of eBOSS QSO sample (orange) and the clustering wedges of BOSS DR12 galaxies (light blue) for a flat ΛCDM model. The joint constraints are shown in green, with Planck in dark blue for comparison.

H_0 whereas the ones on the right correspond to the alternative basis discussed in Sec. 3.1 of σ_{12}, ω_m, and ω_{DE}.

Regardless of the parameter space considered, we find all of our data sets to be in good agreement with each other. The largest deviation between the joint BOSS + eBOSS constraints and those recovered from Planck can be observed in the matter density Ω_m, which displays a difference at the 1.7σ level. Nevertheless, this deviation does not indicate a similarly significant disagreement in the physical matter density preferred by these probes, as the value of ω_m recovered by our clustering constraints matches that of Planck within 0.8σ. This suggests that the differences seen in Ω_m are related to the posterior distributions on h recovered from these data sets. Indeed, looking at our h-independent parameter space, we see that the marginalised constraint of the physical dark energy density also differs from the value preferred by Planck by 0.8σ, with clustering measurements preferring slightly higher ω_{DE}, which translates into a higher value for H_0 and a lower Ω_m.

Tröster et al. (2020) found that the clustering measurements from BOSS wedges prefer a 2.1σ lower value of σ_8 as compared to Planck. Here we confirm the low preference, albeit with much lower significance due to the differences in the modelling of the power spectrum, for both σ_8 and σ_{12} (consistent with Planck at the 1.1σ and 1.3σ level, respectively). The increased consistency between these results is mainly due to the tighter constraints enabled by the use of the co-evolution relation of equation (2.74), which restricts the allowed region of the parameter space to higher values of σ_8 and σ_{12}, as can be seen in Fig. 3.3. The constraints on σ_8 and σ_{12} recovered from eBOSS are at similar levels of agreement with

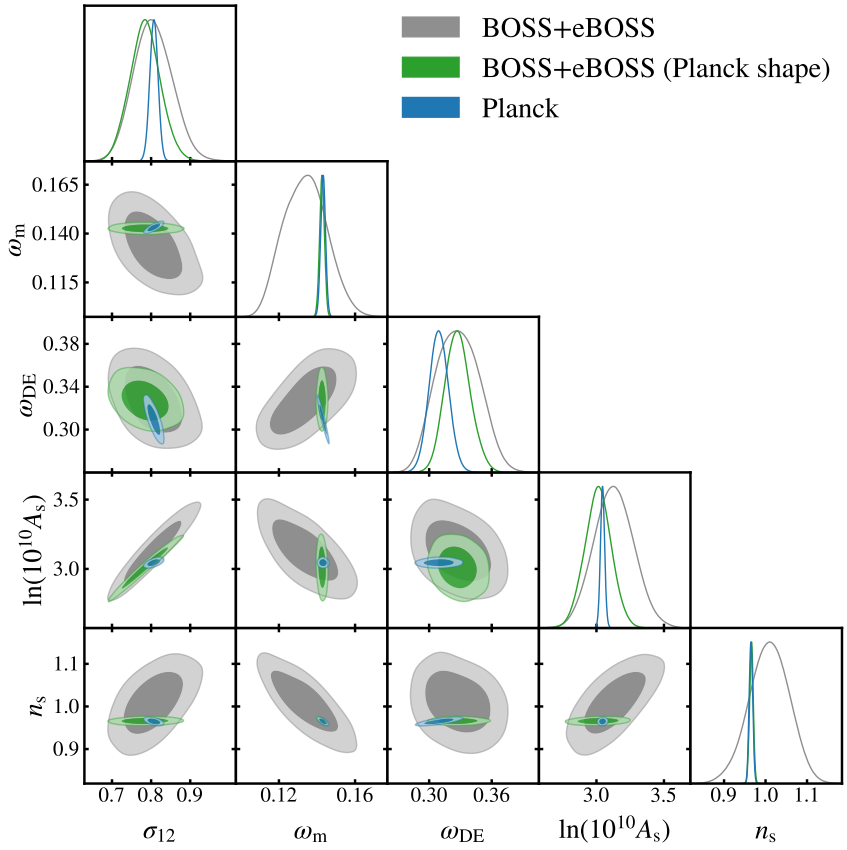

Figure 3.7: When an informative prior is imposed on BOSS + eBOSS for the shape parameters ω_b, ω_c, and n_s so as to match the power spectrum shape obtained by Planck, the recovered constraints (in green) on the evolution parameters are in a good agreement with Planck (dark blue) with a slightly more significant deviation in ω_{DE} only: BOSS + eBOSS prefer a $\sim 1.2\sigma$ higher value of ω_{DE} than Planck.

Planck, however, the values recovered are 1.3σ and 1.2σ *higher* than the CMB results. This is also consistent with the most recent analysis by Hou et al. (2021) and Neveux et al. (2020), who found the inferred growth rate $f\sigma_8$ to be $\sim 2\sigma$ higher than the ΛCDM model with the best-fitting Planck parameters. The combination of the clustering measurements from BOSS and eBOSS is, therefore, in an overall excellent agreement with Planck - with differences at the level of 0.05σ for σ_8 and 0.04σ for σ_{12}.

As discussed in Sec. 3.1, the shape parameters ω_b, ω_c, and n_s are all tightly constrained by Planck with posterior distributions that are in complete agreement with those inferred from the other cosmological probes considered here. We can, therefore, study the improvement in the constraints on the evolution parameters ω_{DE} and A_s that are obtained from the LSS probes when the shape of the power spectrum is constrained to match that of Planck's cosmology. As described in Sec. 3.1, we achieve this by adding an informative Gaussian prior on the shape parameters based on our Planck runs and repeating our analysis with

an otherwise identical set up.

The results of this exercise are shown in Fig. 3.7. As the two data sets were already in a good agreement across the parameter space, including the shape parameters, imposing additional priors simply adds constraining power on the degenerate evolution parameters, most notably $\omega_{\rm DE}$ (degenerate with $\omega_{\rm m}$), which is recovered to be slightly higher than the Planck value to compensate the slight shifts in σ_{12} and $\ln(10^{10}A_{\rm s})$ to lower values.

3.4.2 Consistency with Planck

When looking at marginalised posteriors we are limited by our selection of the parameter space as well as the associated projection effects and, while we can use the standard deviation to quantify agreement on a particular parameter value, this becomes inappropriate when larger parameter spaces are considered. We, therefore, wish to further explicitly quantify the agreement between eBOSS + BOSS and Planck using a tension metric, as has become standard in cosmological analyses.

First, we want to establish agreement over the whole parameter space considered. In order to do this, we use the suspiciousness tension metric, S, introduced by (Handley & Lemos, 2019). The main advantages of using suspiciousness include the fact that it measures the agreement between two data sets across the entire parameter space, similarly to the Bayes factor R. However, unlike R, the suspiciousness is by construction insensitive to prior widths, as long as the posterior is not prior-limited. Given two data sets, A and B, the suspiciousness quantifies the mismatch between them by comparing the relative gain in confidence in data setA when data setB is added (as measured by R) with the unlikeliness of the two data sets ever matching as measured by the information ratio I, that is

$$\ln S = \ln R - \ln I. \tag{3.6}$$

Following the method described in Heymans et al. (2021), we redefine $\ln R$ and $\ln I$ in terms of the expectation values of the log-likelihoods $\langle \ln \mathcal{L} \rangle$ and evidences Z. The evidences, however, cancel out and we are able to calculate S from the expectation values only:

$$\ln S = \langle \ln \mathcal{L}_{\rm A+B} \rangle_{P_{\rm A+B}} - \langle \ln \mathcal{L}_{\rm A} \rangle_{P_{\rm A}} - \langle \ln \mathcal{L}_{\rm B} \rangle_{P_{\rm B}} \,. \tag{3.7}$$

The value of S can then be interpreted using the fact that, for Gaussian posteriors, the difference $d - 2\ln S$, where d is the Bayesian model dimensionality, is χ^2_d distributed. We calculate d for each of the data sets separately, $d_{\rm A}$ and $d_{\rm B}$, and their combination, $d_{\rm A+B}$, as described in Handley & Lemos (2019) and combine the results as $d = d_{\rm A} + d_{\rm B} - d_{\rm A+B}$.

Applying this procedure to eBOSS + BOSS and Planck, we find $\ln S = 0.41 \pm 0.07$ with a Bayesian dimensionality of d=4.5 ± 0.4, which correctly indicates that there are approximately 5 cosmological parameters shared between the two data sets. This can then be related to a p-value of $p = 0.52 \pm 0.02$ or a tension of $0.64 \pm 0.03\sigma$, which is consistent with the $0.76 \pm 0.05\sigma$ tension between Planck and BOSS alone found by Tröster et al. (2020) and indicates a good agreement between these data sets.

In addition to the suspiciousness, we want to use a tension metric that allows for a greater control to focus only on a selected subset of parameters. For this purpose, we use the update difference-in-mean statistic, $\mathcal{Q}_{\rm UDM}$, as described in Raveri & Hu (2019) and implemented in TENSIOMETER[1] (Lemos et al., 2021). This statistic extends the simple difference in means, where the difference in mean parameter values $\hat{\boldsymbol{\theta}}$ measured by two data sets is weighted by their covariance $\mathbf{C}$. The 'update' in UDM refers to the fact that instead of comparing data setA with data setB, we consider the updated information in the combination A + B with respect to A by means of

$$\mathcal{Q}_{\rm UDM} = \left(\hat{\boldsymbol{\theta}}_{\rm A+B} - \hat{\boldsymbol{\theta}}_{\rm A}\right)^t \left(\mathbf{C}_{\rm A} - \mathbf{C}_{\rm A+B}\right)^{-1} \left(\hat{\boldsymbol{\theta}}_{\rm A+B} - \hat{\boldsymbol{\theta}}_{\rm A}\right). \tag{3.8}$$

This has the advantage of the posterior of A + B being more Gaussian than that of B alone. For Gaussian distributed parameters, $\mathcal{Q}_{\rm UDM}$ is chi-square distributed with a number of degrees of freedom given by the rank of $(\mathbf{C}_{\rm A} - \mathbf{C}_{\rm A+B})$. The calculation of $\mathcal{Q}_{\rm UDM}$ may be performed by finding the Karhunen–Loéve (KL) modes of the covariances and re-expressing the cosmological parameters in this basis. This transformation allows us to reduce the sampling noise by imposing a limit to the eigenvalues of the modes that are considered and in this way cutting out those that are dominated by noise (which represent the directions in which adding B does not improve the constraints with respect to A). The number of remaining modes corresponds to the degrees of freedom with which $\mathcal{Q}_{\rm UDM}$ is distributed. For our tension calculations we, therefore, only select the modes α whose eigenvalues λ_α satisfy :

$$0.05 < \lambda_\alpha - 1 < 100. \tag{3.9}$$

This corresponds to requiring that a mode of the base data set is updated by at least 5 per-cent. We subsequently find that there are 2 modes being constrained when Planck is updated by both probe combinations considered in this work (BOSS + eBOSS and BOSS + eBOSS + DES).

For BOSS + eBOSS we get $\mathcal{Q}_{\rm UDM} = 2.0$ for the full parameter space, resulting in a 'tension' with Planck of 0.90σ - only slightly higher than what S suggests.

3.4.3 Joint analysis with DES data

Following Tröster et al. (2020), we want to further investigate the constraints from multiple low-redshift probes together by adding a weak lensing data set - in this case, the 3×2pt measurements from DES Y1. Tröster et al. (2020) used the suspiciousness statistic and showed that the combination of BOSS clustering and KiDS-450 shear measurements are in $\sim 2\sigma$ tension with Planck. The most recent KiDS-1000 3×2pt analysis (Heymans et al., 2021), where the BOSS galaxy sample was used for galaxy clustering and galaxy-galaxy lensing measurements, also found a similar level of tension when the entire parameter space is considered. As DES Y1 measurements have no overlap with either BOSS or eBOSS, we

[1] https://github.com/mraveri/tensiometer

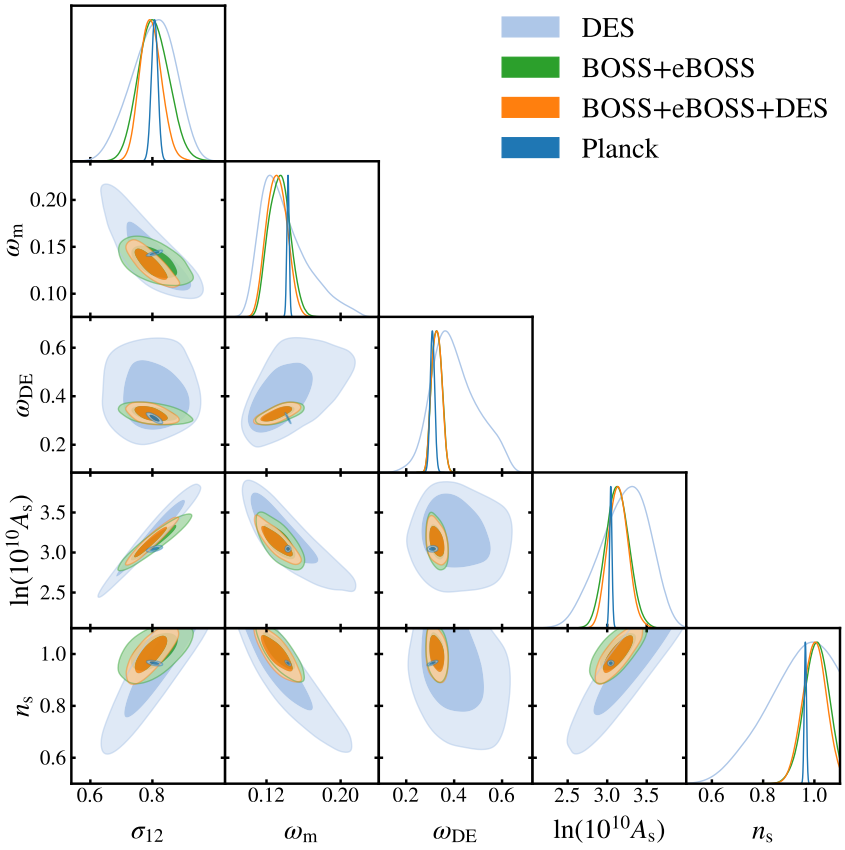

Figure 3.8: In orange - 'low-redshift' constraints for flat ΛCDM obtained from combining BOSS + eBOSS clustering (green) with DES 3×2pt (light blue) and compared with Planck (dark blue). While we obtain a good consistency overall, we note the slight discrepancy between the low redshift probes and Planck contours in $\log(10^{10}A_s) - \sigma_{12}$ and ω_m–σ_{12} projections, reminiscent of the tension seen in σ_8–Ω_m plane.

can treat these data sets as independent and easily combine them to test whether we also find a similar trend.

The resulting constraints are presented in Fig. 3.8. We confirm that DES is in good agreement with eBOSS + BOSS (with $\ln S = -1.08 \pm 0.05$, which corresponds to a $1.3 \pm 0.08\sigma$ tension) and it is, therefore, safe to combine them. The addition of DES data to the analysis provides only slightly tighter constraints with respect to eBOSS + BOSS, with the greatest improvement in σ_{12}, and an overall good agreement with Planck.

Nevertheless, it is worth noting that, when considering the two dimensional posterior projections, there are two parameter combinations in particular for which the 1σ contours of DES + BOSS + eBOSS and Planck do not overlap. The slight discrepancy we observe in the ω_m – σ_{12} plane is reminiscent of the 'σ_8 tension' seen in Ω_m – σ_8 and is larger than the discrepancy displayed by either of the probes individually. In addition to that, we also see a similarly slight disagreement in the $\ln(10^{10}A_s)$ – σ_{12} plane. The projection of

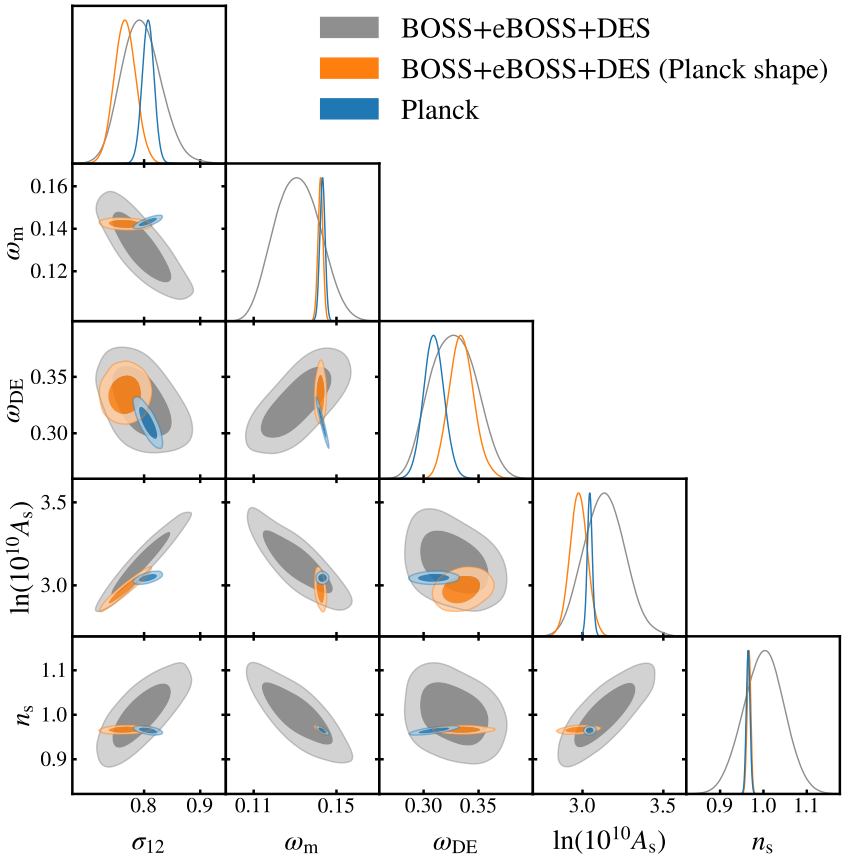

Figure 3.9: Constraints on flat ΛCDM models from the full combination of low-redshift probes (DES + BOSS + eBOSS) obtained after imposing a Planck prior on shape parameters n_s, ω_b and ω_c (orange contours). The constraints from the original uninformative prior analysis (grey contours) and Planck (dark blue contours) are shown for comparison. The results show similar trends as in the case of BOSS + eBOSS. There is a shift to higher values of ω_{DE} that leads to a lower power spectrum amplitude today, σ_{12}, and, to a lesser extent, a lower $\log(10^{10}A_s)$.

A_s with σ_{12} (as opposed to σ_8) allows us to recover the tight degeneracy between the two parameters which exposes how, for a given present-day clustering amplitude, low-redshift probes prefer a higher initial power spectrum amplitude.

We find that adding the DES Y1 3×2pt measurements worsens the agreement with Planck with respect to the results obtained from the combination of BOSS and eBOSS alone. We obtain a suspiciousness of $\ln S = -1.86 \pm 0.04$, corresponding to a tension of $1.54 \pm 0.08\sigma$. When considering the UDM statistic across the entire shared parameter space, we find $\mathcal{Q}_\mathrm{UDM} = 6.3$ distributed with 2 degrees of freedom, which translates into a tension at the 1.9σ level. As for the case of the clustering-only constraints, $\mathcal{Q}_\mathrm{UDM}$ indicates a greater level of tension than S.

Lemos et al. (2021) found that the DES Y1 3×2pt measurements alone are in a 2.3σ tension with Planck, as measured by $\mathcal{Q}_\mathrm{UDM}$. This increases to $2.4 \pm 0.02\sigma$ when using the suspiciousness statistic. These levels of tension are comparable with what we find from the full combination of low-redshift probes. The tension between Planck and weak lensing data sets is usually interpreted as a reflection of tension in the parameter combination $S_8 = \sigma_8(\Omega_\mathrm{m}/0.3)^{0.5}$, that is taken to describe the 'lensing strength'. The S_8 value that we recover from the joint low redshift probes is also about 2σ lower than the Planck constraint (see Table 4.2). Nevertheless, as we see in Fig. 3.8, there is a comparable discrepancy in $\log(10^{10}A_\mathrm{s})$ – σ_{12} plane. We can use $\mathcal{Q}_\mathrm{UDM}$ in order to quantify and compare the level of tension present in these two-dimensional projections by calculating it for a subset of shared parameter space. We find that the amount of tension in both Ω_m – σ_8 and its h-independent equivalent is $\sim 2.0\sigma$, whereas $\log(10^{10}A_\mathrm{s}) - \sigma_{12}$ displays a slightly higher tension of 2.5σ.

We also repeated our fitting procedure with an additional Gaussian prior on the parameters controlling the shape of the power spectrum to be consistent with Planck, as described in Sec. 3.1. The resulting posteriors are shown in Fig. 3.9. We observe the same general trends as from the analysis of our clustering data alone discussed in Sec. 3.4.1. However, the prior on the shape parameters leads to larger shifts in the evolution parameters. This is expected, as DES data on their own cannot constrain the shape parameters well. Adding the informative priors breaks the denegeracies between shape and evolution parameters and increases the constraining power significantly. This, in turn, exposes any discrepancies in the evolution parameters. The values of σ_{12} and $\ln(10^{10}A_\mathrm{s})$ preferred by our low-redshift probes when an informative prior is imposed are, respectively, 1.89σ and 1.22σ lower than the corresponding Planck values. Meanwhile, the recovered value for ω_DE is 1.73σ higher.

3.5 Discussion

The flat ΛCDM constraints from the low-redshift probes presented in Sec. 3.4 show a consistent picture. Updating the power spectrum model and supplementing the clustering measurements with eBOSS data brings the joint BOSS + eBOSS constraints to a better agreement with Planck than the BOSS-only results from Tröster et al. (2020). These

constraints are not significantly modified when these data are combined with DES, resulting in a good overall consistency with Plank across the entire parameter space, as indicated by both S and $\mathcal{Q}_{\rm UDM}$.

Nevertheless, when considering specific two-dimensional projections we still see intriguing differences, mainly driven by the lensing data. Although the constraints in the σ_{12} – $\omega_{\rm m}$ plane obtained using BOSS + eBOSS and DES data separately do not show the discrepancy with Planck that characterizes the results in their h-dependent counterparts of σ_8 and $\Omega_{\rm m}$, the full combination of low-redshift probes tightens the degeneracy between these parameters and leads to constraints that are just outside the region of the parameter space preferred by Planck.

We also see differences in the $\log(10^{10}A_{\rm s})$ – σ_{12} plane between DES and Planck, which are inherited by the full combination of low-redshift data sets. The tight relation between these parameters, which is not seen when using σ_8, illustrates the closer link between σ_{12} and the overall amplitude of density fluctuations obtained by eliminating the ambiguity caused by the dependency on h. For a given value of σ_{12}, Planck measurements prefer a lower initial amplitude of density fluctuations than DES, suggesting a discrepancy in the total growth of structures predicted by these two data sets.

Within the context of a ΛCDM model, the key parameter controlling the growth of structure at low redshift is the physical dark energy density. Indeed, as can be seen in Fig. 3.8 the posterior distribution of $\omega_{\rm DE}$ recovered from DES extends to significantly higher values than the one obtained using Planck CMB measurements. The tendency of the low-redshift data to prefer a higher value of $\omega_{\rm DE}$ than that of Planck can be seen more clearly in the results obtained after imposing a prior on the shape parameters shown in Fig. 3.9. In this case, we find $\omega_{\rm DE} = 0.335 \pm 0.011$ using BOSS + eBOSS + DES while Planck data lead to $\omega_{\rm DE} = 0.3093 \pm 0.0093$. A higher value of $\omega_{\rm DE}$ corresponds also to a higher value of h. Therefore, this difference is also interesting in the context of the Hubble tension, as many of the proposed solutions to this issue focus on modifying the dark energy component.

The analysis of the consistency between low- and high-redshift data has been focused on the comparison of constraints on S_8, which depends on the present-day value of σ_8. Fig. 3.10 shows the redshift evolution of $S_8(z)$ predicted by Planck and the combination of all low-redshift data sets. These curves are consistent at high redshift during matter domination and start to diverge at $z < 1$ to reach a difference at the 2σ level at $z = 0$. However, as this redshift is not probed by any LSS data set, the value of $S_8(z = 0)$ is an extrapolation based on the assumption of a ΛCDM background evolution. Extending this extrapolation to $a > 1$, the difference between the two cosmologies continues to increase and becomes even more significant. Therefore, quoting the statistical significance of any discrepancy in the recovered values of $S_8(z = 0)$ might not be the best characterization of the difference in the cosmological information content of these measurements.

As discussed before, DES and Planck data appear to prefer different evolutions for the growth of cosmic structure, which in a ΛCDM universe depends on $\omega_{\rm m}$ and $\omega_{\rm DE}$. As the former is exquisitely constrained by Planck for general parameter spaces, the latter is perhaps the most interesting parameter to consider. As $\omega_{\rm DE}$ is constant in redshift for a

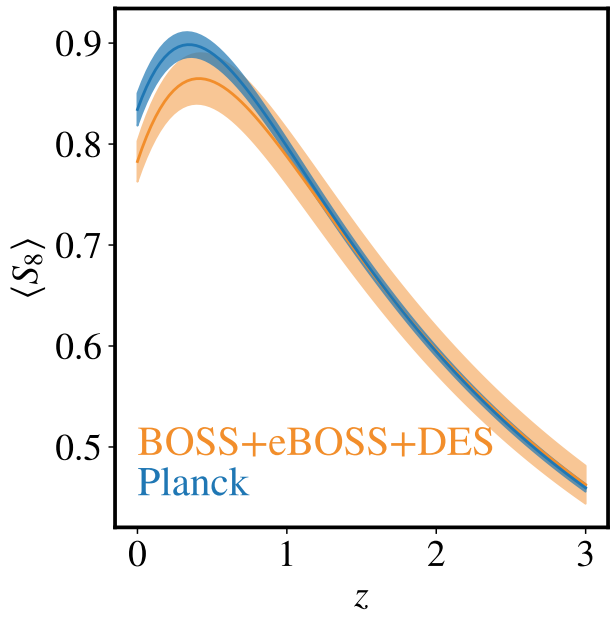

Figure 3.10: Comparison of the inferred mean value for $S_8(z)$ (solid lines) and their corresponding 68 per-cent confidence level (shaded area) corresponding to the combination of BOSS + eBOSS + DES (orange) and Planck (blue).

ΛCDM universe, the deviations in the value of this parameter recovered from different data sets could be used as an indication of their consistency within the standard cosmological model.

3.6 Summary

In the work presented in this chapter, we obtained constraints on the parameters of the standard ΛCDM model from the joint analysis of anisotropic clustering measurements in configuration space from BOSS and eBOSS. In particular, we used the information of the full shape of the clustering wedges of the final BOSS galaxy samples obtained by Sánchez et al. (2017) and the legendre multipoles of the eBOSS DR16 QSO catalogue of Hou et al. (2021). We updated the recipes to describe the non-linear matter power spectrum and the non-local bias parameters with respect to those used in the BOSS-only analyses of Sánchez et al. (2017) and Tröster et al. (2020). We directly compared our theoretical predictions for different cosmologies against the BOSS and eBOSS clustering measurements, without the commonly used RSD and BAO summary statistics. We focus on cosmological parameters that can be classified either as shape or evolution parameters (Sanchez et al., 2021), such as the physical matter and dark energy densities, instead of other commonly used quantities such as $\Omega_{\rm m}$ and $\Omega_{\rm DE}$ that depend on the value of h. Our constraints from the combination of BOSS + eBOSS represent improvements ranging from 20 to 25 per-cent with respect to those of Tröster et al. (2020) and are in excellent agreement with Planck, with the

suspiciousness and updated difference in means tension metrics indicating agreement at the level of 0.64σ and 0.90σ, respectively.

We combined the clustering data from BOSS and eBOSS with the 3×2pt correlation function measurements from DES Y1 to obtain joint low-redshift cosmological constraints that are also consistent with the ΛCDM Planck results, albeit with larger deviations (1.54σ and 2.00σ differences as inferred from S and $\mathcal{Q}_{\rm UDM}$, respectively). We do see interesting discrepancies in certain parameter combinations at the level of 2σ or more, such as the $\Omega_{\rm m}$ – σ_8 and $\omega_{\rm m}$ – σ_{12} planes, and, more significantly, in the $\log(10^{10}A_{\rm s})$ – σ_{12} projection. For a given value of σ_{12}, low-redshift probes (mostly driven by DES) prefer a higher amplitude of primordial density fluctuations than Planck, indicating differences in the total growth of structure predicted by these data sets.

We further tested the impact of imposing a Gaussian prior on $\omega_{\rm b}$, $\omega_{\rm c}$, and $n_{\rm s}$ representing the constraints on these shape parameters recovered from Planck data. Such prior leads to a significant improvement in the constraints on the evolution parameters, such as $\omega_{\rm DE}$ and $A_{\rm s}$. In this case, we find that the full combination of low-redshift data sets prefers a value of the physical dark energy density $\omega_{\rm DE}$ that is 1.7σ higher than that preferred by Planck. This discrepancy, which is also related to the amount of structure growth preferred by these data sets, offers and interesting link with the H_0 tension, as it points to a higher value of h being preferred by the low-redshift data.

Chapter 4

Beyond-ΛCDM Constraints

In the previous chapter, we explored the consistency between the combined galaxy clustering measurements from BOSS galaxies eBOSS quasars and weak lensing '3×2pt' measurements from DES Y1 within the flat ΛCDM model. We showed that, when the data sets are compared in the physical parameter space, Planck prefers slightly more total structure growth than the low redshift cosmological probes. By imposing informative priors on the cosmological parameters that define the shape of the power spectrum, we furthermore demonstrated that this discrepancy could be linked to the differences in the values of the physical dark energy density $\omega_{\rm DE}$ preferred by the two sets of probes.

In light of the inconsistencies within ΛCDM among the different probes, a number of extensions to the base ΛCDM model may be considered, as has already become standard in many major surveys (Spergel et al., 2003; Campbell et al., 2013; Kitching et al., 2014; Abbott et al., 2019; Alam et al., 2021; Tröster et al., 2021b; Planck Collaboration et al., 2020; DES Collaboration et al., 2022). While the constraints derived from BAO and RSD summary statistics have mostly focused on combining clustering with CMB or supernovae data, full-shape galaxy clustering analyses have been shown to provide competitive results beyond ΛCDM without requiring combination with any additional probes (Chudaykin et al., 2021). Importantly, unlike when using these summary statistics, full shape analyses are not susceptible to the bias due to the h^{-1}Mpc units (RSD effects are usually summarised as a combination of linear growth rate f and σ_8, leading to underestimation of the true error, as shown in Sánchez, 2020).

With this motivation in mind, the analysis presented in this chapter extends the full shape BOSS galaxy and eBOSS quasar clustering analysis described in Chapter 3 and presents the current physical parameter space constraints for extensions to ΛCDM. In particular, we are interested in models where w is allowed to take values other than $w = -1$, and in the resulting constraints for curvature, neutrino mass, and time-varying equation of state parameter in such cosmologies.

Our data and modelling choices remain largely the same as in Chapter 3 (however, we do not explore combinations with weak lensing, but rather make use of supernovae measurements, as they provide significant additional constraining power at low redshifts, important for evolving dark energy models). We review the model extensions considered in

Table 4.1: Priors used in our analysis. U indicates a flat uniform prior within the specified range, the nuisance parameter priors are listed in the bottom section of the table. Unless stated otherwise, the priors on the cosmological and clustering nuisance parameters match those presented in Chapter 3.

Parameter	Prior
$\Omega_{\rm b}h^2$	$U(0.019, 0.026)$
$\Omega_{\rm c}h^2$	$U(0.01, 0.2)$
$100\theta_{\rm MC}$	$U(0.5, 10.0)$
τ	$U(0.01, 0.8)$
$\ln(10^{10}{\rm A_s})$	$U(1.5, 4.0)$
$\rm n_s$	$U(0.5, 1.5)$
w	$U(-3, -0.3)$
$w_{\rm a}$	$U(-2, 2)$
$\Omega_{\rm K}$	$U(-0.3, 0.3)$
$\sum m_\nu$	$U(0.0, 2.0)$
b_1	$U(0.5, 9.0)$
b_2	$U(-4, 8.0)$
$a_{\rm vir}$	$U(0.0, 12.0)$
$\sigma_{\rm err}$ (eBOSS only)	$U(0.01, 6.0)$

this analysis in Section 4.1. The results for each of the cosmologies considered are presented in Section 4.2, together with constraints on selected parameters in Table 4.2 (additional parameter constraints are presented in Appendix B). We make use of the simplest extension considered, wCDM, to illustrate the advantages of physical parameter space in such extended models and discuss this in Section 4.2.1. Our conclusions are presented in Section 4.3.

4.1 Extended model overview

When considering extended cosmologies, we expect our chosen parameter space to be most relevant for the cases where the ΛCDM assumptions about dark energy are relaxed, as the physical dark energy density $\omega_{\rm DE}$ is not well constrained by the CMB or large scale structure probes and depends on the assumed dark energy model. This statement is especially true for Planck, which probes the Universe at the redshift of recombination. The dimensionless Hubble parameter, h, is defined by the sum of all energy contributions from baryons ($\omega_{\rm b}$), cold dark matter ($\omega_{\rm c}$), neutrinos (ω_ν), dark energy, and curvature (ω_K) :

$$h^2 = \omega_{\rm b} + \omega_{\rm c} + \omega_\nu + \omega_{\rm DE} + \omega_K, \tag{4.1}$$

with dark energy comprising the majority of the total energy budget today. Therefore, when we introduce additional freedom to dark energy modelling, this is also reflected in the constraints on h and any parameter that is defined through it. In this analysis, we,

therefore, allow the dark energy equation of state parameter $w = p_{\rm DE}/\rho_{\rm DE}$ to deviate from its ΛCDM value of $w = -1$ and treat it as a free parameter for all extensions considered in order to explore the effects on the physical parameter space constraints [1].

In addition to the basic wCDM model with constant w, we also consider a more general parametrization where w is allowed to evolve with the scale factor a (Chevallier & Polarski, 2001; Linder, 2003):

$$w = w_0 + w_a(1 - a). \tag{4.2}$$

Here w_0 and w_a are free parameters; we refer to this case as w_aCDM model. We also explore wKCDM - non-flat models with $\omega_k \neq 0$. Here, as with the other energy budget components, we are interested in physical curvature density $\omega_k = \Omega_k h^2$. Finally, we investigate the constraints on the neutrino mass sum $\sum m_\nu$, by allowing it to vary freely instead of fixing it to the fiducial value of $\sum m_\nu = 0.06$eV, corresponding to the minimum value allowed by neutrino oscillation experiments under normal hierarchy (Otten & Weinheimer, 2008). We refer to this model as $w\nu$CDM.

As in Chapter 3, we use CosmoMC (Lewis & Bridle, 2002) to perform Monte Carlo Markov chain (MCMC) sampling. For the linear-theory matter power spectrum prediction, CosmoMC uses CAMB (Lewis et al., 2000), adapted to compute the theoretical model for anisotropic clustering measurements. In addition to the nuisance parameters, listed in Table 3.1, we sample over the basis cosmological parameters used by CosmoMC:

$$\boldsymbol{\theta}_{\rm base} = \left(\omega_{\rm b}, \omega_{\rm c}, \Theta_{\rm MC}, A_{\rm s}, n_{\rm s}, w_0, w_a, \Omega_k, \sum m_\nu\right), \tag{4.3}$$

where $\Theta_{\rm MC}$ is 100 times the approximate angular size of the sound horizon at recombination. For each of the models described in this section, we only vary the relevant extended parameters, fixing the rest to their fiducial values, as described above. We impose flat and uninformative priors, except for $\omega_{\rm b}$, with the priors for ΛCDM parameters matching those used in the analysis carried out in Chapter 3. Our flat prior for $\omega_{\rm b}$ informs clustering measurements, as they do not constrain this parameter on their own, and is 25 times wider than the corresponding Planck constraint. We also need to specify the allowed values of the Hubble parameter, h. In order to be consistent with our ΛCDM constraints, we choose the same range of $0.5 < h < 0.9$. While this range is somewhat restrictive for Planck on its own for varying dark energy cosmologies, these limits have little effect on the physical parameter space that we consider, and are mostly uninformative once Planck is combined with clustering or when considering clustering alone. Finally, these limits are motivated by the direct measurements of H_0, which fall well within this range (Riess et al., 2021; Kourkchi et al., 2020; Birrer et al., 2020; Reid et al., 2013; Hagstotz et al., 2022). A summary of all cosmological priors used in this analysis is presented in Table 4.1.

Table 4.2: Marginalised posterior constraints (mean values with 68 per-cent confidence interval, for $\sum m_\nu$ - 95 per-cent confidence interval) derived from Planck CMB and the full shape analysis of BOSS + eBOSS clustering measurements on their own, as well as in combination with each other and with Pantheon supernovae Ia measurements (SN). All of the models considered here vary the dark energy equation of state parameter w. In addition to this, w_aCDM also varies w_a, allowing for the equation of state parameter that evolves with redshift, wKCDM varies curvature, and $w\nu$CDM varies neutrino mass sum $\sum m_\nu$. Note that for wKCDM the joint BOSS+eBOSS+Planck constraints should be interpreted bearing in mind that BOSS+eBOSS and Planck are discrepant in this parameter space (Figure 4.4, top). Further constraints, including those on parameters defined through h (relative densities, σ_8), are available in Appendix B.

		Planck	BOSS+eBOSS
wCDM	σ_{12}	0.816 ± 0.011	$0.775^{+0.055}_{-0.066}$
	$\omega_{\rm DE}$	$0.509^{+0.15}_{-0.054}$	$0.352^{+0.033}_{-0.044}$
	w	$-1.41^{+0.11}_{-0.27}$	-1.10 ± 0.13
w_aCDM	σ_{12}	0.816 ± 0.012	$0.768^{+0.053}_{-0.061}$
	$\omega_{\rm DE}$	$0.494^{+0.17}_{-0.062}$	$0.356^{+0.042}_{-0.059}$
	w_0	$-1.22^{+0.33}_{-0.39}$	-1.09 ± 0.30
	w_a	< -0.330	$-0.13^{+1.1}_{-0.94}$
wKCDM	σ_{12}	0.896 ± 0.029	$0.754^{+0.056}_{-0.062}$
	$\omega_{\rm DE}$	$0.323^{+0.073}_{-0.20}$	$0.394^{+0.046}_{-0.053}$
	w	$-1.57^{+0.67}_{-0.38}$	$-0.921^{+0.15}_{-0.093}$
	ω_K	$-0.0116^{+0.0029}_{-0.0036}$	-0.057 ± 0.037
$w\nu$CDM	σ_{12}	$0.810^{+0.019}_{-0.012}$	$0.767^{+0.053}_{-0.064}$
	$\omega_{\rm DE}$	$0.508^{+0.15}_{-0.061}$	$0.353^{+0.036}_{-0.046}$
	w	$-1.43^{+0.16}_{-0.26}$	$-1.16^{+0.16}_{-0.13}$
	$\sum m_\nu$ (eV)	< 0.321	< 1.34

		BOSS+eBOSS+Planck	BOSS+eBOSS+Planck+SN
wCDM	σ_{12}	0.804 ± 0.010	0.8023 ± 0.0097
	$\omega_{\rm DE}$	$0.341^{+0.020}_{-0.023}$	0.329 ± 0.012
	w	$-1.066^{+0.057}_{-0.052}$	-1.033 ± 0.031
w_aCDM	σ_{12}	0.807 ± 0.011	0.805 ± 0.010
	$\omega_{\rm DE}$	$0.322^{+0.026}_{-0.039}$	0.330 ± 0.012
	w_0	$-0.87^{+0.27}_{-0.22}$	-0.955 ± 0.086
	w_a	-0.60 ± 0.68	$-0.34^{+0.36}_{-0.30}$
wKCDM	σ_{12}	0.809 ± 0.011	0.804 ± 0.010
	$\omega_{\rm DE}$	$0.346^{+0.020}_{-0.024}$	0.327 ± 0.012
	w	$-1.108^{+0.078}_{-0.067}$	-1.044 ± 0.036
	ω_K	-0.0012 ± 0.0013	-0.0006 ± 0.0011
$w\nu$CDM	σ_{12}	$0.796^{+0.016}_{-0.012}$	$0.799^{+0.014}_{-0.011}$
	$\omega_{\rm DE}$	$0.346^{+0.020}_{-0.025}$	0.329 ± 0.012
	w	$-1.102^{+0.086}_{-0.058}$	$-1.040^{+0.038}_{-0.033}$
	$\sum m_\nu$ (eV)	< 0.300	< 0.211

4.2 Cosmology constraints

We are mainly interested in clustering constraints from BOSS+eBOSS as well as their combination with Planck. Where found informative, we supplement the clustering constraints with those from SNe Ia. The summary of our results on the main parameters of interest is shown in Table 4.2 with further constraints available in Appendix B. We present our parameter constraints in terms of marginalised posterior mean values with corresponding 68 per-cent confidence intervals (95 per-cent for $\sum m_\nu$). As we perform MCMC sampling to obtain our constraints, we only have a noisy estimate of the true best-fit values, which, however, are all within an order of a standard deviation from the corresponding means. Because of this, and due to the fact that we find our likelihood surface to be fairly flat, with all fits within 1σ of mean cosmological parameter values providing excellent fits, we choose to not present the best-fit values among our results.

4.2.1 Evolving dark energy - wCDM

ΛCDM assumes a cosmological constant-like behaviour for dark energy with a fixed $w = -1$. Nevertheless, w may be allowed to deviate from this value and be treated as a free parameter. As the simplest of the models considered here, we will use our results for wCDM to illustrate the behaviour of our data in the physical parameter space for this class of cosmologies. Figure 4.1 presents our constraints from BOSS+eBOSS and Planck on their own (light and dark blue respectively) as well as their combination (in red) with constraints on the standard parameter space shown in the left panel for comparison.

Comparing the two sets of panels in Figure 4.1, it is clear that the posterior degeneracy directions for H_0 (whose value is determined by the sum of the physical densities of all the components) are set by the $\omega_{\rm DE}$ component. For Planck, the constraint on $\omega_{\rm DE}$ is set by the prior - as discussed before, Planck does not probe the redshfits at which dark energy becomes dominant directly, but rather is able to provide model-dependent constraints based on $D_{\rm A}(z_\star)$. The CMB observations, therefore, do not constrain dark energy density once the evolution of this component is not well defined.

In contrast, $\omega_{\rm m}$ is set by the scale dependence on the amplitude of the CMB power spectra (for a fixed acoustic angular scale) and is not sensitive to the assumptions on dark energy. Following H_0, any parameter defined through it also exhibits similar degeneracies, as they are effectively averaged over the posterior of H_0 - as a result, σ_8 and $\Omega_{\rm m}$ are not well constrained by Planck either.

Nonetheless, importantly, that does not mean that Planck is unable to measure the clustering amplitude today - the wCDM constraint on σ_{12} has the same precision as in ΛCDM model, as illustrated in Figure 4.2. The lack in constraining power on σ_8 is, therefore, an artefact of using h^{-1}Mpc units. To understand why Planck does not lose constraining power on σ_{12} even in extended cosmologies, one can first note that w and σ_{12}

[1]While our wide uninformative w prior is in line with the commonly adopted range, it may be noted that extremely negative values for this parameter violate the Null Energy Condition, as noted in, for example, Colgáin & Sheikh-Jabbari (2021)

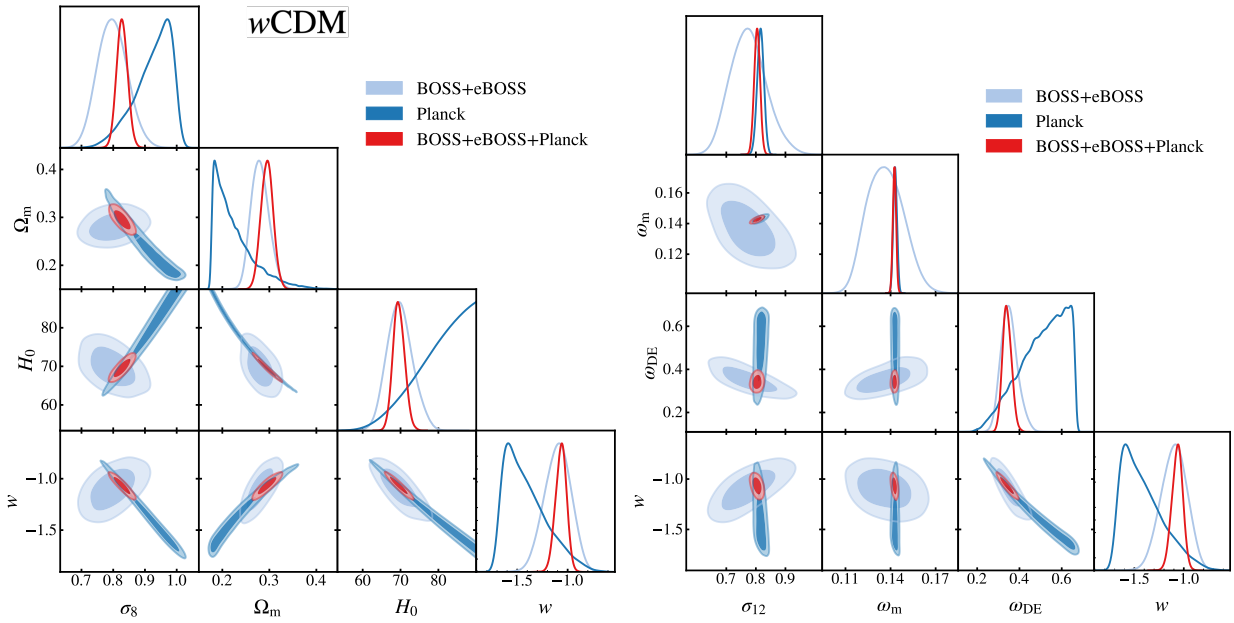

Figure 4.1: Marginalised posterior contours in the 'traditional' and h-independent parameter spaces derived from the full shape of anisotropic clustering measurements of BOSS DR12 galaxies in combination with eBOSS quasars (light blue) and CMB measurements by Planck (dark blue) for a wCDM model. The joint constraints are shown in red. In physical parameter space, Planck is able to constrain the clustering amplitude today σ_{12} even in models with free w.

show almost no correlation for this probe. This behaviour arises because the change in w is compensated by a change in $\omega_{\rm DE}$, as is evident from Planck's constraints in the $w-\omega_{\rm DE}$ plane: only certain combinations of these parameters, set by $D_{\rm A}(z_\star)$, are allowed, with the resulting degeneracy corresponding to a constant σ_{12}. This result means that a preference for more negative w closely corresponds to an increase in $\omega_{\rm DE}$.

Our analysis is, therefore, the first one to quote a CMB constraint on clustering amplitude today in cosmologies with varying w. The Planck best-fit value of $\sigma_{12} = 0.816 \pm 0.011$ that we find for the wCDM model is slightly higher than that for ΛCDM (0.807 ± 0.011) - this increase is because the higher values of $\omega_{\rm DE}$ allowed in wCDM correspond to a more negative w, thus the dark energy content is lower at the start of the epoch when this component becomes relevant, which results in slightly more total structure growth. Nevertheless, as σ_{12} is mostly determined by the physical matter density, as discussed above, this shift in the clustering amplitude value is minimal.

The advantage of our physical parameter space is most evident for Planck due to its lack of constraining power on H_0; nonetheless, even for clustering probes, the precision on physical parameter constraints degrades less, compared to their h-dependent counterparts, once dark energy model assumptions are relaxed.

While clustering on its own prefers a mean equation of state parameter value that is

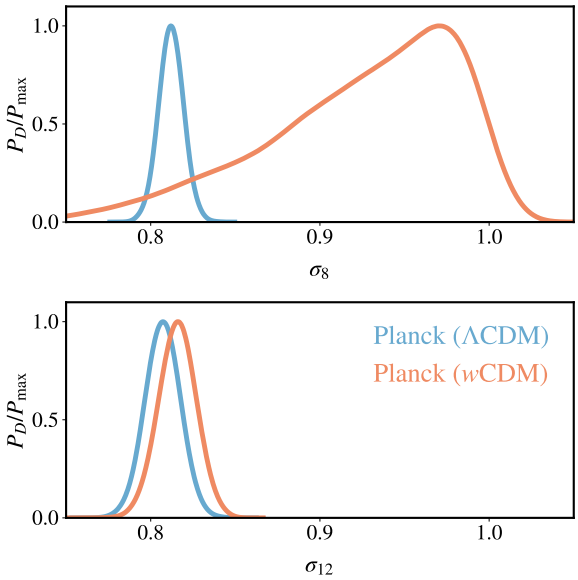

Figure 4.2: Upper panel: one-dimensional marginalised Planck posteriors for σ_8 for ΛCDM (blue) and wCDM (orange) cosmologies . Lower panel: the corresponding posteriors of σ_{12}. The difference between the two panels is due to h^{-1}Mpc units used to define the scale at which the linear density field variance is measured for σ_8. As Planck does not constrain H_0 well once w is allowed to freely vary, the resulting posterior, over which the clustering amplitude is effectively averaged, is extremely wide and results in degraded constraint on σ_8. The parameter σ_{12} is not affected by this issue, as it is defined on a Mpc scale - the precision of the measurement for Planck is the same in both cosmologies shown.

compatible with $w = -1$ ($w = -1.10 \pm 0.13$), for the combination BOSS+eBOSS+Planck, the fiducial value is just outside of the 68% confidence limit. This result appears because of the significant volume of Planck's posterior corresponding to models with high dark energy content, which shifts the mean w. The addition of low redshift information from clustering rules out such models and brings the joint constraints closer to ΛCDM.

Comparing our BOSS+eBOSS+Planck constraints with previously published full shape analysis of BOSS clustering wedges by Sánchez et al. (2017), who obtain $w = -0.991^{+0.062}_{-0.047}$ for BOSS+Planck (2015), reveals that our updated analysis shifts the mean w by $\sim 1\sigma$ towards more negative values. This result may be attributable to a number of differences between the analyses, most notably the updated CMB measurements from Planck.

In addition to this approach, Brieden et al. (2022) performed a reconstructed power spectrum multipole analysis of BOSS DR12 LRG and eBOSS QSO samples. In this analysis, the information from BAO and RSD summary statistics is complemented by additional summary statistic derived from the shape of the power spectrum (ShapeFit, Brieden et al., 2021). Our BOSS+eBOSS(+Planck) constraint $w = -1.10 \pm 0.13$ ($w = -1.066^{+0.057}_{-0.052}$) agrees well with that of Brieden et al. (2022): $w = -0.998^{+0.085}_{-0.073}$ ($w = -1.093^{+0.048}_{-0.044}$), showing the robustness of these results.

4.2.2 Evolving dark energy equation of state - w_aCDM

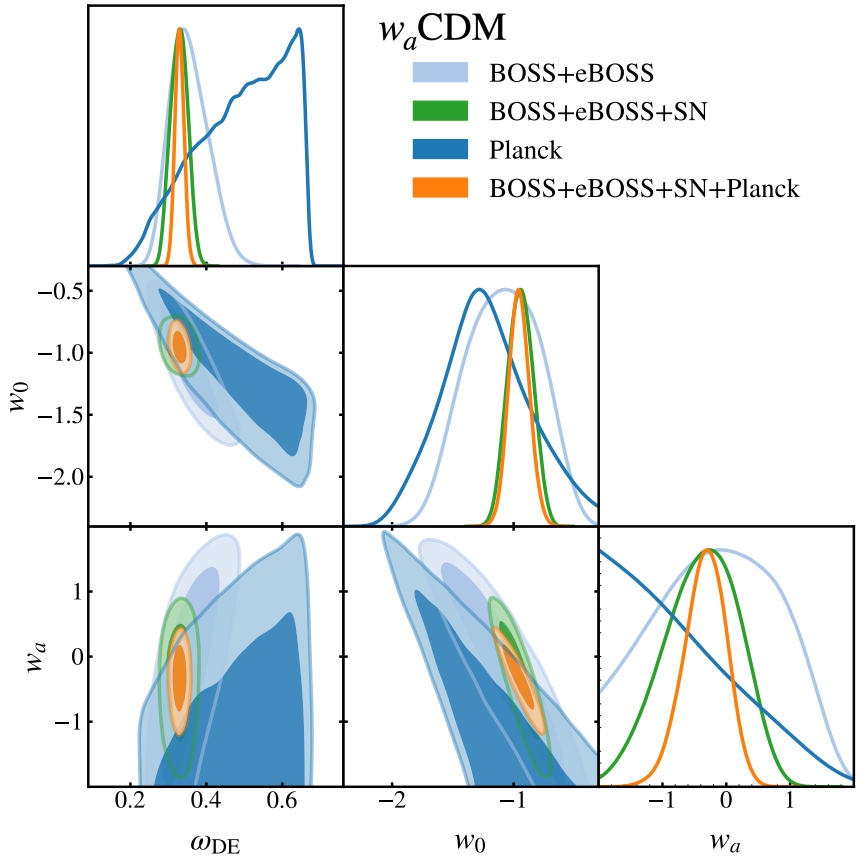

Figure 4.3: Marginalised posterior contours for dark energy parameters in w_aCDM, where the dark energy equation of state parameter w is allowed to evolve in time, as defined in equation (4.2). We show constraints from the full-shape clustering analysis of BOSS DR12 galaxies in combination with eBOSS quasars (light blue), their combination with Pantheon SN Ia measurements (green), CMB constraints by Planck (in dark blue), and the combination of all four data sets (in orange).

We further generalise the dark energy description allowing its equation of state parameter to vary with time, as defined in equation (4.2). Figure 4.3 shows our constraints for the dark energy parameters: w, w_a, ω_{DE}. Here, we combine our clustering constraints with SNe Ia, which provide background constraints for the lowest redshifts and are, therefore, extremely useful for probing the evolution of dark energy.

The additional freedom in the equation of state model has minimal impact on the constraints on σ_{12} and ω_{DE}. However, the addition of SNe Ia data halves the error on dark energy density with the resulting constraint of $\omega_{DE} = 0.330 \pm 0.012$. All of the data set combinations considered recover a value of w_0 that is consistent with -1, although with significantly larger uncertainty than in wCDM. Planck does not constrain w_a on its own,

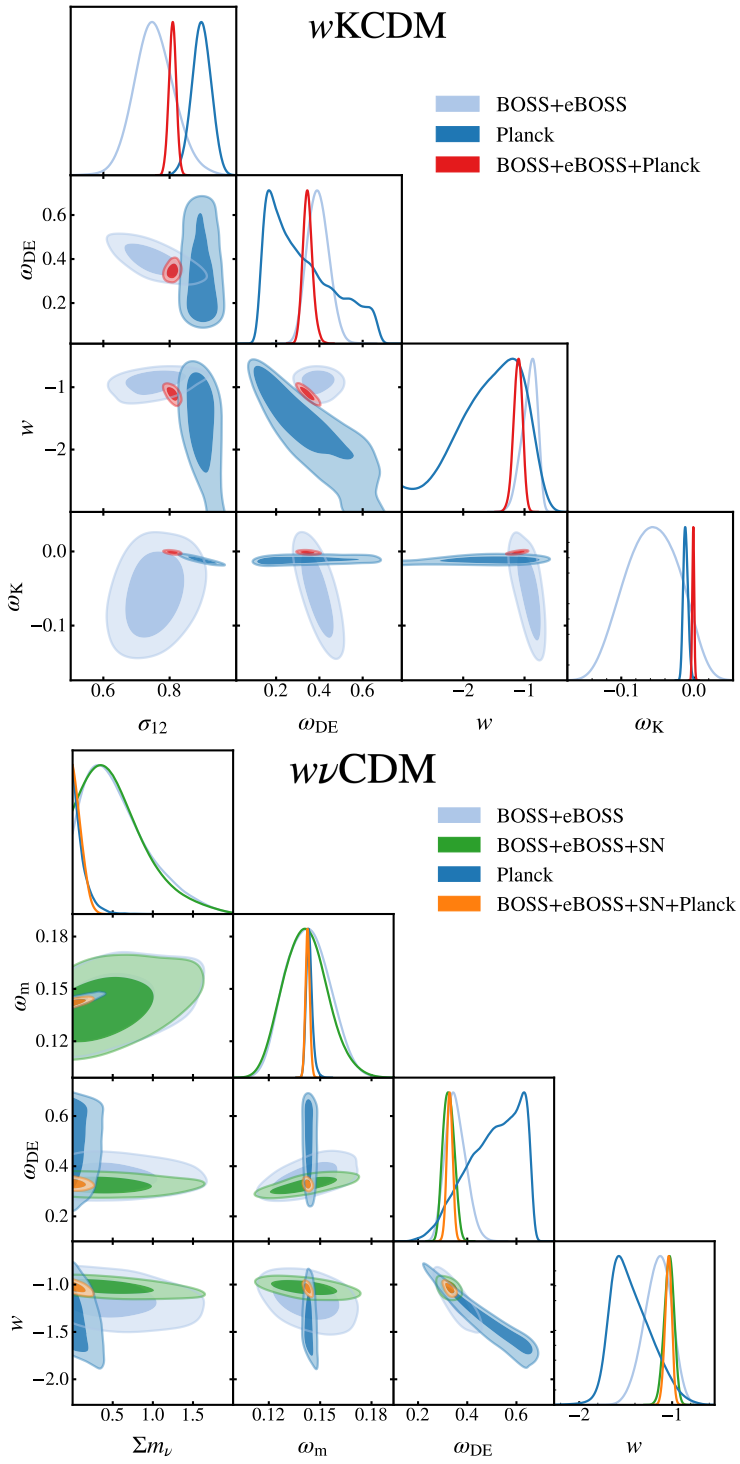

Figure 4.4: Marginalised posterior contours for dark energy parameters and physical matter and curvature densities in wKCDM (*top*) and the neutrino mass sum $\sum m_\nu$ in $w\nu$CDM (*bottom*). *Top*, wKCDM: The constraints from the full-shape clustering analysis of BOSS DR12 galaxies in combination with eBOSS quasars are shown in light blue, CMB constraints by Planck are displayed in dark blue and the combination of the two sets of probes is shown in red. Note the discrepancy between Planck and BOSS+eBOSS in σ_{12} as well as w-ω_{DE} plane. *Bottom*, $w\nu$CDM: Here, we explore varying w together with $\sum m_\nu$. We show constraints from the full-shape clustering analysis of BOSS DR12 galaxies in combination with eBOSS quasars (light blue), their combination with Pantheon SN Ia measurements (green), CMB constraints by Planck (in dark blue) and the combination of all four data sets (in orange).

but combining it with the clustering and supernovae data yields a value compatible with no evolution (for BOSS+eBOSS+Planck+SN $w_a = -0.34^{+0.36}_{-0.30}$).

We may compare our constraints with those from the completed SDSS consensus analysis by Alam et al. (2021), although note that they use additional data sets, including eBOSS luminous red galaxy and emission line galaxy samples as well as BAO from the Lyα forest, and use summary statistics for BAO and RSD to obtain the joint constraints. The consensus analysis also uses reconstructed BAO, while we perform no reconstruction in this work. The quoted constraints for combined Planck+Pantheon SNe, SDSS BAO+RSD and DES 3 $\times$ 2pt data are $w_0 = -0.939 \pm 0.073$ and $w_a = -0.31^{+0.28}_{-0.24}$, which are in excellent agreement with our BOSS+eBOSS+Planck+SN result.

Chudaykin et al. (2021) also performed a full shape analysis using a model based on the Effective Field Theory of Large Scale Structure (EFT, Baumann et al., 2012). They analysed the BOSS DR12 luminous red galaxy redshift space power spectrum multipoles in combination with BAO measurements from post-reconstructed power spectra of BOSS DR12 supplemented with a number of additional BAO measurements from SDSS, including those from the eBOSS QSO sample and additionally augmented by adding supernovae Ia measurements from Pantheon (as also done in this work). Their final constraints of $w_0 = -0.98^{+0.10}_{-0.11}$ and $w_a = -0.32^{+0.63}_{-0.48}$ are tighter than ours (most likely due to the additional BAO data) but in an excellent agreement with our BOSS+eBOSS+SN results: $w_0 = -0.94^{+0.20}_{-0.19}$ and $w_a = -0.40^{+1.0}_{-1.2}$. It is additionally important to note that EFT-based constraints have been shown to depend on the counterterm prior choices (Carrilho et al., 2023; Simon et al., 2022).

4.2.3 Non-zero curvature - wKCDM

We explore what occurs when, in addition to varying w (but with $w_a = 0$), we also allow for non-flat models. The resulting constraints are shown in Figure 4.4 (top). Here, together with the dark energy parameter constraints, we also display the physical curvature density $\omega_\mathrm{K} = \Omega_\mathrm{K} h^2$.

It is interesting that the Planck data constrain the physical curvature well, with the mean value of $\omega_\mathrm{K} = -0.0116^{+0.0029}_{-0.0036}$, indicating a strong preference for non-zero curvature. We compare this result with the constraint on the h-dependent equivalent, $\Omega_\mathrm{K} = -0.030^{+0.018}_{-0.010}$; note how physical units allow us to detect the deviation from flatness at a higher significance (4σ for ω_K versus 1.6σ for Ω_K). This preference for a closed Universe is a known feature of Planck data and believed to be related to the lensing anomaly (Planck Collaboration et al., 2020). Nonetheless, our physical curvature constraint indicates the most significant deviation from flat Universe yet, which is especially interesting bearing in mind that, in addition to curvature, we are also varying w and would, therefore, expect a somewhat more significant preference for a closed Universe in fixed $w = -1$ case (however, as seen in Figure 4.4, dark energy parameters are almost independent of ω_K for Planck; therefore, we expect the change in the result to be minimal).

Recently, Glanville et al. (2022) reported that clustering data alone may show a 2σ preference for a closed universe. There, a full shape analysis based on EFT is performed on

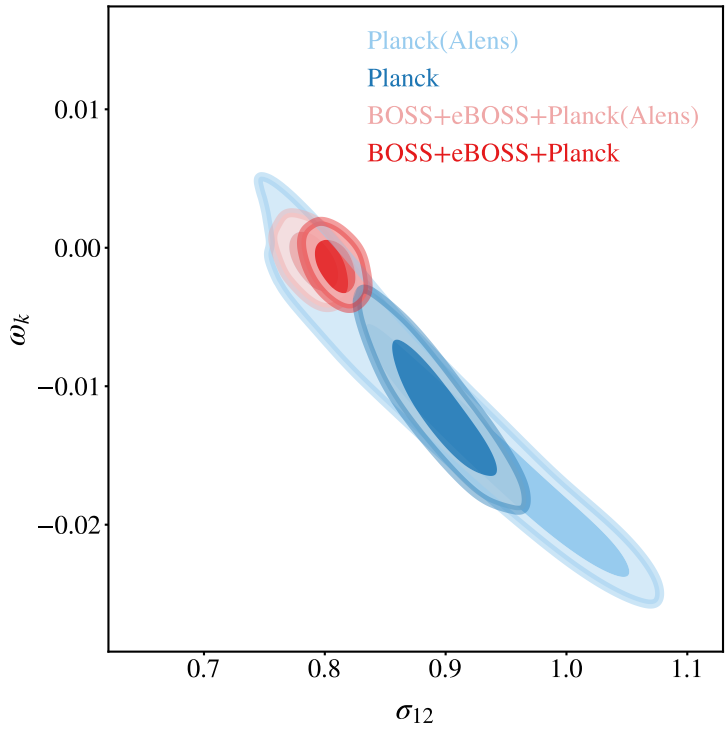

Figure 4.5: Marginalised posterior contours for σ_{12} and $\omega_{\rm K}$ from Planck and its combination with BOSS+eBOSS in wKCDM. We compare two cases - one with varying $A_{\rm lens}$ (light blue and pink) and a fiducial one, with $A_{\rm lens}$ fixed to 1 (dark blue and red).

the power spectra multipoles from the full combined 6dFGS, BOSS, and eBOSS catalogues. The analysis by Chudaykin et al. (2021), however, finds a less significant deviation of $\sim 1\sigma$. Neither of these analyses vary w, which allows for somewhat tighter constraints than what we expect for wKCDM (as seen in Figure 4.4, clustering exhibits some degeneracy between the two parameters). The mean value of $\omega_{\rm K}$ preferred by BOSS+eBOSS in our analysis, $\omega_{\rm K} = -0.057 \pm 0.037$, also deviates from 0, but is consistent with flatness at 95% confidence level, indicating no significant preference for a closed universe. In terms of dark energy constraints, the effect of allowing a free varied curvature for clustering is to allow for larger values of $\omega_{\rm DE}$.

CMB and clustering data sets are highly complementary in wKCDM, with Planck providing a measurement on curvature and BOSS+eBOSS constraining dark energy parameters. Nevertheless, the two data sets are discrepant within this cosmology. This behaviour is most clear from the $w - \omega_{\rm DE}$ projection in which the 2σ regions of the two sets of contours show little overlap. As discussed before, this particular degeneracy is defined by the clustering amplitude today, so this discrepancy is also reflected in the σ_{12} constraints, with Planck preferring a 2.4σ higher value. This model displays the greatest shift in Planck's σ_{12} out of all of the cosmologies considered in this work.

A discrepancy between Planck and BAO measurements and Planck and full shape

analysis of clustering measurements was also found in previous work that varied curvature but kept w fixed (Di Valentino et al., 2020; Handley, 2021; Vagnozzi et al., 2021; Glanville et al., 2022). Our analysis, therefore, demonstrates that the degradation of constraining power when varying w does not provide a solution to this tension (although see also Bel et al., 2022, for "clustering ratio" based analysis which shows agreement between clustering and CMB even in curved models).

Planck's lensing anomaly, which is related to the preference for non-zero curvature, can be characterised by the phenomenological parameter $A_{\rm lens}$, which scales the amplitude of the lensing power relative to the physical value. In the absence of systematics or non-standard physics, $A_{\rm lens} = 1$ and is highly degenerate with the measured cosmological parameters that set the amplitude of the power spectrum at late times - σ_{12} and $\omega_{\rm K}$. Figure 4.5, illustrates how allowing $A_{\rm lens}$ to freely vary extends Planck's posterior contours to include $\omega_{\rm K} = 0$ within 2σ and, therefore, recovers flat ΛCDM. The extension of Planck's posterior distribution of σ_{12} to lower values allows for a reconciliation with the constraints from BOSS+eBOSS and reduces the mean σ_{12} inferred from the combined BOSS+eBOSS+Planck analysis. Varying $A_{\rm lens}$ allows to compensate for the excess lensing and brings Planck in line with clustering measurements in wKCDM. This behaviour is expected and is consistent with existing analyses (for example, Di Valentino et al. (2021) demonstrated how varying $A_{\rm lens}$ allows Planck to be more consistent with flatness and brings it to a better agreement with the BAO measurements for cosmological models with varying curvature, w, and neutrino mass sum). Here we additionally note that the inclusion of $A_{\rm lens}$ does significantly degrade Planck's ability to constrain σ_{12}.

4.2.4 Massive neutrinos - $w\nu$CDM

Finally, we vary the neutrino mass sum, $\sum m_\nu$: here, once again, we supplement our clustering and CMB data sets with Pantheon supernovae. As $\sum m_\nu$ exhibits a degeneracy with the dark energy equation of state (for a more detailed discussion see Hannestad, 2005), varying w is expected to degrade the resulting constraints. The addition of SNe Ia, therefore, allows to improve the precision of our constraint through providing a measurement of w.

While clustering alone does not provide a tight upper limit for the neutrino mass ($\sum m_\nu < 1.34$ eV from BOSS+eBOSS at 95% confidence), its combination with Planck does allow for some improvement as compared with Planck alone ($\sum m_\nu < 0.300$ eV for joint constraints versus $\sum m_\nu < 0.321$ eV for Planck). It might, nevertheless, initially seem surprising that the improvement is rather minimal (especially, compared to the constraints obtained from combined Planck full shape clustering analyses in ΛCDM, such as Ivanov et al., 2020a; Tanseri et al., 2022). This result can be understood by noting that, due to the degeneracy between the physical matter density, $\omega_{\rm m}$, and $\sum m_\nu$, a precise measurement of $\omega_{\rm m}$ is required to shrink the upper limit on $\sum m_\nu$. While clustering constraints on either of these parameters are much looser than those of Planck, it may, nonetheless, be able to improve on Planck's measurement of $\omega_{\rm m}$ by excluding some of the cosmologies in $\omega_{\rm m} - \omega_{\rm DE}$ space allowed by Planck. This effect happens to be more significant in a ΛCDM cosmology,

where the full shape BOSS+eBOSS analysis yields a lower mean value of $\omega_{\rm m}$, as compared to Planck, resulting in a lower $\omega_{\rm m}$ from a combined measurement and, therefore, reducing the maximum limit of $\sum m_\nu$. As shown in Figure 4.4, in the cosmology with varying w the BOSS+eBOSS contour is perpendicular to Planck in $\omega_{\rm m} - \omega_{\rm DE}$, resulting in minimal effect on the constraint on $\omega_{\rm m}$ and, consequently, little improvement on the upper mass limit for $\sum m_\nu$.

Our tightest constraint then arises from the combination of BOSS+eBOSS+Planck+SN, which places the upper limit of $\sum m_\nu < 0.211$ eV at 95% confidence. This improvement is due to the fact that the supernovae constraint on $\omega_{\rm DE}$ does result in a tighter constraint on $\omega_{\rm m}$, which, in turn, shrinks the $\omega_{\rm m}$, and $\sum m_\nu$ degeneracy.

The SDSS consensus analysis (Alam et al., 2021) provides a lower neutrino mass sum limit for wνCDM of $\sum m_\nu < 0.139$ eV (Planck+BAO+RSD+SN+DES, 95% upper limit); nevertheless, these constraints do include additional data (reconstructed BAO and RSD measurements from full SDSS data as well as measurements from the Dark Energy Survey, DES) and are, therefore, not directly comparable. As described above, the final limit is extremely sensitive to $\omega_{\rm m}$ and, therefore, not only the statistical power of a particular combination of measurements but also the position and orientation of the contours (which may differ for full shape and BAO-only analyses). Here we only vary neutrino mass and, while there are additional non-standard neutrino properties that could be explored, given our limited constraining power in varying dark energy cosmologies, it is unlikely we might be able to obtain meaningful constraints. It has additionally been shown that in the cosmological constant scenarios, full shape analysis of BOSS and eBOSS galaxy clustering recovers no deviations from standard neutrino properties (Kumar et al., 2022).

4.2.5 Discussion

In this work we performed a full shape analysis of the anisotropic two-point clustering measurements from BOSS galaxy and eBOSS QSO samples together with Planck CMB and Pantheon SNe Ia measurements and explored extensions to the ΛCDM cosmological model. In particular, we were interested in models with free dark energy equation of state parameter w and the resulting constraints in physical parameter space.

We demonstrated that CMB recovers a tight degeneracy in the $w - \omega_{\rm DE}$ parameter space and is able to constrain the linear density field variance, σ_{12}, as well as the physical curvature density $\omega_{\rm K}$, to high precision. The apparent lack of constraining power when using σ_8 is, therefore, only an artifact of using h^{-1}Mpc units when defining the scale on which the density field variance is measured. This approach results in averaging over the posterior of h that Planck does not constrain in evolving dark energy models. We subsequently presented the first CMB measurements of the clustering amplitude today in cosmologies with varying w and found that the clustering amplitude tends to increase for such models. This behaviour is because a more negative w requires a lower initial $\omega_{\rm DE}$ value to reach the same constraint at redshift zero. This observation also rules out the evolving dark energy models considered here as a potential way to bring Planck's predicted amount of structure growth closer to weak lensing observations on their own, as such extensions do

not affect the initial amplitude of matter fluctuations and the clustering amplitude today σ_{12} is set by the $w - \omega_{\rm DE}$ degeneracy and is well measured even in the extended models.

When, in addition to w, the curvature is also allowed to vary, BOSS+eBOSS and Planck become discrepant, most significantly in the $w - \omega_{\rm DE}$ plane and, subsequently, in the resulting values of σ_{12}, with Planck preferring a 2.4σ higher value than BOSS+eBOSS. Varying dark energy models are, therefore, not able to bring the two probes in a better agreement for curved cosmologies. In addition to this result, our physical curvature density constraint for Planck $\omega_{\rm K} = -0.0116^{+0.0029}_{-0.0036}$ prefers a curved Universe at 4σ significance, which is 2.4σ higher than what is found using $\Omega_{\rm K}$.

It is encouraging that the extended model constraints that we derive from our full shape analysis are compatible with a ΛCDM cosmology (with the greatest deviation seen for $\omega_{\rm K}$ but still within $\omega_{\rm K} = 0$ at 95% confidence) as well as with previous clustering analyses. We derive the 95% upper limit for the neutrino mass sum of $\sum m_\nu < 0.211$ eV (BOSS+eBOSS+Planck+SN), which is a higher value than that of the SDSS consensus analysis (though the two constraints are not directly comparable, as the consensus analysis makes use of a more extensive data set). When w is varied freely, clustering alone only allows for a modest improvement in the upper limit of $\sum m_\nu$.

Our analysis demonstrates the strength of the physical parameter space in constraining extended cosmologies. While we are currently still unable to place tight constraints on the dark energy parameters directly, we were able to show how even the high-redshift observations place limits on the allowed behaviours. We were also able to provide a consistent picture of the current state of full-shape clustering constraints, which were shown to be highly complementary to the CMB measurements. With CMB providing information on physical matter and curvature densities, as well as setting a strict limit on the allowed clustering amplitude values and clustering offering a way to measure dark energy, we may hope that the Stage-IV surveys will be able to confidently exclude large regions of the extended parameter space.

4.3 Summary

In this work, we performed a full shape analysis of the anisotropic two-point clustering measurements from BOSS galaxy and eBOSS QSO samples together with Planck CMB and Pantheon SNe Ia measurements and explored extensions to the ΛCDM cosmological model. In particular, we were interested in models with free dark energy equation of state parameter w and the resulting constraints in physical parameter space.

We demonstrated that CMB recovers a tight degeneracy in the $w - \omega_{\rm DE}$ parameter space and is able to constrain the linear density field variance, σ_{12}, as well as the physical curvature density $\omega_{\rm K}$, to high precision. The apparent lack of constraining power when using σ_8 is, therefore, only an artifact of using h^{-1}Mpc units when defining the scale on which the density field variance is measured. This approach results in averaging over the posterior of h that Planck does not constrain in evolving dark energy models. We subsequently presented the first CMB measurements of the clustering amplitude today in

cosmologies with varying w and found that the clustering amplitude tends to increase for such models. This behaviour is because a more negative w requires a lower initial $\omega_{\rm DE}$ value to reach the same constraint at redshift zero. This observation also rules out the evolving dark energy models considered here as a potential way to bring Planck's predicted amount of structure growth closer to weak lensing observations on their own, as such extensions do not affect the initial amplitude of matter fluctuations and the clustering amplitude today σ_{12} is set by the $w - \omega_{\rm DE}$ degeneracy and is well measured even in the extended models.

When, in addition to w, the curvature is also allowed to vary, BOSS+eBOSS and Planck become discrepant, most significantly in the $w - \omega_{\rm DE}$ plane and, subsequently, in the resulting values of σ_{12}, with Planck preferring a 2.4σ higher value than BOSS+eBOSS. Varying dark energy models are, therefore, not able to bring the two probes in a better agreement for curved cosmologies. In addition to this result, our physical curvature density constraint for Planck $\omega_{\rm K} = -0.0116^{+0.0029}_{-0.0036}$ prefers a curved Universe at 4σ significance, which is 2.4σ higher than what is found using $\Omega_{\rm K}$.

It is encouraging that the extended model constraints that we derive from our full shape analysis are compatible with a ΛCDM cosmology (with the greatest deviation seen for $\omega_{\rm K}$ but still within $\omega_{\rm K} = 0$ at 95% confidence) as well as with previous clustering analyses. We derive the 95% upper limit for the neutrino mass sum of $\sum m_\nu < 0.211$ eV (BOSS+eBOSS+Planck+SN), which is a higher value than that of the SDSS consensus analysis (though the two constraints are not directly comparable, as the consensus analysis makes use of a more extensive data set). When w is varied freely, clustering alone only allows for a modest improvement in the upper limit of $\sum m_\nu$.

Chapter 5

Emulating the two-point correlation function with evolution mapping

The analyses performed in the previous chapters rely on evaluating the full-shape galaxy two-point correlation function at every point of the sampled parameter space. This is a computationally expensive process, in particular, when considering extended cosmologies with additional free parameters. When performing Bayesian inference, a converged MCMC chain typically requires the order of $\mathcal{O}(10^5 - 10^6)$ likelihood evaluations, which means that a small improvement in a single model calculation can amount to a significant difference in total computing time. This significance further increases if we take into account the fact that, in addition to the final cosmological analysis, a number of sampling runs are required for model tuning and validation purposes. This constitutes a relevant bottleneck in the process of extracting cosmological constraints from the full shape of the galaxy clustering measurement. Emulators present a solution to this problem: given a sample of predictions within the parameter space of interest, one can use machine learning techniques to train an emulator that can then provide a theory model for any parameter combination within the training range at a fraction of the time it takes to perform a full calculation. This chapter presents the details of the development of the full-shape two-point correlation function emulator based on the evolution-mapping approach.

The idea of using emulation for theory modelling in cosmological analyses was first introduced by Heitmann et al. (2006) and Habib et al. (2007) as a way to replace costly numerical simulations and has been gaining popularity since. A number of emulators have been developed to aid the predictions in cases where the physics is not well understood and the theory prediction must be obtained from simulations, for instance, when modelling the power spectrum enhancement in modified gravity scenarios (for example, Ramachandra et al., 2021; Arnold et al., 2022), its suppression due to baryonic effects (Schneider et al., 2020) or modelling small scale clustering (Zhai et al., 2023; Yuan et al., 2022; Nishimichi et al., 2019; Heitmann et al., 2014; Angulo et al., 2021; Euclid Collaboration et al., 2019, 2021).

Alternatively, the emulation approach can also be used to speed up the calculation of analytically known theory predictions: this is particularly advantageous in cases where the

evaluation of theoretical models is particularly expensive. This is typically the case for perturbative approaches, such as the ones used in this work, which involve a number of integrals over a range of wavevectors (see Section 2.4.5 for the expressions for loop corrections for the matter power spectrum in equations (2.65) and (2.66) as well as the bias corrections described in Section 2.5.1). As a result, there has been a considerable effort put into developing emulators for perturbative model theory predictions for large-scale structure analyses (see DeRose et al., 2022; Donald-McCann et al., 2023). Additionally, significant progress has been made in developing a hybrid approach, making use of perturbative recipes to obtain biased tracer power spectra from the component cross-spectra measured from N-body simulations (for example, Zennaro et al., 2023; Aricò et al., 2021; Kokron et al., 2021; Pellejero Ibañez et al., 2022).

The challenge in building an emulator lies in balancing the coverage for the range of theory predictions required for a particular analysis with the prediction accuracy. For a theory emulator, matching the high accuracy of perturbation theory models is a particularly stringent demand and the ability to meet this requirement is inevitably affected by the parameter space used in the emulator training process. This includes both the number of free parameters of the model as well as the size of their respective prior range. This often means that the emulators only cover narrow parameter ranges and a limited number of cosmological models (for example, the vast majority of existing emulators only consider flat models). Adopting the already mentioned hybrid approach of combining simulated cross-spectra and perturbative expansions is one of the ways to reduce the emulated parameter space, because only the component power spectra need to be emulated, while the corresponding bias prefactors are incorporated exactly using the analytical expression. The pure-theory emulators by Donald-McCann et al. (2023) and Aricò et al. (2021) employ an equivalent approach and factorise all model parameters related to galaxy bias as well. However, that still leaves a large parameter space of cosmological parameters and redshifts that we might be interested in.

A key insight that allows us to significantly reduce the dimension of the cosmological parameter space is provided by the evolution mapping idea. As already described in Section 2.6, when cosmological parameters are expressed in physical units (i.e. not through h), their effect on the power spectrum can be clearly identified as either describing its shape or amplitude. In the linear case, the parameters that describe the amplitude of the power spectrum (evolution parameters and redshift) follow an exact degeneracy (as long as the power spectrum is also expressed in the units of Mpc). This degeneracy can be exploited to reduce the required parameter space by, instead of considering each of the evolution parameters separately, performing a mapping from a model with matching shape parameters. This can be done by relabeling the redshifts that correspond to the required value of σ_{12} (which correctly describes the amplitude of the linear power spectrum) for the desired evolution parameter combination.

The first emulator based on the evolution mapping approach, `COMET`, was introduced by Eggemeier et al. (2023). The emulator employs Gaussian Processes (GP) to predict galaxy clustering power spectrum multipoles in real and redshift spaces. `COMET` emulates the perturbative models' terms, making use of a similar approach as in Donald-McCann et al.

(2023) and Aricò et al. (2021) where the bias dependence is factorised out. The emulator is built to provide predictions for two theory models that differ in their treatment of the RSD: in the effective field theory (EFT) approach the real-to-redshift space mapping is performed through a full perturbative expansion, in contrast, the velocity-generating function (VDG) model retains the non-perturbative nature of this mapping. In both cases, Eggemeier et al. (2023) demonstrate that the power spectrum predictions follow the degeneracy for evolution parameters and `COMET` is trained in terms of σ_{12} and f (to perform the redshift space mapping) in combination with the shape parameters. The AP parameters are further applied in a separate step, in this way further reducing the emulation parameter space.

The clear improvements in the emulator validity range allowed by the inclusion of evolution mapping can be seen when comparing `COMET` to similar theory emulators `EFTEMU` by Donald-McCann et al. (2023) and `EmulateLSS` DeRose et al. (2022) which implement the EFT model and the Lagrangian perturbation theory respectively. While `EFTEMU` supports five cosmological parameters ω_b, ω_c, h, A_s and n_s, `COMET` additionally allows for the most common extensions: varied Ω_K, w_0, w_a. It is crucial to note that the extensions are available without having to explicitly train the emulator on these evolution parameters, retaining high accuracy of emulation (in contrast, see, for example, Euclid Collaboration et al., 2021, who find that including w_0 increases complexity and expense of creating the emulator significantly). In addition to a more limited parameter space support, the ranges allowed for the base ΛCDM parameters by both `EmulateLSS` and `EFTEMU` are, with the exception of n_s, smaller than what is supported by `COMET`. `EmulateLSS` is furthermore more restrictive than `EFTEMU`: it fixes n_s and does not cover the full galaxy bias and counterterm parameter space. Moreover, these emulators make predictions for a fixed background cosmology and fixed redshifts. `COMET` is able to cover the extensive parameter space without compromising emulation accuracy which reaches around $\sim 0.1\%$ for the synthetic data that covers 10 times the *Euclid* volume and with BOSS-like galaxy bias parameters. This is slightly better than the performance of `EmulateLSS` and about an order of magnitude more accurate than `EFTEMU`.

While the performance of `COMET` is extremely promising, its predictions are currently limited to the Fourier space. Similarly to the analyses carried out in Chapters 3 and 4, we can still use `COMET` for modelling the two-point correlation function by simply Fourier transforming the emulator output. Nonetheless, this action adds additional computational time to each theory evaluation. In addition to this, if not performed carefully, the Fourier transform can introduce additional errors, due to the fact that the function transformed is only defined over a limited range of k, while the Fourier transform integral should be performed over the full k-space. This limited k-range creates a step-like feature in Fourier space that may show up as wiggles in the transformed function in the configuration space.

In order to avoid these issues we set out to extend `COMET` for the two-point correlation function analyses by implementing direct emulation in the configuration space. This is achieved by Fourier transforming the original `COMET` training set (which, however, is re-calculated to extend to higher k values) and training the emulator on these configuration space theory predictions. In this chapter, we begin by presenting the theoretical model that we wish to emulate (EFT), as it is implemented in `COMET`, and describe how evolution

mapping guides the emulator design. We proceed with a description of the training data set and, finally, test the configuration space COMET on the validation data set as well as by using it as the theory model to analyse galaxy clustering wedges measured on MINERVA - the galaxy mocks that were used in Section 3.2.

5.1 Theory modelling

In this section we describe the modelling of the redshift-space galaxy power spectrum, highlighting any instances where our approach does not exactly follow Eggemeier et al. (2023), with the configuration space predictions simply obtained by performing a Fourier transform, as described in Section 5.2.4.

5.1.1 Perturbation theory models

COMET emulates the galaxy power spectrum multipole predictions for two different perturbative models, which mainly differ in their treatment of the RSD. In the EFT (Baumann et al., 2012) approach, the finger-of-god effect due to the small-scale virialised galaxy velocities is captured by a set of so-called counterterms (for analyses using EFT see Ivanov et al., 2020b; d'Amico et al., 2020). Alternatively, the FoG can be accounted for by a damping factor that represents the velocity difference generating (VDG) function: this is the approach (albeit with some minor differences from the version implemented in COMET) that was used in the analysis presented in Chapters 3 and 4 as well as in the BOSS and eBOSS full shape analyses by Sánchez et al. (2017), Grieb et al. (2017) and Hou et al. (2021) (for a more detailed description of the model and the explicit expression for the damping function $F_{\rm FoG}$ see equation (2.105) and the accompanying discussion in Section 2.5.2). The two models are related by:

$$P_{\rm gg,VDG}(\boldsymbol{k}) = F_{\rm FoG}(\boldsymbol{k})[P_{\rm gg,EFT}(\boldsymbol{k}) - \Delta P(\boldsymbol{k})], \tag{5.1}$$

where $\Delta P(\boldsymbol{k})$ accounts for the fact that in the VDG model the mapping from real to redshift space does not contain the full perturbative expansion, as it does in the EFT. The two models are, therefore, closely related, however, the VDG option in configuration space requires an additional emulated parameter - the small-scale velocity distribution kurtosis parameter $a_{\rm vir}$. For this reason, we develop the first configuration space version of COMET for the EFT model only, leaving the VDG implementation for future work.

The EFT galaxy-galaxy power spectrum model, $P_{\rm gg,EFT}(\boldsymbol{k})$, can be generally written down as:

$$P_{\rm gg,EFT}(\boldsymbol{k}) = P^{\rm tree}_{\rm gg,SPT}(\boldsymbol{k}) + P^{\rm 1-loop}_{\rm gg,SPT}(\boldsymbol{k}) + P^{\rm stoch}_{\rm gg}(\boldsymbol{k}) + P^{\rm ctr}_{\rm gg}(\boldsymbol{k}), \tag{5.2}$$

with the first two terms corresponding to the leading and next-to-leading contributions from SPT (see Section 2.4.5 for a more in-depth description of SPT in the context of the real space matter power spectrum). This expression is obtained by performing an expansion of real-to-redshift space mapping, taking into account the non-linear evolution of

the matter and velocity fields and galaxy bias. The third term, $P_{\rm gg}^{\rm stoch}(\boldsymbol{k})$, represents the stochastic contribution from highly non-linear scales that is not captured by the perturbative expansion. This term essentially acts as a constant noise term and accounts for highly local effects, which in configuration space would contribute to separations close to zero. We, therefore, do not model $P_{\rm gg}^{\rm stoch}(\boldsymbol{k})$ for the two-point correlation function emulator. Finally, the last contribution, $P_{\rm gg}^{\rm ctr}(\boldsymbol{k})$, is given by a series of counterterms and, as mentioned before, captures the FoG effect, as well as any further effects on the one-loop terms exerted by the modes not modelled by the perturbation theory.

In the following subsections, we provide a brief overview of the non-vanishing components listed in equation (5.2). We omit the description of the galaxy bias model, as it matches the modelling presented in Section 2.5.1 (although with an additional term accounting for the higher derivative bias which will be absorbed into one of the counterterms), but discuss the redshift-space mapping, the counterterm contribution, and the infrared resummation.

5.1.2 Redshift-space power spectra

In this section, we will recap the mapping to redshift space, as described in the EFT model.

The relationship between the density fields in real and redshift space in the non-linear case (and assuming the plane-parallel approximation as well as the approximation for the Jacobian $J = |1 - f\nabla_\parallel u_\parallel| \simeq 1 - f\nabla_\parallel u_\parallel$) can be written down as (Scoccimarro et al., 1999):

$$\delta_{\rm s}(\boldsymbol{k}, z) = \int_{\boldsymbol{x}} {\rm e}^{i\boldsymbol{k}\cdot\boldsymbol{x}} {\rm e}^{-ifk_\parallel u_\parallel(\boldsymbol{x})} D_{\rm s}(\boldsymbol{x}), \tag{5.3}$$

where $D_{\rm s}(\boldsymbol{x}) \equiv \delta_{\rm g(\boldsymbol{x})} + f\nabla_\parallel u_\parallel(\boldsymbol{x})$. The expression for $\delta_{\rm s}(\boldsymbol{k}, z)$ can be expanded (following the same logic as in the real space SPT case, as described in Section 2.4.5) to obtain the perturbation theory kernels in redshift-space, Z_n, such that the resulting SPT expressions for the power spectrum in redshift space at leading and next-to-leading order are:

$$P_{\rm gg,SPT}^{\rm tree}(\boldsymbol{k}) = Z_1(\boldsymbol{k})^2 P_{\rm lin}(k), \tag{5.4}$$

$$\begin{aligned} P_{\rm gg,SPT}^{1-{\rm loop}} =& 2\int_{\boldsymbol{q}} [Z_2(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q})]^2 P_{\rm lin}(q) P_{\rm lin}(|\boldsymbol{k}-\boldsymbol{q}|) \\ &+ 6P_{\rm lin}(k)\int_{\boldsymbol{q}} [Z_3(\boldsymbol{q}, -\boldsymbol{q}, \boldsymbol{k})] P_{\rm lin}(q). \end{aligned} \tag{5.5}$$

This is equivalent to the real space SPT expressions in equations (2.64)-(2.66) but with redshift space perturbation kernels:

$$Z_1(\boldsymbol{k}) = b_1 + f\mu_k^2, \tag{5.6}$$

$$\begin{aligned} Z_2(\boldsymbol{k}_1, \boldsymbol{k}_2) =& \mathcal{K}_2(\boldsymbol{k}_1, \boldsymbol{k}_2) + f\mu_k^2 G_2(\boldsymbol{k}_1, \boldsymbol{k}_2) \\ &+ \frac{1}{2} fk\mu_k \left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2) + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right] \end{aligned} \tag{5.7}$$

$$
\begin{aligned}
Z_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \mathcal{K}_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) &+ f\mu_k^2 G_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) \\
&+ \frac{1}{2} f^2 k^2 \mu_k^2 \frac{\mu_2 \mu_3}{k_2 k_3} (b_1 + f\mu_1^2) \\
&+ fk\mu_k \frac{\mu_3}{k_3} \Big[b_1 F_2(\boldsymbol{k}_1, \boldsymbol{k}_2) + f\mu_{12}^2 G_2(\boldsymbol{k}_1, \boldsymbol{k}_2) \Big] \\
&+ fk\mu_k \frac{\mu_{23}}{k_{23}} \Big(b_1 + f\mu_1^2 \Big) G_2(\boldsymbol{k}_2, \boldsymbol{k}_3) \\
&+ fk\mu_k \frac{\mu_1}{k_1} \Big[\frac{b_2}{2} + \gamma_2 K(\boldsymbol{k}_2, \boldsymbol{k}_3) \Big],
\end{aligned}
\tag{5.8}
$$

with $K(\boldsymbol{k_1}, \boldsymbol{k_2}) = (\boldsymbol{k}_1 \cdot \boldsymbol{k}_2)^2 / (k_1^2 k_1^2) - 1$ (equivalent to $S(\boldsymbol{k_1}, \boldsymbol{k_2})$ in the real space bias expansion described in Section 2.5.1). The mapping of the tree-level contribution, described by Z_1, therefore, simply corresponds to the Kaiser factor, whereas the one-loop cross-spectra and bispectra kernels Z_2 and Z_3 account for the non-linear corrections due to mode coupling. For the sake of conciseness, we omit the full form of the real-space kernels $\mathcal{K}_n$, these expressions can be found in the Appendix A of Eggemeier et al. (2023).

5.1.3 Counterterms

The 1-loop integral in equation (5.5) is performed over all scales $\boldsymbol{q}$, including the deeply non-linear regime where perturbation theory breaks down. In this regime, the assumption of perfectly pressureless fluid no longer applies, as dark matter undergoes shell-crossing and the baryonic pressure is present and impacts the clustering of dark matter on larger scales. The usual way to correct for the small-scale physics effects in the large-scale limit is to introduce a series of *counterterms* with freely varied amplitude pre-factors which in redshift space scale as $\sim \mu_k^{2n} k^2 P_{\mathrm{lin(k)}}$ with $n = 0, 1, 2$ (Senatore & Zaldarriaga, 2014; Desjacques et al., 2018).

The first counterterm accounts for the leading effect from a breakdown of the perfect fluid approximation for the matter field (Pueblas & Scoccimarro, 2009; Baumann et al., 2012; Carrasco et al., 2012). Additionally, this counterterm also absorbs the effect of the higher derivative bias, as it has the same scaling as that of the relevant leading order bias contribution $\nabla^2 \delta$ [1]. The second and third counterterms account for the relevant velocity bias effects.

The leading order (LO) counterterm contribution is, therefore, introduced as a set of three free parameters c_0, c_2 and c_4:

$$
P_{\mathrm{gg}}^{\mathrm{ctr,LO}}(k, \mu_k) = -2 \sum_{n=0}^{2} c_{2n} L_{2n}(\mu_k) k^2 P_{\mathrm{lin}}(k). \tag{5.9}
$$

[1] The next-to-leading order higher-derivative bias terms can be neglected, assuming that the scale at which these contributions become important is of similar order as the non-linearity scale of the matter field.

As seen here, each counterterm c_{2n} mainly contributes to a single multipole of the same order. This is because, in this approach, each Legendre polynomial of order $2n$ replaces the scaling with μ_k^{2n}, which corresponds to a linear transformation of the counterterm parameters.

The one-loop perturbation theory model, as described in the previous section neglects the fact that the mapping from real to redshift space can produce non-negligible corrections coming from the contributions from short-scale velocities (FoG) whose characteristic scale can be significantly lower than the scale controlling gravitational non-linearities. In order to correct for this effect Ivanov et al. (2020b) introduce an additional counterterm to capture the redshift-space non-linearities at next-to-leading order (NLO):

$$P_{\rm gg}^{\rm ctr,NLO}(k,\mu_k) = c_{\rm nlo}(\mu_k k f)^4 P_{\rm gg,SPT}^{\rm tree}(\boldsymbol{k}). \tag{5.10}$$

This term is also included in the Fourier space version of `COMET`. Nonetheless, as $P_{\rm gg}^{\rm ctr,NLO}(k,\mu_k)$ scales as k^4, the term diverges with k, meaning that it is not possible to perform its Fourier transform and that its most significant contribution in configuration space will be at zero separation, $s = 0$. For these reasons, we choose to omit this term when constructing the configuration space emulator. While this omission will inevitably limit the range of scales to which our theory is applicable, in Section 5.3.2 we show that, even without including this contribution, our emulator can recover the true cosmology of our validation mock galaxy catalogues. In principle, the most appropriate method to account for this small-scale velocity contribution in a way that is also compatible with configuration space modelling would be to use a damping function as in the VDG approach. However, as mentioned earlier, this would require expanding the emulator parameter space by an additional free parameter, which is why we leave the implementation of the VDG model in configuration space emulator for future work.

5.1.4 Infrared resummation

So far the model as described here does not account for the large-scale bulk flows that smear the BAO signal via the large-scale relative displacement field. This means that, while we are able to provide a fairly accurate description of the broadband, the modelling of the BAO shows a non-negligible amplitude error (see, for example Baldauf et al., 2015).

The effect from the fluctuations at larger scales $q < k$ can be taken into account at each k by performing a resummation of corrections. At leading order this results in a damping factor that only acts on the BAO wiggles. It is, therefore, standard to separate the linear matter power spectrum into smooth (broadband) and wiggly components $P_{\rm nw}$ and $P_{\rm w}$ (Baldauf et al., 2015; Blas et al., 2016):

$$P_{\rm lin} = P_{\rm nw}(k) + P_{\rm w}(k). \tag{5.11}$$

The split is carried out (following Vlah et al., 2016; Osato et al., 2019) by rescaling the originally proposed expression for the featureless power spectrum $P_{\rm EH}$ by Eisenstein & Hu

(1998), adjusted to match the broadband amplitude for the linear power spectrum:

$$P_{\rm nw}(k) = P_{\rm EH}(k)\mathcal{F}\Big[\frac{P_{\rm lin}(k)}{P_{\rm EH}(k)}\Big]. \tag{5.12}$$

Here $\mathcal{F}$ is a Gaussian bandfilter of the form:

$$\mathcal{F}[f(k)] = \frac{1}{\sqrt{(2\pi)}\log_{10}(\lambda)}\int d\Big(\log_{10}(q)\Big) f(q) \times \exp\Big[-\frac{\Big(\log_{10}(k) - \log_{10}(q)\Big)^2}{2\Big(\log_{10}(\lambda)\Big)^2}\Big], \tag{5.13}$$

where $\log_{10}\Big(\lambda/(h^{-1}{\rm Mpc})\Big) = 0.25$.

At the leading order, we can explicitly rewrite equation (5.11) so that the infrared-resummed power spectrum is defined as the sum of the no-wiggle component and a damped wiggly part with the damping factor calculated assuming Zel'dovich approximation:

$$P_{\rm mm}^{\rm IR-LO}(k) = P_{\rm nw}(k) + {\rm e}^{-k^2\Sigma^2}P_{\rm w}(k), \tag{5.14}$$

with Σ as the two-point function of the relative displacement field evaluated at the BAO scale, such that:

$$\Sigma^2 = \frac{1}{6\pi^2}\int_0^{k_{\rm s}} P_{\rm nw}(q)\Big[1 - j_0\Big(\frac{q}{k_{\rm osc}}\Big) + 2j_2\Big(\frac{q}{k_{\rm osc}}\Big)\Big]dq. \tag{5.15}$$

The n-th order spherical Bessel function j_n here is evaluated at wavelengths scaled with respect to the wavemode $k_{\rm osc} = 1/l_{\rm osc}$ which corresponds to the reference BAO scale $l_{\rm osc} = 110\,h^{-1}{\rm Mpc}$ and $k_{\rm s}$ is the ultraviolet integration limit. This limit is traditionally set at an intermediate scale to avoid performing integration over the full $q < k$ (as the perturbative IR expansion breaks at large k). When calculating the predictions for `COMET` training set this limit is set to $k_{\rm s} = 0.14\,{\rm Mpc}^{-1}$.

At next-to-leading order, the IR-resummed matter power spectrum is a sum of the standard contributions from one-loop corrections sourced by higher powers of the density field, however, it also receives additional contributions:

$$P_{\rm mm}^{\rm IR-NLO}(k) = P_{\rm nw}(k) + \Big(1 + k^2\Sigma^2\Big){\rm e}^{-k^2\Sigma^2}P_{\rm w}(k) + P^{\,1-{\rm loop}}\Big[P_{\rm mm}^{\rm IR-LO}\Big](k). \tag{5.16}$$

Here the square brackets mark the fact that the one-loop integrals are evaluated using the leading order IR-resummed power spectrum, as given in equation (5.14), instead of $P_{\rm lin}$.

The mapping of IR-resummed power spectrum from real to redshift space is similar to the non-resummed case described in Section 5.1.2, but here the damping factor also acquires a dependence on the angle to the line-of-sight μ_k. The leading order term then can be written down as (Ivanov & Sibiryakov, 2018):

$$P_{\rm gg}^{\rm s,IR-LO}(k,\mu_k) = \Big(b_1 + f\mu_k^2\Big)^2\Big[P_{\rm nw}(k) + {\rm e}^{-k^2\Sigma^2_{\rm tot}(\mu_k)}P_{\rm w}(k)\Big] \tag{5.17}$$

with

$$\Sigma^2_{\rm tot}(\mu_k) = \Big[1 + f\mu_k^2(2+f)\Big]\Sigma^2 + f^2\mu_k^2(\mu_k^2-1)d\Sigma^2 \tag{5.18}$$

and

$$d\Sigma^2 = \frac{1}{2\pi^2}\int_0^{k_s} P_{\rm nw}(q) j_2\Big(\frac{q}{k_{\rm osc}}\Big)dq. \tag{5.19}$$

The next-to-leading order contribution can be written as:

$$\begin{aligned} P_{\rm gg}^{\rm s,IR-NLO}(k,\mu_k) =& \Big(b_1 + f\mu_k^2\Big)^2\Big[P_{\rm nw}(k) + \Big(1 + k^2\Sigma^2_{\rm tot}(\mu_k)\Big)\mathrm{e}^{-k^2\Sigma^2_{\rm tot}(\mu_k)}P_{\rm w}(k)\Big] \\ &+ P_{\rm gg}^{\rm s,1-loop}[P_{\rm nw}](k) \\ &+ \mathrm{e}^{-k^2\Sigma^2_{\rm tot}(\mu_k)}\Big(P_{\rm gg}^{\rm s,1-loop}[P_{\rm nw}+P_{\rm w}](k) - P_{\rm gg}^{\rm s,1-loop}[P_{\rm nw}](k)\Big) \end{aligned} \tag{5.20}$$

with the square brackets once again indicating that the loop terms are evaluated by plugging in either the sum of smooth and wiggled contributions or at the smooth contribution only, instead of $P_{\rm lin}$.

5.2 Emulator design

The design of the configuration-space extension to `COMET` mimics that of the Fourier space original. There are three main features that allow `COMET` to achieve superior performance compared to similar emulators. First of all, the use of evolution mapping provides a straightforward recipe for calculating any prediction that corresponds to a rescaling of power spectrum amplitude at fixed shape parameters, i.e., `COMET` is able to provide predictions at any redshift or ΛCDM extensions with evolving dark energy or curvature without expanding the training parameter space. Second, the exact treatment of Alcock-Paczynski distortions allows the user to choose an arbitrary fiducial cosmology, as this effect does not need to be emulated. Finally, while this feature is not unique to `COMET`, the contributions proportional to all possible galaxy bias parameter combinations are emulated separately, which further limits the training parameter space to cosmology parameters only. In the following sections, we describe each of these features and how they apply in configuration space in greater detail.

5.2.1 Evolution mapping

The idea behind evolution mapping in the context of the linear power spectrum has already been introduced in Section 2.6. This approach can also be extended to perturbation theory models as follows. The redshift-space tracer power spectrum model, as discussed in Section 5.1, can be written down as a sum of tree-level and one-loop terms which can be further grouped to make up a series of terms $P_{\mathcal{B}}(\boldsymbol{k}|z)$ that are proportional to unique combinations of galaxy bias (or counterterm) parameters $\mathcal{B}$. Each of these terms is a functional of the linear power spectrum and depends on the growth rate of structures $f(z)$ through the

damping term in the IR resummation step (as shown in equation (5.18)). The same is also true for the two-point correlation function so that its components: $\xi_{\mathcal{B}}(\boldsymbol{s}|z)$ satisfy:

$$\xi_{\mathcal{B}}(\boldsymbol{s}|z,\Theta_{\rm s},\Theta_{\rm e}) = \xi_{\mathcal{B}}\{\boldsymbol{s}|f(z,\Theta_{\rm s},\Theta_{\rm e}),\xi_{\rm lin}(r|\Theta_{\rm s},\sigma_{12}(z,\Theta_{\rm s},\Theta_{\rm e}))\}, \tag{5.21}$$

where $\xi_{\rm lin}$ is just the Fourier transform of $P_{\rm lin}$. As a result, the cosmology and redshift dependence of a perturbation theory prediction is fully determined by the shape parameters, the growth rate f, and σ_{12}. This means that redshifts or evolution parameters need not be explicitly emulated, but rather we simply need to perform a scaling to the correct σ_{12}.

5.2.2 Alcock-Paczynski distortions in COMET

As described in Section 2.5.4, in order to convert from observed galaxy redshifts and angular positions into distances, we need to assume a fiducial cosmology. Any mismatch between the assumed and the true cosmologies will show up as a distortion of the observed two-point correlation function, which needs to be modelled. The amplitude of this distortion is a function of both shape and evolution parameters, such that (with primed quantities evaluated at the fiducial cosmology):

$$q_{\parallel} = \frac{H'(z_{\rm eff}|\Theta'_{\rm s},\Theta'_{\rm e})}{H(z_{\rm eff}|\Theta_{\rm s},\Theta_{\rm e})}, \qquad q_{\perp} = \frac{D_{\rm M}(z_{\rm eff}|\Theta_{\rm s},\Theta_{\rm e})}{D'_{\rm M}(z_{\rm eff}|\Theta'_{\rm s},\Theta'_{\rm e})}. \tag{5.22}$$

Due to the mixing of both shape and evolution parameters and in order to be able to correct for any choice of fiducial cosmology, the effect of the AP distortions is modelled analytically. The two-point correlation function components $\xi_{\mathcal{B}}(\boldsymbol{s}|z)$ are emulated without including the AP effect and the separations and the angles are subsequently rescaled, as described in Section 2.5.4, equations (2.113) and (2.114).

5.2.3 Emulated quantities

Following Eggemeier et al. (2023), we emulate the two-point correlation function contributions $\xi_{\mathcal{B}}(\boldsymbol{s}|z)$. The emulated terms correspond to the Fourier transforms of the terms in the original `COMET` with the exception of the terms involving the $c_{\rm NLO}$ counterterm: as described in Section 5.1.3, these are omitted from the configuration space model. There are, therefore, a total of 14 different contributions, listed in Table 5.1, with linear and one-loop terms for $\mathcal{B} = b_1$ and $\mathcal{B} = 1$ emulated together and for $\mathcal{B} = b_1^2$ emulated separately to retain the option of calculating the IR resummed linear two-point correlation function.

Multipoles and anisotropic two-point correlation function reconstruction

Equivalently to the Fourier space emulator, in configuration space, we also project the full two-dimensional two-point correlation function $\xi(s,\mu)$ into three multipoles and emulate these compressed quantities instead: i.e., we emulate three sets of the components listed in Table 5.1, each corresponding to the contributions to a different multipole.

$\mathcal{B}$	b_1^2	b_1	1	b_1^2	$b_1 b_2$	$b_1\gamma_{21}$	$b_1\gamma_{21}$	b_2^2	$b_2\gamma_2$	γ_2^2	b_2	γ_2	γ_{21}	c_0	c_2	c_4
linear	✓	✓	✓											✓	✓	✓
1-loop		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			

Table 5.1: Bias contributions to the two-point correlation function at linear and one loop order. Each of the terms corresponds to a Fourier transform of the equivalent power spectrum term from Eggemeier et al. (2023).

We project the 2D correlation function components such that:

$$\xi_{\mathcal{B},\ell}(s) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu L_\ell(\mu)\xi_{\mathcal{B}}(s,\mu) \tag{5.23}$$

and emulate the components for the monopole, quadrupole and hexadecapole, which are then used to reconstruct $\xi(s,\mu)$ using the Legendre expansion. The two-dimensional two-point correlation function reconstructed from the three multipoles is, nevertheless, not exact because, first, the redshift-space $\xi(\mu,s)$ contains terms up to μ^8 and, second, the IR damping factor (as given in equation (5.18)) depends on μ^2, which leads to non-zero multipoles for all even l. The error that arises due to neglecting higher-order multipoles is more significant for quadrupole and hexadecapole and for the VDG model (as shown in Eggemeier et al., 2023), nonetheless, following the design of the Fourier space `COMET`, we also apply an approximate correction by additionally including the $\ell = 6$ multipole $\xi_{\mathcal{B},6}$ evaluated at a fixed redshift $z = 1$ and for a fixed set of ΛCDM parameters matching the cosmology from *Planck* TT,TE,EE+lowE+lensing constraints (Planck Collaboration et al., 2020):

$$\begin{aligned}\xi_{\mathcal{B}}(\boldsymbol{s}|z,\Theta_{\rm s},\Theta_{\rm e}) \approx &\sum_{\ell=0}^{2}\xi_{\mathcal{B},2\ell}(s|z,\Theta_{\rm s},\Theta_{\rm e})L_{2\ell}(\mu)+\\ &\xi_{\mathcal{B},6}\Big\{s|f(z,\Theta_{\rm s},\Theta_{\rm e}),\xi_{\rm lin}\Big(s|\Theta_{\rm s}^{\rm Planck},\sigma_{12}(z,\Theta_{\rm s},\Theta_{\rm e})\Big)\Big\}L_6(\mu).\end{aligned} \tag{5.24}$$

As stated in the above expression, the $\ell = 6$ multipole depends on $f(z,\Theta_{\rm s},\Theta_{\rm e})$ and $\sigma_{12}(z,\Theta_{\rm s},\Theta_{\rm e})$, whose variation with shape and evolution parameters is correctly accounted for. In order to reduce the error in our approximation, we can factor out this dependence by splitting up each of the $\xi_{\mathcal{B},6}$ terms into contributions with different powers of f and scaling their amplitude such that:

$$\xi_{\mathcal{B},6} = \left(\frac{\sigma_{12}(z,\Theta_{\rm s},\Theta_{\rm e})}{\sigma_{12}(z=1,\Theta_{\rm s}^{\rm Planck},\Theta_{\rm e}^{\rm Planck})}\right)^{2L}\xi_{\mathcal{B},6}|_{\rm Planck}, \tag{5.25}$$

with $L = 1$ for the linear terms and $L = 2$ for all of the one-loop terms (as listed in Table 5.1). The remaining effect of our approximation is that when calculating $\xi_{\mathcal{B},6}$, $\xi_{\rm lin}$ is always evaluated at fixed shape parameters and that the IR damping term dependence on f and σ_{12} is computed at $z = 1$ and Planck cosmology shape parameters.

Parameter	Min. emulator range	Max. emulator range
ω_b	0.0205	0.02415
ω_c	0.085	0.155
n_s	0.92	1.01
σ_{12}	0.2	1.0
f	0.5	1.05

Table 5.2: The ranges of validity for the configuration space `COMET` extension

Summary

Following the evolution mapping expression in equation (5.21), our `COMET` extension consists of two emulators. First, we emulate the components for each of the three two-point correlation function multipoles at different combinations of shape parameters, f, and σ_{12} and, second, we emulate σ_{12} for combinations of shape parameters with fixed evolution parameter values [2]:

1. $\xi_{\mathcal{B},l}(s|\Theta_\mathrm{s}, \sigma_{12}, f)$ for $l = 0, 2, 4$ and for all $\mathcal{B}$ in Table 5.1
2. $\sigma_{12}(z = 1, \Theta_\mathrm{s}, \Theta_\mathrm{e}^\mathrm{fixed})$

This is two fewer emulators than in the Fourier space `COMET`, which also includes emulators for the VDG model parameter σ_v and the linear power spectrum P_lin. The latter is needed because in the Fourier space `COMET`, the quantity emulated by emulator 1 is the ratio of the multipoles with the linear power spectrum P_lin, which is done in order to reduce the dynamical range (and, therefore, improve the accuracy) of the emulated quantity. This approach is, however, not appropriate for the two-point correlation function, because its amplitude spans both positive and negative values across the range of s and so taking a ratio with ξ_lin would result in a function that diverges at the points where it crosses zero. We, nonetheless, find that even without the step of taking the multipole ratio our emulator is able to successfully recover the validation data set, as discussed in Section 5.3. The multipole contributions $\xi_{\mathcal{B},l}$ are emulated on a range of scales spanning separations from 10 Mpc to 250 Mpc (to account for a range of fiducial cosmologies with AP distortions).

5.2.4 Parameter space and training

Parameter space

The configuration space extension to `COMET` is constructed over a total of five cosmological parameters: the growth rate f, required to obtain predictions in redshift space, σ_{12}, which defines the amplitude of the power spectrum (and, therefore, also the two-point correlation

[2]The particular choice of $\Theta_\mathrm{e}^\mathrm{fixed}$ is not significant but the σ_{12} emulator here matches that of the Fourier space `COMET` and is therefore constructed at $z = 1, h = 0.695, A_\mathrm{s} = 2.2078559$ with all other possible evolution parameters set to zero

function) at a given cosmology and for a given redshift, and the three shape parameters $\omega_\mathrm{b}, \omega_\mathrm{c}$, and n_s. The parameter ranges covered by the training set match those of the Fourier space `COMET` and are given in Table 5.2. For the shape parameters, ω_b, ω_c and n_s the ranges were chosen to span approximately 12, 30 and 11σ interval around the *Planck* 2018 best-fit values respectively, whereas the choice for the minimum and maximum values of f and σ_{12} is determined by the need to capture the evolution across a range of redshifts and combinations of evolution parameters.

In terms of the growth rate, the chosen range supports any redshift for $\omega_c \gtrsim 0.107$ and the most extreme values of shape parameters put the lower boundary limitation $z \gtrsim 0.1$. However, the limiting factor that determines the redshift range of validity for `COMET` is the allowed values of σ_{12}. As shown in Figure 1 in Eggemeier et al. (2023), the chosen σ_{12} range puts the upper redshift boundary at around $z \sim 3$.

In addition to the emulator's native parameter space, `COMET` supports cosmology inputs in terms of $A_\mathrm{s}, h, w_0, w_\mathrm{a}, \Omega_\mathrm{K}$.

Training data

The training data for the configuration space `COMET` extension is obtained by Fourier transforming a modified set of the training data for the original power spectrum emulator.

The Fourier space training set is built on a Latin Hypercube and consists of 1500 samples covering the parameter space listed in Table 5.2. The sampling step was repeated 10,000 times to find an optimal set that maximises the minimum Euclidean distances between any two of its points. All of the model ingredients were evaluated using a numerical integrator and starting from `CAMB`-generated linear input power spectra.

In order to obtain noiseless configuration space predictions, two additional steps had to be taken before performing a Fourier transform of the original training data. The relation between Fourier space multipoles and their configuration space equivalents is given by:

$$\xi_{\mathcal{B},\ell}(r) = \frac{1}{2\pi^2} \int P_{\mathcal{B},\ell}(k) j_\ell(kr) k^2 dk, \tag{5.26}$$

where the integration is assumed to be over the full range of k. Nevertheless, in practice, the theory predictions only span a limited k range and, in particular, the truncation of the prediction at the maximum k limit shows up as a feature in Fourier space. This will subsequently result in an additional short-wavelength mode imposed over the configuration space theory prediction and show up as wiggles. In order to reduce this effect, we extend the Fourier-space measurements from $k_\mathrm{max} = 0.35\,\mathrm{Mpc}^{-1}$ to $k_\mathrm{max} = 15\,\mathrm{Mpc}^{-1}$ and apply an additional suppression factor, such that:

$$P'_{\mathcal{B},\ell}(k) = \exp\Big[- \frac{k}{k_\mathrm{cut}} \Big]^2 P_{\mathcal{B},\ell}(k), \tag{5.27}$$

where k_cut is a parameter that sets the scale of suppression. Our choice of k_cut is based on testing our $\xi_\ell(r)$ predictions obtained using different values of k_cut against the result obtained when Fourier transforming the full multipoles (in which we expect to have a smaller

Fourier transform error, as their calculation involves fewer transforms and the functions themselves are generally smoother than the individual power spectrum components). The test is performed by assuming a reference set of bias parameter values which match those listed in Section 5.3.1. We find that the resulting difference is the greatest on the monopole and on small scales and increases with increasing $k_{\rm cut}$ until converging to a constant value at $k_{\rm cut} \approx 2.0$. Based on these tests we choose a conservative value of $k_{\rm cut} = 2.5$. We calculate our training set for the separations s covering the range of $10\,{\rm Mpc} < s < 250\,{\rm Mpc}$ with 180 bins spaced linearly in the range of $10\,{\rm Mpc} < s < 200\,{\rm Mpc}$ and 20 more linearly spaced bins in the range of $200\,{\rm Mpc} < s < 250\,{\rm Mpc}$.

Finally, following the same approach as when building the Fourier space emulator, in order to reduce the dynamical range of the theory predictions the emulator is trained on, we pre-process each component by subtracting their mean training set value and dividing by the variance. For the σ_{12} emulator we additionally take the logarithm of each point in the training set before performing the normalisation as described above. For the $\xi_{\mathcal{B},l}(s)$ emulator we instead multiply the components by s^2 before the normalisation.

Gaussian process emulation

The emulator is built using Gaussian Processes (GP), as described in Rasmussen & Williams (2006). A key ingredient in GP emulation is the kernel function $K(\boldsymbol{x}, \boldsymbol{x'})$ - a description of the covariance between point $\boldsymbol{x}$ and point $\boldsymbol{x'}$ in the training set. For our configuration space emulator, we opt for the same kernel function as the one used to build the Fourier space `COMET` (which was itself chosen by comparing the performance of a number of typical functional forms found in the literature):

$$K(\boldsymbol{x}, \boldsymbol{x'}) = \exp\Big(-\frac{r^2}{2}\Big) + (1+\sqrt{3}r)\exp(-\sqrt{3}r) \tag{5.28}$$

with $r^2 = \sum_{i=1}^{d}(x_i - x_i')/l_i^2$. Here d is the dimension of the parameter space and $\boldsymbol{l}$ represents the hyperparameters that characterise the length scales of typical features in the training data. The first term of the kernel function is the squared exponential kernel, whereas the second term is the Matérn kernel of degree $\nu = 3/2$. The hyperparameters may differ for the two kernel function components and their values are optimised by maximising the log-likelihood of the GP models with respect to the training data. We repeat this step ten times, each with a different random initialisation. The GP emulation is implemented using the publicly available package `GPy` [3].

5.3 Validation

This section presents the results of the validation tests performed on the configuration space extension to the `COMET` emulator. In the first part of the section, we perform tests on the synthetic validation data set at a single redshift and with statistical uncertainties

[3] https://gpy.readthedocs.io/en/deploy/

Parameter	Minimum	Maximum
ω_b	0.02100	0.02365
ω_c	0.095	0.145
n_s	0.93	1.00
h	0.55	0.85
A_s	0.8	3.0

Table 5.3: The parameter space used to generate the validation sample.

that correspond to a volume 10 times larger than expected for the *Euclid* galaxy survey. In the second part of the section, we use our emulator for a cosmological MCMC analysis of galaxy clustering wedges measured on the MINERVA mocks and demonstrate that we can recover the input cosmology.

5.3.1 Recovering validation theory models

The validation sample consists of $P_{\mathcal{B},\ell}$ components calculated for a set of 1500 flat ΛCDM cosmologies defined by the parameters $\omega_\mathrm{b}, \omega_\mathrm{c}, n_\mathrm{s}, h$ and A_s. The parameter combinations were obtained by drawing random points within the minimum and maximum values as defined in Table 5.3. The validation set is generated for the redshift of $z = 0.904588119$ and a fiducial flat ΛCDM cosmology with $h = 0.67, \omega_\mathrm{b} = 0.0219961$ and $\omega_\mathrm{c} = 0.1212029$. When generating the validation set, the AP distortions are incorporated exactly (i.e., the approximation described in Section 5.2.3 is not used), however, just like in the training sample, we neglect the contribution from any terms involving the counterterm c_NLO (see Section 5.1.3).

In order to compare the validation set with the emulator predictions, the validation set $P_{\mathcal{B},\ell}$ components are used to construct the full multipoles in configuration space. We first calculate the power spectra multipoles by assuming the bias parameter values following the same assumptions as in the validation procedure performed by Eggemeier et al. (2023): the linear bias parameter is set by the relation $b_1 = \sqrt{1+z}$, which represents the expected bias of Hα galaxies selected by *Euclid* (Rassat et al., 2008; di Porto et al., 2012) and which is then used to determine the rest of the bias parameters through the following relations. For the tidal bias parameters γ_2 and γ_{21}, we use the excursion set and the coevolution relations by Sheth et al. (2013) and Eggemeier et al. (2019), respectively (which also mimics the modelling choices for BOSS and eBOSS cosmological analyses presented in this thesis). The value for the quadratic bias parameter b_2 is set based on the peak-background-split relation from Lazeyras et al. (2016) and all of the counterterm parameters are set to zero. Once the power spectrum multipoles are constructed, we apply the same damping factor as in equation (5.27) and perform a Fourier transform to obtain the configuration space prediction.

The statistical error for the validation set multipoles is computed assuming Gaussian errors and *Euclid* survey characteristics. In order to provide a conservative estimate of the

emulation accuracy, we scale the errors to represent 10 times the survey volume. To have a more representative error estimate, we furthermore perform a scaling with A_s, which is meant to correct for the fact that the initial variance is computed for a fixed cosmology and redshift, i.e., a fixed expected signal amplitude. The final error σ_i assigned to the validation multipoles at each cosmology i, therefore, follows $\sigma_i = (A_\mathrm{s}^i/A_\mathrm{s}^{fid})\sigma_\mathrm{Euc}/\sqrt{10}$, where A_s^{fid} is the value of A_s used to calculate the *Euclid* error σ_Euc and A_s^i is the value of A_s of the validation cosmology i.

In Figure 5.1 we present the mean error (solid line) and the standard deviation (vertical bars) for the multipole recovery. We calculate these quantities over the full validation sample and display them in terms of the scaled *Euclid* error σ, as described above. As illustrated in the plots, the greatest error appears on small scales, where the variance of the error across the validation cosmologies is also larger. At the minimum scales that the emulator is trained on ($s = 10$ Mpc), the error reaches $-1 \pm 4\sigma$ for ξ_0, $2 \pm 5\sigma$ for ξ_2 and $-4 \pm 6\sigma$ for ξ_4 - as a result, we do not recommend using the emulator at such small scales. The situation improves considerably at the minimum scales of $s = 20$ Mpc (marked by the vertical dashed line): here the validation set is accurately recovered with the error of approximately $-0.2 \pm 0.2\sigma$ for ξ_0, $0.2 \pm 0.3\sigma$ for ξ_2 and $-0.2 \pm 0.2\sigma$ for ξ_4. The mean emulation error and the standard deviation subsequently swiftly decrease further with increasing s until around $s = 230$ Mpc after which there is another slight increase, mostly affecting the monopole.

In order to get a better idea in terms of the rate of failure to recover the validation cosmology within 1σ, we plot a cumulative histogram of the fraction of samples versus the maximum difference between the emulated multipole and the validation set theory prediction, as shown in Figure 5.2. Here we consider the maximum error over, first, the full emulator range (starting with $s = 10$ Mpc, dashed lines) and, second, the reduced range, starting with $s = 20$ Mpc (solid lines). We see that, even for the full range of scales, 68% of the samples (indicated by the first dashed vertical line) recover the validation set at an accuracy of 1σ or better. Nonetheless, the rest of the samples show larger errors. As expected from the mean error results in Figure 5.1, the rate of failure decreases significantly, once the smallest scales are removed and we consider the minimum separation of $s = 20$ Mpc. In this case, all of the hexadecapole predictions as well as all but one prediction of monopole and quadrupole are recovered to better than 1σ accuracy, with 68% of samples showing a maximum error of $\sim 0.3\sigma$ or less and 98% samples showing a maximum error of $\sim 0.6\sigma$ or less. In general, all multipoles are recovered with similar accuracy, although the hexadecapole shows a somewhat better emulation performance, mainly due to the fact that the *Euclid* error used to assess the accuracy for ξ_4 is the greatest and the emulation errors for this multipole are highly localised at the smallest scales below $s = 20$ Mpc, as also seen in 5.1.

Based on the results presented in this section, we conclude that our emulator can accurately recover theory predictions for all three multipoles in the cosmologies covered by the validation set on scales $20\,\mathrm{Mpc} < s < 250\,\mathrm{Mpc}$.

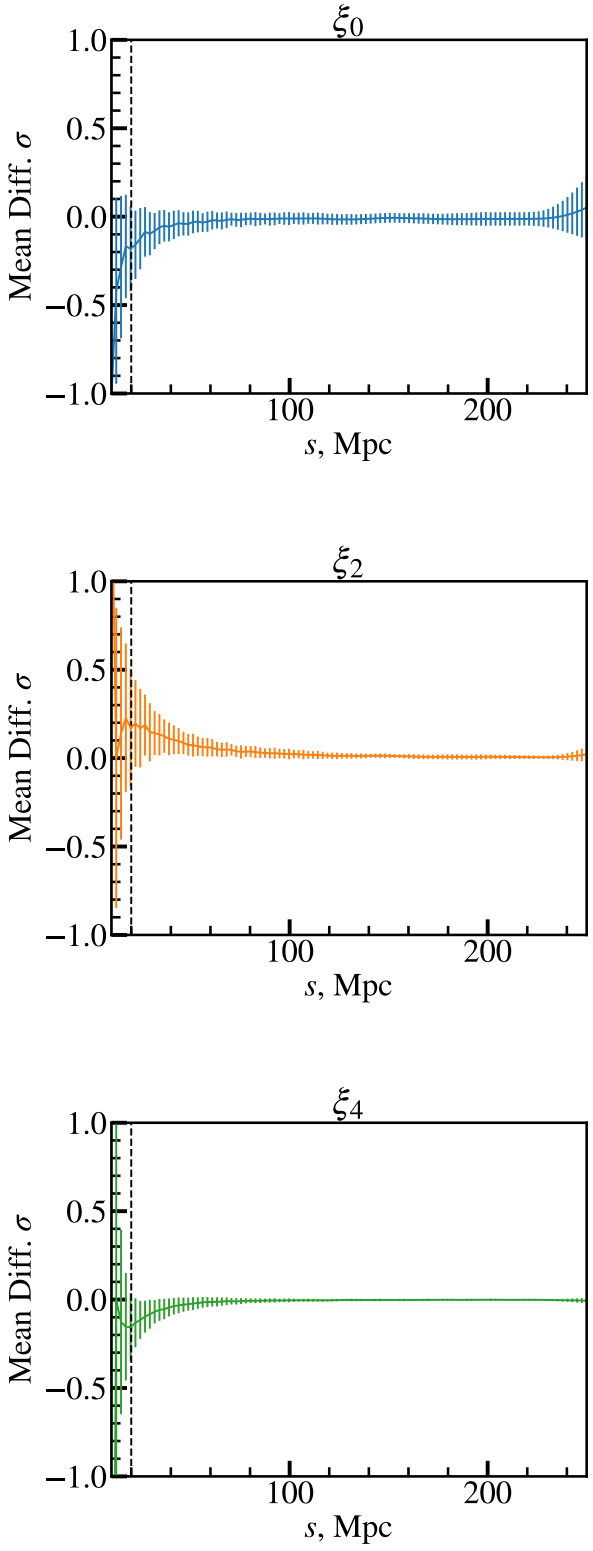

Figure 5.1: The mean difference between validation set and emulated multipole for the monopole (ξ_0, top), quadrupole (ξ_2, middle) and hexadecapole (ξ_4, bottom). The coloured vertical lines indicate the standard deviation of the difference which is calculated over all cosmologies in the validation set and is expressed in units of the standard deviation σ, corresponding to a measurement in 10 times the *Euclid* volume. The dashed line marks the scale of $s = 20$ Mpc. The multipoles are calculated using the fiducial cosmology and bias parameter values described in Section 5.3.

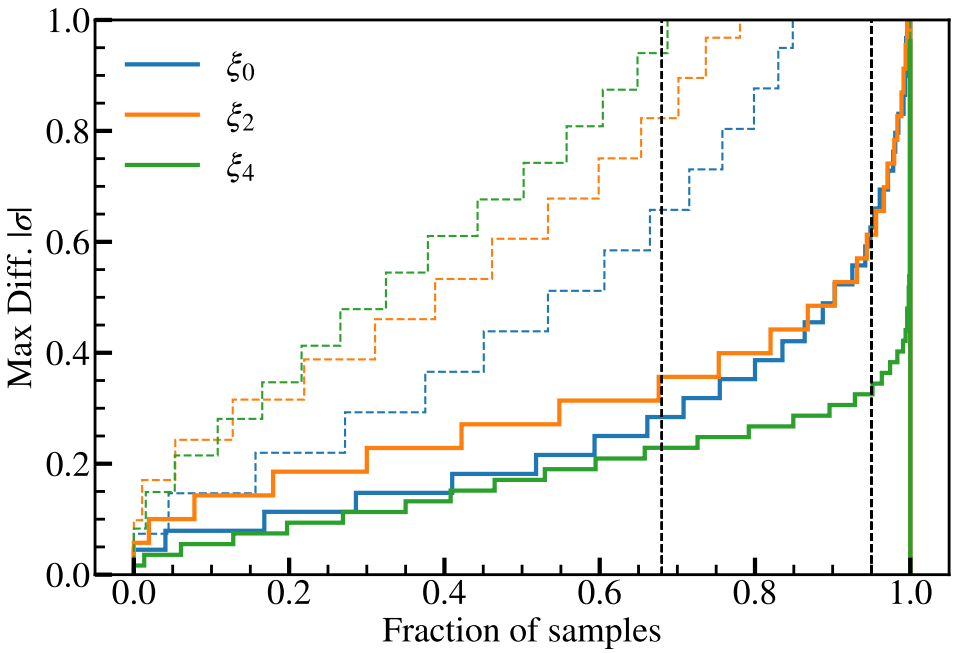

Figure 5.2: Cumulative histogram of the fraction of samples with a given maximum absolute difference from the validation set multipole ξ_0 (blue), ξ_2 (orange) and ξ_4 (green). The dashed coloured lines show the result when the minimum scale taken into account matches the minimum scale the emulator is trained on $s = 10$ Mpc, whereas the solid lines show the results with the recommended minimum scale $s = 20$ Mpc. The vertical dashed lines mark 68% and 95% of the validation samples.

5.3.2 Mock cosmological analysis

While the emulator showed an excellent performance predicting the multipoles of the validation set, it is important to bear in mind that the validation theory predictions were produced neglecting the short-scale velocity contribution characterised by $c_{\rm NLO}$. In order to further confirm that the emulated theory model is sufficient to recover unbiased cosmological constraints, we use our emulator to analyse galaxy clustering wedges measured on Minerva mocks, which were used to validate the theory model for BOSS galaxy clustering and which are described in Section 3.2. We perform the MCMC sampling using the publicly available `COBAYA` package (Torrado & Lewis, 2021). Our priors are listed in Table 5.4 and are chosen to match the ranges of the cosmological parameters of the training set or the cosmological analyses carried out in Chapters 3 and 4 (though note that the resulting priors are narrower than the ones used in the cosmological analyses due to the more limited shape parameter range that the emulator is valid for). We use the co-evolution relations for tidal bias parameters γ_2 and γ_{21} (as described in Section 2.5.1).

We present the resulting parameter constraints in Figure 5.3. For comparison, we also

Parameter	Prior
ω_b	U(0.0205, 0.02415)
ω_c	U(0.085, 0.155)
n_s	U(0.92, 1.01)
h	U(0.50, 0.90)
$\ln10^{10}\mathrm{A_s}$	U(1.5, 4.0)
b_1	U(0.5, 9.0)
b_2	U(-4, 8.0)
c_0	$\mathcal{N}(0, 30)$
c_2	$\mathcal{N}(0, 30)$
c_4	$\mathcal{N}(0, 30)$

Table 5.4: The priors used for the cosmological analysis of the MINERVA wedges with the configuration-space extension of `COMET`. Here U marks the uniform prior with the minimum and maximum limits presented in the brackets. $\mathcal{N}(x, y)$ indicates the normal distribution with mean (x) and standard deviation (y). The priors for the shape parameters were chosen to match the training sample ranges listed in Table 5.2, whereas the ranges for h, $\ln10^{10}\mathrm{A_s}$, b_1 and b_2 were taken from the cosmological analyses in Chapters 3 and 4, as listed in Table 3.1.

display the constraints obtained from the model validation performed in Section 3.2 (Figure 3.3). Note that the constraints taken from Figure 3.3 are not only obtained with different modelling of both the non-linear power spectrum (RESPRESSO for the exact model, versus EFT for the emulator-based predictions) and the small scale RSD (the EFT-based model differs from the VDG prescription as stated in equation (5.1) and we additionally neglect the c_NLO term.), but also wider priors. As a result, we do not expect the two sets of contours to match.

The cosmological constraints obtained using emulator-based modelling recover the input cosmology within 1σ. The resulting contours are, generally, tighter than those in the original validation analysis because of the narrower priors, however, the input cosmology recovery is somewhat more biased. This is particularly evident in the constraints for σ_{12} and A_s, indicating that the amplitude of the power spectrum is modelled to be somewhat high. This might be the result of a number of effects: as mentioned before, EFT-based analyses have been shown to depend on the priors used (Carrilho et al., 2023; Simon et al., 2022). Our choice of priors for the counterterms is somewhat more narrow than in many cosmological analyses. Indeed, Carrilho et al. (2023) in particular show that using wider priors on nuisance parameters (including the counterterms) shifts the A_s constraints to lower values. Furthermore, our incomplete modelling of the FoG effect may also influence the final constraints. This test suggests that in order to improve the accuracy with which the input cosmology is recovered, further investigation on the analysis setup is needed.

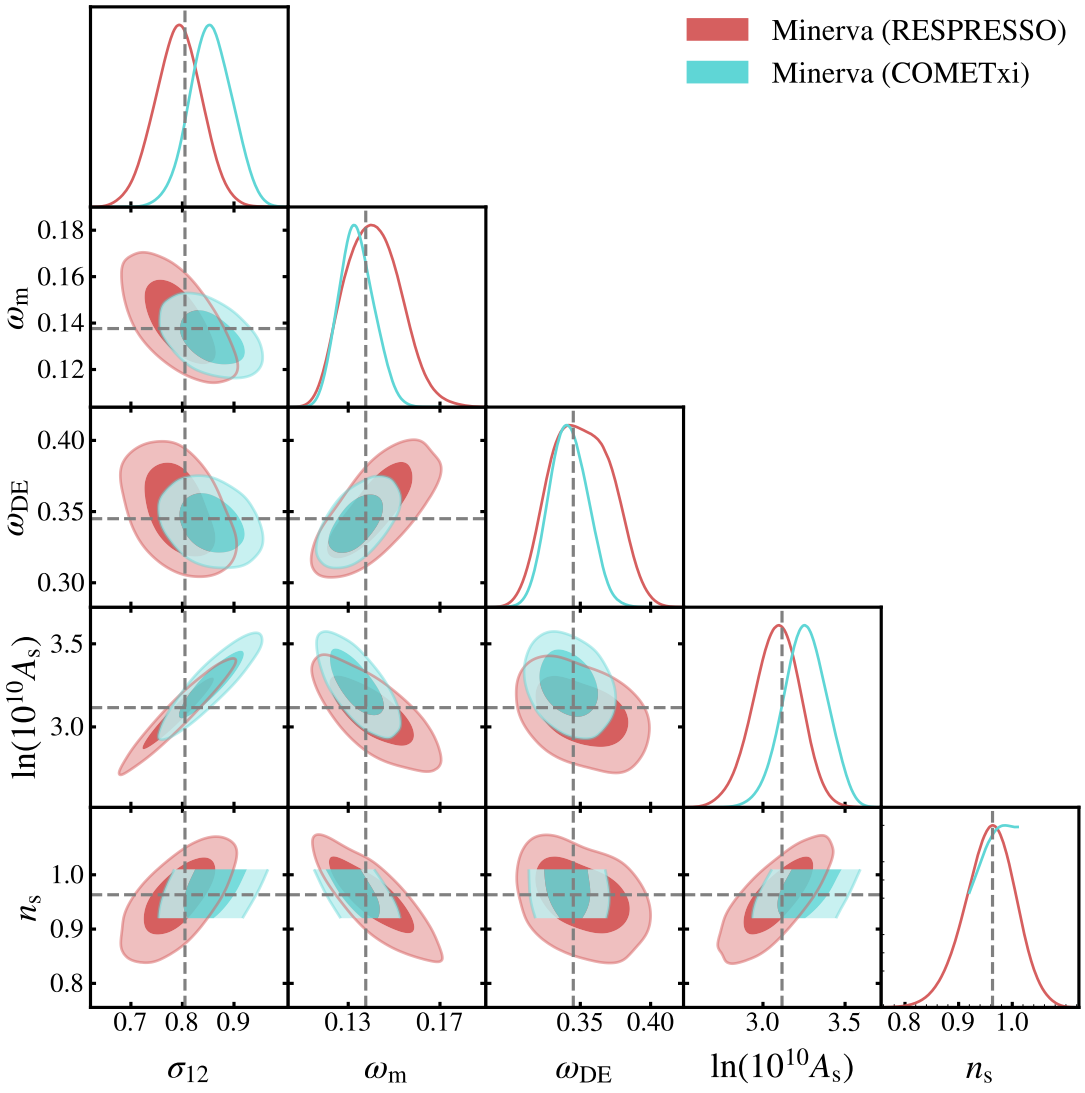

Figure 5.3: Flat ΛCDM constraints from galaxy clustering wedges as measured on MINERVA mocks. We show the constraints obtained using our configuration space extension of `COMET` - `COMETxi`. For the reference we also provide the constraints obtained when validating the RESPRESSO based power spectrum model, used for analyses in Chapters 3 and 4. The fiducial cosmology values are indicated by the dotted lines.

5.4 Summary

In this chapter, we presented the development of a configuration space extension of the evolution-mapping-based emulator `COMET`. We use gaussian processes to train an emulator on the Fourier-transformed EFT predictions for the power spectrum components (as listed in Table 5.1 and throughout Section 5.1). Our emulator produces predictions for the redshift-space galaxy two-point correlation function multipoles ξ_0, ξ_2, ξ_4, and, following the original design, our extension is built to support arbitrary fiducial cosmologies as well as a continuous range of redshifts up to approximately $z < 3$. This is done by training an emulator on the theory predictions as a function of shape parameters, σ_{12} and f and then scaling the resulting multipole components to the target amplitude, which is itself obtained from a σ_{12} emulator trained on different combinations of shape parameters.

We validate our emulator for the cosmological parameter ranges that match those of the Fourier space `COMET` (and as listed in Table 5.3) and for the scales of $20\,\mathrm{Mpc} < s < 250\,\mathrm{Mpc}$. We show that for these ranges the emulated multipoles are consistent with the validation theory predictions at the level of below 1σ, where σ stands for an error of a survey with the volume of 10 times that of *Euclid*. We note that 98% of validation samples are recovered at an accuracy of 0.6σ or better.

Our emulator validation on cosmological analysis of galaxy clustering in the Minerva mocks confirms that the emulator is able to recover the input cosmology, however, we note that the constraints obtained for both A_s and σ_{12} are rather high with the fiducial values falling within the lower limit of the one standard deviation interval of the respective posteriors. However, this result is expected to depend on the priors on the nuisance parameters used and requires further investigation. We also note that in the current implementation, the configuration space modelling lacks a complete treatment of the FoG effect which can also have an effect on the final constraints. The VDG treatment of the RSD from small-scale velocity contributions is more appropriate for the analyses in configuration space (as we saw, in perturbation theory the terms that are used to correct for this effect diverge and, therefore, cannot be accurately Fourier transformed) and the results from the mock analyses further confirm the importance of such modelling choices in the final constraints. The speedup that our emulator provides will allow us to test the effect of any such modelling assumptions more easily, leading to more robust analyses.

The configuration space extension allows for a speedup in theory predictions not only with respect to the traditional exact computation but also compared to the analysis using the Fourier space emulation and performing the transform to configuration space. Our training set is pre-processed to ensure an accurate Fourier transform which further increases the robustness of configuration-space analyses with `COMET`. The average time for one evaluation of all three multipoles for 300 separation bins is around 5ms.

The other available two-point correlation function emulators in the literature mostly focus on small-scale clustering and are, therefore, trained on simulations and halo model-based statistics (for example, Zhai et al., 2019). Perhaps the closest to our work in terms of the range of scales emulated is the Dark Emulator (Nishimichi et al., 2019; Cuesta-Lazaro et al., 2023), which covers the separations of $0.01h^{-1}\mathrm{Mpc} < r < 150h^{-1}\mathrm{Mpc}$. The

emulator is based on the DarkQuest simulation suite and employs the halo model to be able to provide predictions at significantly smaller scales than our `COMET` extension. Nevertheless, it is currently only available for fixed redshifts and does not include redshift-space predictions. Dark Emulator is also considerably slower (the quoted single evaluation time is 300ms compared to ours 5ms) and, while it also allows for sampling of the halo model parameters, it covers a more limited cosmological parameter space (which includes the ΛCDM parameters and w but does not allow for its time evolution or curvature). For the ΛCDM parameters, Dark Emulator supports narrower ranges for all cases but n_s. Our configuration space emulator is, therefore, uniquely suitable for full-shape galaxy clustering analyses up to mildly-non-linear scales. It could additionally be used in tandem with Dark Emulator to cover a wider range of scales and test the consistency between the halo model and perturbation theory-based predictions on the scales on which the two emulators overlap.

Chapter 6

Summary and outlook

Galaxy clustering is a well-established cosmological probe, offering some of the most precise low-redshift cosmological parameter constraints. However, in order to extract and correctly interpret the full information from increasingly richer data sets, it is important to reassess the traditional methods and assumptions used and be ready to incorporate new approaches into our analyses. This thesis offers the insights gained from one such new approach - evolution mapping, and presents its applications to the anisotropic galaxy clustering analyses.

In Chapter 3 we presented the full-shape analysis of BOSS galaxy and eBOSS quasar clustering within ΛCDM cosmologies. We analysed the galaxy clustering wedges from Sánchez et al. (2017) and quasar clustering multipoles from Hou et al. (2021). Together the two sets of measurements allowed us to trace the background evolution of the Universe as well as the structure growth through an extensive redshift range of $0.2 < z < 2.2$. Our full-shape model contained several important updates from the original analyses. First, the more accurate non-linear power spectrum predictions were obtained using RESPRESSO, matching the analysis of Hou et al. (2021) but unlike in Sánchez et al. (2017) who used gRPT-based modelling. Second, we used co-evolution relations from Eggemeier et al. (2019) to set the values of the tidal bias parameters - this allowed us to reduce the sampled parameter space and, therefore, improve the precision of the resulting constraints. We validated this model on Minerva and OuterRim mocks and confirmed that it recovers unbiased input cosmology.

We were interested in how the recovered galaxy clustering constraints compare with and complement the constraints from weak lensing two-point statistics (the 3×2pt measurements) from DES Y1 and CMB power spectra measurements by Planck. We furthermore performed our analysis in the physical parameter space - i.e., we were interested in cosmological parameters not defined through the dimensionless Hubble parameter, h. We illustrated, by focusing on the posteriors of σ_8, the clustering amplitude defined on h^{-1}Mpc scale, how the constraints on this parameter obtained from different probes cannot be meaningfully compared, as they each measure a weighted clustering amplitude over different ranges of scales, defined by the corresponding h posterior. We furthermore demonstrated that, when using the physical parameter space instead, we are able to correctly

recover the tight degeneracy between the initial and final amplitudes of density fluctuations $\log(10^{10}A_{\rm s})$ and σ_{12}. This subsequently allowed us to demonstrate that, first, the clustering amplitude, σ_{12} recovered by all probes is consistent with each other and, second, in the physical parameter space the σ_8 tension instead shows up as tension in the amount of total structure growth preferred by the combined low redshift probes and Planck, with Planck preferring a greater amount. We found that the tension in the $\log(10^{10}A_{\rm s}) - \sigma_{12}$ parameter subspace between the combined BOSS+eBOSS+DES measurements and Planck is 2.5σ, which is similar to the level of tension in S_8 for these data sets.

Finally, we carried out an additional consistency test by repeating our base analysis with narrow priors, taken from Planck posterior constraints, on the shape parameters $\omega_{\rm b}, \omega_c, n_{\rm s}$. The additional information from Planck allowed us to obtain tighter constraints on the parameters that determine the power spectrum amplitude. The increased constraining power resulted in low redshift probes showing a preference for 1.89σ and 1.22σ lower σ_{12} and $\log(10^{10}A_{\rm s})$ values than those obtained by Planck. Additionally, the low redshift probes recovered a 1.73σ higher value of $\omega_{\rm DE}$. While these differences are not significant, they are consistent with the tension in the amount of total growth observed in the base analysis with greater dark energy density suppressing the structure growth. However, as none of our probes provides tight constraints on $\omega_{\rm DE}$, any disagreements in the inferred value could not be measured directly.

Motivated by these findings, in Chapter 4, we repeated our analysis for ΛCDM extensions with varying dark energy equation of state parameter w. In addition to this, we explored models with the time evolution of the equation of state parameter (varying $w_{\rm a}$), curvature ($\Omega_{\rm K}$) and the neutrino mass sum ($\sum m_\nu$). Once again, we performed the analysis in the physical parameter space and demonstrated its clear advantages for these cosmologies, as they represent a case where constraints on H_0 are particularly degraded. This is especially true for Planck and we illustrated that by showing that clustering amplitude is constrained by the CMB probes even in the cosmologies with varying dark energy density, as long as it is measured on a physical scale defined in Mpc. In this work, we, therefore, presented the first-ever CMB constraints on the linear density field variance, as measured by σ_{12}, in cosmologies with varying w. We showed that this value is defined by the $w - \omega_{\rm DE}$ degeneracy, which corresponds to the constant angular diameter distance to the last scattering surface. We obtained Planck wCDM constraint of $\sigma_{12} = 0.816 \pm 0.011$, which is only slightly higher than the ΛCDM value and which degrades little even when w_a is allowed to vary as well.

In addition to this, we demonstrated that CMB can also constrain the physical curvature density of the Universe, $\omega_{\rm K}$, even when w is allowed to vary. We confirmed that Planck prefers negative curvature of $\omega_{\rm K} = -0.0116^{+0.0029}_{-0.0036}$, as reported in previous work, however, the physical curvature result carries the higher significance of 4σ. Nevertheless, we also confirmed that BOSS+eBOSS, which constrain $\omega_{\rm K}$ to be consistent with 0, is discrepant with Planck when the curvature is allowed to vary, even in the evolving dark energy density scenarios. As reported in previous work, the closed Universe result by Planck can be linked back to its preference for excess lensing of the temperature power spectrum (characterised by $A_{\rm lens} > 1$).

Overall, with the exception of Planck's curvature constraint, we found no significant preference for any of the extended parameters taking non-ΛCDM values. For the neutrino mass sum, we found that the combination of Planck+BOSS+eBOSS only offers a modest improvement in comparison to Planck's constraints alone. Our best upper mass limit, therefore, comes from the combination Planck+BOSS+eBOSS+SN and is $\sum m_\nu < 0.211$, which is less constraining than the SDSS consensus analysis, which, however, makes use of a more extensive data set.

In the final Chapter 5, we proceeded to make use of the evolution mapping approach, which we now applied to the power spectrum modelling with the goal of building an emulator for the redshift-space galaxy two-point correlation function multipoles. We followed the design of `COMET`, the evolution-mapping-based Gaussian process emulator for the power spectrum multipoles, and built its extension. The configuration-space emulator provides predictions for the two-point correlation function monopole, quadrupole, and hexadecapole, based on the EFT recipe for the non-linear power spectrum. The use of evolution mapping allowed us to reduce the parameter space on which the emulator needed to be trained. This was reflected both by its performance and flexibility in terms of the number of cosmological models and redshifts emulated. Our configuration space emulator, just like the original `COMET`, allows for the varied standard ΛCDM parameters ω_b as well as the common extensions $\Omega_\mathrm{K}, w_0, w_a$. In addition to the extensive cosmological parameter coverage, the emulator also supports arbitrary fiducial cosmologies and a wide redshift range, stretching up to the redshift of $z \sim 3$.

We validated our emulator on a synthetic data set which was assigned the errors corresponding to 10 times the volume of *Euclid*. We confirmed that the validation multipoles were recovered within 1σ for all but one case in the quadrupole and hexadecapole for the scales of $20\,\mathrm{Mpc} < s < 250\,\mathrm{Mpc}$. Moreover, 98% of the validation samples were recovered with an accuracy of $\sim 0.6\sigma$ or better and 68% of samples were recovered with an accuracy of $\sim 0.3\sigma$ or better. We furthermore used our emulator to perform the cosmological analysis of galaxy clustering wedges measured on the Minerva mocks. We found that the resulting constraints recovered the input cosmology within 1σ, albeit with somewhat high values for the power spectrum amplitude parameters A_s and σ_{12}. This suggests that further work is needed to determine the influence of the priors as well as the modelling of the virialised small-scale galaxy velocity contribution.

First and foremost our future work will, therefore, include the development of an emulator for the configuration space VDG model to further investigate the influence that the modelling of the small-scale velocity contribution has on the cosmological parameter constraints. Our emulator currently exhibits substantial errors on the small-scale theory predictions for $10\,\mathrm{Mpc} < s < 20\,\mathrm{Mpc}$. We would like to further investigate how this can be improved to extend the validity range of the emulator. One of the possible avenues to explore is implementing additional weights on the different scales emulated such that, ideally, we can obtain an emulation accuracy that scales with the measurement errors expected.

We are interested in making use of `COMET` as well as the ideas behind evolution mapping to further explore how the full-shape clustering information can be appropriately summarised. In addition to the fact that the summary parameters do not contain the

information from the shape of the two-point correlation function (Brieden et al., 2022), another issue is related to the use of $f\sigma_8$ parameter combination to characterise the RSD effect on the two-point correlation function. As described throughout this thesis and as pointed out in Sánchez (2020), the use of h^{-1}Mpc units introduces an additional dependence on the choice of h. This is particularly an issue when the value of $f\sigma_8$ is obtained from the so-called "fixed template" analysis where the parameters describing the shape and the amplitude of the matter power spectrum are kept fixed. There has been some evidence that there is a mismatch between the constraints obtained from the fixed template approach as compared to the equivalent results obtained from fitting the measurements directly while varying the cosmological parameters (see, for example, Maus et al., 2023). This mismatch can be traced back to the fact that in the fixed template analysis, $f\sigma_8$ is measured at a fixed h, which reduces the resulting error as compared to the case where h is varied and marginalised over. An interesting avenue for future work is then to update the RSD constraints from SDSS galaxy clustering measurements by re-expressing them in terms of $f\sigma_{12}$ and explore how these evolve with redshift and compare with the full shape constraints.

It would also be interesting to repeat our cosmological analysis presented in Chapter 3 updated with the most recent weak lensing measurements from KiDS-1000 (Heymans et al., 2021) and DES Y3 (Collaboration et al., 2021) as well as the additional growth of structure constraints from the SDSS $f\sigma_{12}$ measurements.

Nevertheless, ultimately, it is the upcoming galaxy surveys that promise the constraining power required to more decisively confirm whether any discrepancy between the high and low redshift data sets exists and whether it points towards models beyond the standard ΛCDM. The advent of new large, high-quality data sets (the so-called Stage-IV surveys) such as the ones obtained by the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al., 2016), the ESA space mission *Euclid* (Laureijs et al., 2011), and the Legacy Survey of Space and Time (LSST) at the Rubin Observatory (Ivezić et al., 2019), will allow us to combine multiple probes and significantly tighten our cosmological constraints. An impressive example of the unprecedented size of these data sets is the recently started DESI, which has already obtained spectra of more galaxies than available in the full BOSS sample. At the end of its five-year operation, DESI is expected to have collected a total of 30 million galaxy and quasar redshifts. The space-based *Euclid* mission that was launched in July 2023 and has already obtained its very first images is expected to observe around 10 billion sources with 30 million of them expected to be used for galaxy clustering studies. Finally, the photometric LSST survey will offer complimentary observations of the immense sample of 20 billion galaxies over the period of 10 years.

The discussion of the consistency between different data sets has so far been limited to the best-constrained parameters with cosmological analyses employing the methods developed when the constraining power available was a fraction of what is expected from future observations. The use of h^{-1} Mpc units and the parameters defined through them is one example of such a tradition. As we move on to the analysis of Stage IV data sets, it would be beneficial to shift our focus towards the quantities that more closely represent the cosmological information content of those data, or that have a more direct

physical interpretation. The future surveys described here will be used to complement each other and it is, therefore, more important than ever that this is done in a consistent and robust manner, bearing in mind the assumptions that have gone in when developing the frameworks used to analyse each separate probe. The work presented in this thesis has demonstrated that tighter constraints on ΛCDM extensions are available even with the current data - we may hope that by making use of the future observations we will be able to increase the precision of these measurements and reduce the extended parameter space. In order to perform these analyses we will need efficient and accurate pipelines and emulation will be a key ingredient in developing them as well as testing the various model assumptions. While the first Stage IV data is already being collected, the validity of the currently available emulators is still very limited - we hope that our work represents a significant step towards fulfilling the demands of these near-future analyses. Cosmology has come a long way since the very first observations. In order to make the most of the potential offered by the incredible data sets that are about to be available, we must be ready to question our methods and meet the challenges offered by the new generation of surveys.

Appendix A

Joint analysis with KiDS-450 data

In this section, we present the joint analysis of the anisotropic clustering measurements from BOSS and eBOSS together with cosmic shear measurements from KiDS-450, Hildebrandt et al., 2016.

We use cosmic shear measurements from the Kilo-Degree Survey, (Kuijken et al., 2015; Hildebrandt et al., 2016; Fenech Conti et al., 2017), hereafter referred to as KiDS. The KiDS data are processed by THELI (Erben et al., 2013) and Astro-WISE (Begeman et al., 2012; de Jong et al., 2017). Shears are measured using lensfit (Miller et al., 2013), and photometric redshifts are obtained from PSF-matched photometry and calibrated using external overlapping spectroscopic surveys (see Hildebrandt et al. (2016)).

The KiDS-450 weak lensing data set consists of tomographic shear measurements from four redshift bins spanning the total range of $0.1 < z \leq 0.9$ and the corresponding source redshift distributions estimated from the weighted direct calibration ('DIR') for each bin (Lima et al., 2008). We use the recommended scale cuts and use the angular bins with $\theta < 72\,\mathrm{arcmin}$ for $\xi_+(\theta)$ and $\theta > 6\,\mathrm{arcmin}$ for $\xi_-(\theta)$.

It is important to note that here we are using the same DES shear model as in the main analysis. This means that the treatment of the nuisance parameters (namely, the baryonic effects, the photometric redshift uncertainty and the additive and multiplicative bias parameters) differs from the original analysis of KiDS-450. We do not include baryonic effects and our priors for photometric redshift uncertainty match those of Hildebrandt, H. et al. (2020). We impose flat priors for multiplicative bias $U(-0.1, 0.1)$ with the additive bias parameter taken to be zero, mimicking the DES set up. Finally, we also follow the DES intrinsic alignment model, imposing a flat prior on redshift evolution of intrinsic alignment parameter. We compare our final posteriors with the ones obtained from the publicly available KV450 chains (Hildebrandt, H. et al., 2020) and find a good agreement between the two, as shown in Fig. A.1. Moreover, there is a weak tendency for our analysis to prefer lower RMS variance values, which leads to a more conservative assessment of any potential tensions with Planck.

Fig. A.2 shows the combined constraints from the combination of KiDS-450 + BOSS +eBOSS (orange contours), which are in near-perfect agreement with the equivalent combination using DES (grey contours). The suspiciousness statistic shows an agreement between Planck

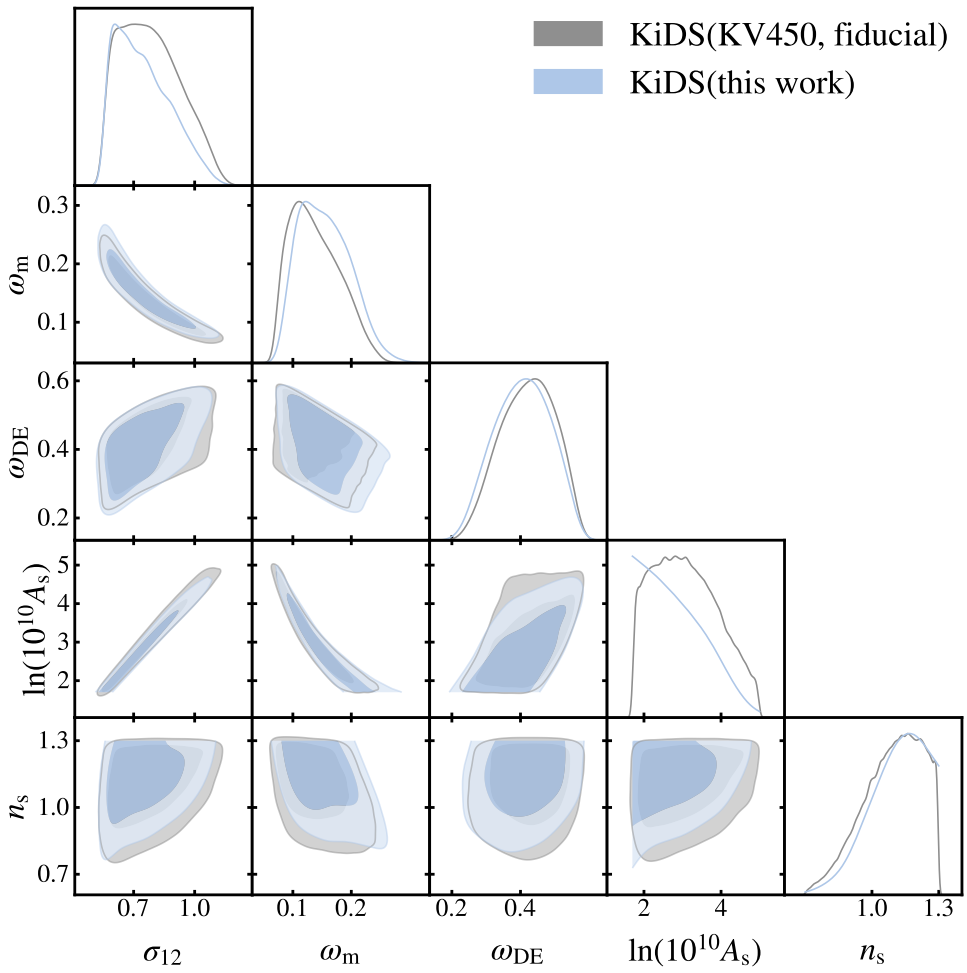

Figure A.1: Comparison of the marginalised posterior distributions between the fiducial KV450 analysis and this work. For this comparison, we adapted the cosmological parameter priors to match those of the fiducial analysis.

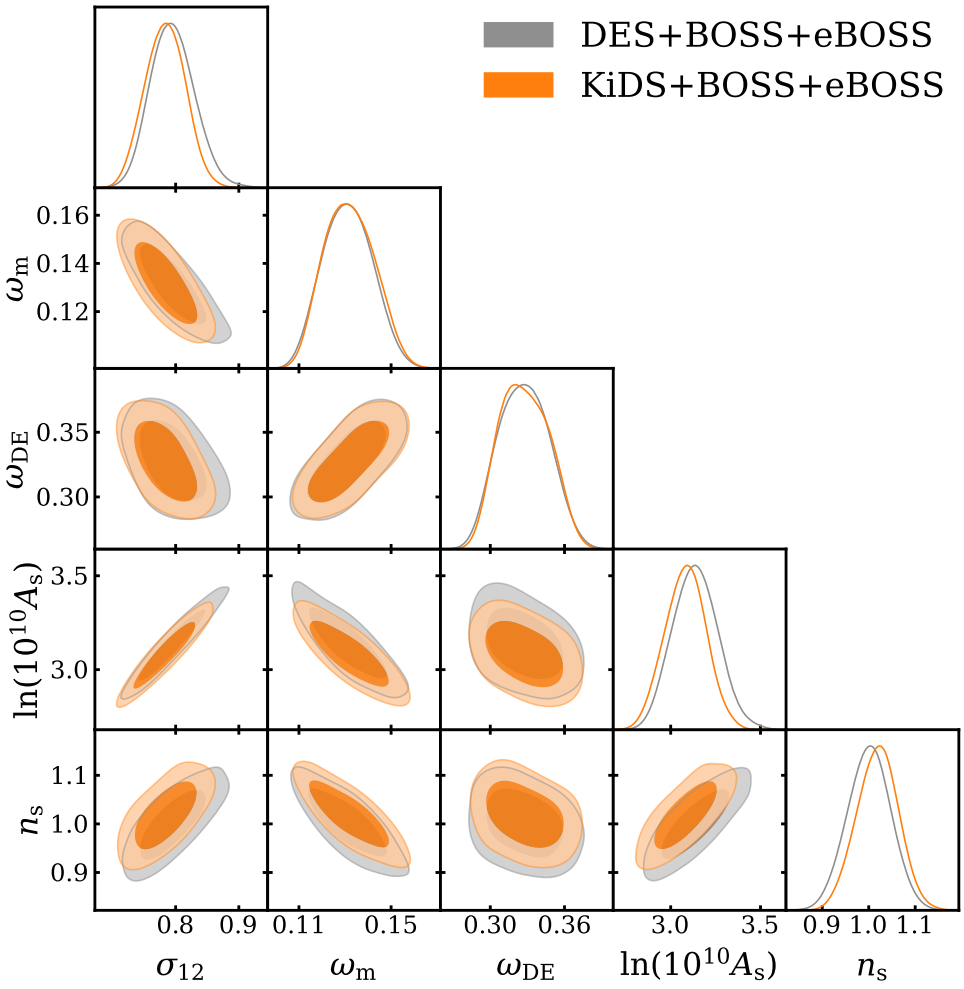

Figure A.2: Comparison of low-redshift constraints obtained by combining BOSS + eBOSS with either DES Y1 3 × 2pt data (grey contours) or the KiDS-450 shear measurements (orange contours).

and this set of low-redshift measurements of $1.5 \pm 0.5\sigma$, which is also consistent with our results from BOSS + eBOSS + DES.

Appendix B

Additional beyond-ΛCDM constraints

In this section, we present the constraints on parameters omitted in Table 4.2, including constraints in the "traditional" parameter space (Table B.1). The physical parameters (i.e., not defined through h units) constrained by our data include the physical matter density $\omega_{\rm m}$, the spectral index $n_{\rm s}$ and the (log) amplitude of initial density fluctuations $\ln 10^{10} {\rm A_s}$. For completeness, we include the traditional parameters σ_8 (linear density field variance as measured on the scale of $8h^{-1}$ Mpc whose physical equivalent is σ_{12}), Hubble parameter H_0 and the relative densities of matter ($\Omega_{\rm m}$), dark energy ($\Omega_{\rm DE}$) and curvature ($\Omega_{\rm K}$).

In general, we expect that the constraints on these parameters are degraded in comparison to their physical equivalents due to averaging over the posterior of H_0, which tends to be less well constrained in these extended cosmologies (this is most evident when comparing σ_{12} with σ_8).

Table B.1: Marginalised posterior constraints (mean values with 68 per-cent confidence interval, for $\sum m_\nu$ - 95 per-cent confidence interval) derived from Planck CMB and the full shape analysis of BOSS + eBOSS clustering measurements on their own, as well as in combination with each other and with Pantheon supernovae Ia measurements (SN). All of the models considered here vary dark energy equation of state parameter w, w_aCDM additionally allows a redshift evolution for w, wKCDM varies curvature and $w\nu$CDM varies neutrino mass sum $\sum m_\nu$. Note that for wKCDM the joint BOSS+eBOSS+Planck constraints should be interpreted bearing in mind that BOSS+eBOSS and Planck are discrepant in this parameter space.

		BOSS+eBOSS	BOSS+eBOSS+Planck	BOSS+eBOSS+Planck+SN
wCDM	σ_8	0.798 ± 0.047	0.828 ± 0.017	0.818 ± 0.012
	H_0	$69.8^{+3.1}_{-3.6}$	$69.6^{+1.4}_{-1.6}$	68.62 ± 0.84
	Ω_m	$0.280^{+0.017}_{-0.021}$	0.295 ± 0.013	0.3026 ± 0.0080
	Ω_Λ	$0.720^{+0.021}_{-0.017}$	0.705 ± 0.013	0.6974 ± 0.0080
	ω_m	$0.137^{+0.011}_{-0.013}$	0.1426 ± 0.0011	0.1424 ± 0.0011
	n_s	0.990 ± 0.055	0.9661 ± 0.0042	0.9665 ± 0.0041
	$\ln 10^{10} \mathrm{A_s}$	3.02 ± 0.21	3.043 ± 0.016	3.044 ± 0.016
w_aCDM	σ_8	0.793 ± 0.045	$0.818^{+0.019}_{-0.022}$	0.822 ± 0.012
	H_0	$70.1^{+3.7}_{-4.4}$	$68.1^{+2.0}_{-2.8}$	68.72 ± 0.86
	Ω_m	$0.281^{+0.026}_{-0.030}$	$0.309^{+0.024}_{-0.021}$	0.3025 ± 0.0081
	Ω_Λ	$0.719^{+0.030}_{-0.026}$	$0.691^{+0.021}_{-0.024}$	0.6975 ± 0.0081
	ω_m	0.138 ± 0.012	0.1428 ± 0.0012	0.1427 ± 0.0011
	n_s	0.983 ± 0.054	0.9656 ± 0.0042	0.9658 ± 0.0041
	$\ln 10^{10} \mathrm{A_s}$	3.00 ± 0.20	3.042 ± 0.016	3.042 ± 0.016
wKCDM	σ_8	0.770 ± 0.049	0.834 ± 0.019	0.819 ± 0.012
	H_0	$68.9^{+2.9}_{-3.4}$	$69.7^{+1.4}_{-1.6}$	68.46 ± 0.91
	Ω_m	0.292 ± 0.019	0.292 ± 0.014	$0.3032^{+0.0077}_{-0.0086}$
	Ω_Λ	0.829 ± 0.073	0.710 ± 0.015	$0.6982^{+0.0085}_{-0.0076}$
	Ω_K	-0.121 ± 0.078	-0.0025 ± 0.0026	-0.0014 ± 0.0024
	ω_m	$0.139^{+0.011}_{-0.013}$	0.1419 ± 0.0013	0.1420 ± 0.0013
	n_s	$0.975^{+0.060}_{-0.053}$	0.9679 ± 0.0046	0.9676 ± 0.0045
	$\ln 10^{10} \mathrm{A_s}$	2.80 ± 0.26	3.041 ± 0.016	3.043 ± 0.016
$w\nu$CDM	σ_8	$0.795^{+0.041}_{-0.048}$	0.822 ± 0.018	$0.816^{+0.016}_{-0.013}$
	H_0	70.3 ± 3.5	$70.0^{+1.5}_{-1.8}$	68.66 ± 0.85
	Ω_m	$0.288^{+0.019}_{-0.022}$	0.293 ± 0.013	0.3028 ± 0.0083
	Ω_Λ	$0.712^{+0.022}_{-0.019}$	0.707 ± 0.013	0.6972 ± 0.0083
	ω_m	$0.143^{+0.012}_{-0.014}$	$0.1433^{+0.0013}_{-0.0016}$	0.1427 ± 0.0013
	n_s	$1.066^{+0.070}_{-0.11}$	0.9659 ± 0.0039	0.9665 ± 0.0041
	$\ln 10^{10} \mathrm{A_s}$	$3.18^{+0.23}_{-0.26}$	3.043 ± 0.016	3.045 ± 0.016

Bibliography

Abbott T. M. C., et al., 2018, Phys. Rev., D98, 043526

Abbott T. M. C., et al., 2019, Phys. Rev. D, 99, 123505

Ahumada R., et al., 2020, ApJS, 249, 3

Alam S., et al., 2015, ApJS, 219, 12

Alam S., et al., 2017, MNRAS, 470, 2617

Alam S., et al., 2021, Phys. Rev. D, 103, 083533

Alcock C., Paczyński B., 1979, Nature, 281, 358

Angulo R. E., Zennaro M., Contreras S., Aricò G., Pellejero-Ibañez M., Stücker J., 2021, MNRAS, 507, 5869

Aricò G., Angulo R. E., Zennaro M., 2021, arXiv e-prints, p. arXiv:2104.14568

Arnold C., Li B., Giblin B., Harnois-Déraps J., Cai Y.-C., 2022, MNRAS, 515, 4161

Ata M., et al., 2018, MNRAS, 473, 4773

Baldauf T., Seljak U. c. v., Desjacques V., McDonald P., 2012, Phys. Rev. D, 86, 083540

Baldauf T., Mirbabayi M., Simonović M., Zaldarriaga M., 2015, Phys. Rev. D, 92, 043514

Baumann D., Nicolis A., Senatore L., Zaldarriaga M., 2012, J. Cosmology Astropart. Phys., 2012, 051

Bautista J. E., et al., 2021, MNRAS, 500, 736

Begeman K., Belikov A., Boxhoorn D., Valentijn E., 2012, Experimental Astronomy, 35

Bel J., Pezzotta A., Carbone C., Sefusatti E., Guzzo L., 2019, Astron. & Astrophys., 622, A109

Bel J., Larena J., Maartens R., Marinoni C., Perenon L., 2022, J. Cosmology Astropart. Phys., 2022, 076

Birrer S., et al., 2020, Astron. & Astrophys., 643, A165

Blanton M. R., et al., 2017, A. J., 154, 28

Blas D., Garny M., Ivanov M. M., Sibiryakov S., 2016, J. Cosmology Astropart. Phys., 2016, 028

Bridle S., King L., 2007, New Journal of Physics, 9, 444

Brieden S., Gil-Marín H., Verde L., 2021, J. Cosmology Astropart. Phys., 2021, 054

Brieden S., Gil-Marín H., Verde L., 2022, J. Cosmology Astropart. Phys., 2022, 024

Campbell H., et al., 2013, The Astrophysical Journal, 763, 88

Carbone C., Petkova M., Dolag K., 2016, J. Cosmology Astropart. Phys., 2016, 034

Carrasco J. J. M., Hertzberg M. P., Senatore L., 2012, Journal of High Energy Physics, 2012, 82

Carrilho P., Moretti C., Pourtsidou A., 2023, J. Cosmology Astropart. Phys., 2023, 028

Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, MNRAS, 297, 692

Catelan P., Porciani C., Kamionkowski M., 2000, MNRAS, 318, L39

Chan K. C., Scoccimarro R., Sheth R. K., 2012, Phys. Rev. D, 85, 083509

Chen S.-F., Vlah Z., White M., 2021, arXiv e-prints, p. arXiv:2110.05530

Chevallier M., Polarski D., 2001, International Journal of Modern Physics D, 10, 213

Chudaykin A., Dolgikh K., Ivanov M. M., 2021, Phys. Rev. D, 103, 023507

Cole S., et al., 2005, MNRAS, 362, 505

Colgáin E. Ó., Sheikh-Jabbari M. M., 2021, Classical and Quantum Gravity, 38, 177001

Collaboration D., et al., 2021, Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing (`arXiv:2105.13549`)

Colless M., et al., 2001, MNRAS, 328, 1039

Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 063519

Crocce M., Scoccimarro R., Bernardeau F., 2012, MNRAS, 427, 2537

Cuesta-Lazaro C., et al., 2023, MNRAS, 523, 3219

DES Collaboration et al., 2022, arXiv e-prints, p. arXiv:2207.05766

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Dawson K. S., et al., 2013, A. J., 145, 10

Dawson K. S., et al., 2016, A. J., 151, 44

DeRose J., Chen S.-F., White M., Kokron N., 2022, J. Cosmology Astropart. Phys., 2022, 056

Desjacques V., 2008, Phys. Rev. D, 78, 103503

Desjacques V., Crocce M., Scoccimarro R., Sheth R. K., 2010, Phys. Rev. D, 82, 103529

Desjacques V., Jeong D., Schmidt F., 2018, J. Cosmology Astropart. Phys., 2018, 035

Di Valentino E., Melchiorri A., Silk J., 2020, Nature Astronomy, 4, 196

Di Valentino E., Melchiorri A., Silk J., 2021, The Astrophysical Journall, 908, L9

Dodelson S., Schmidt F., 2020, Modern Cosmology, doi:10.1016/C2017-0-01943-2.

Donald-McCann J., Koyama K., Beutler F., 2023, MNRAS, 518, 3106

Efstathiou G., et al., 2002, MNRAS, 330, L29

Eggemeier A., Scoccimarro R., Smith R. E., 2019, Phys. Rev. D, 99, 123514

Eggemeier A., Scoccimarro R., Crocce M., Pezzotta A., Sánchez A. G., 2020, Phys. Rev. D, 102, 103530

Eggemeier A., Scoccimarro R., Smith R. E., Crocce M., Pezzotta A., Sánchez A. G., 2021, Phys. Rev. D, 103, 123550

Eggemeier A., Camacho-Quevedo B., Pezzotta A., Crocce M., Scoccimarro R., Sánchez A. G., 2023, MNRAS, 519, 2962

Eisenstein D. J., Hu W., 1998, The Astrophysical Journal, 496, 605

Eisenstein D. J., et al., 2005, The Astrophysical Journal, 633, 560

Eisenstein D. J., et al., 2011, A. J., 142, 72

Erben T., et al., 2013, MNRAS, 433, 2545

Euclid Collaboration et al., 2019, MNRAS, 484, 5509

Euclid Collaboration et al., 2021, MNRAS, 505, 2840

Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, MNRAS, 467, 1627

Freedman W. L., 2021, The Astrophysical Journal, 919, 16

Fry J. N., 1996, The Astrophysical Journal, 461

Garny M., Taule P., 2021, J. Cosmology Astropart. Phys., 2021, 020

Gil-Marín H., et al., 2020, MNRAS, 498, 2492

Glanville A., Howlett C., Davis T. M., 2022, arXiv e-prints, p. arXiv:2205.05892

Gott J. Richard I., Jurić M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, The Astrophysical Journal, 624, 463

Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, MNRAS, 457, 1577

Grieb J. N., et al., 2017, MNRAS, 467, 2085

Gunn J. E., et al., 2006, A. J., 131, 2332

Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, Phys. Rev. D, 76, 083503

Hagstotz S., Reischke R., Lilow R., 2022, MNRAS, 511, 662

Handley W., 2021, Phys. Rev. D, 103, L041301

Handley W., Lemos P., 2019, Phys. Rev. D, 100, 043504

Hannestad S., 2005, , 95, 221301

Hartlap J., Simon P., Schneider P., 2007, Astron. & Astrophys., 464, 399

Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, , 646, L1

Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2014, The Astrophysical Journal, 780, 111

Heitmann K., et al., 2019, The Astrophysical Journal Supplement Series, 245, 16

Heymans C., et al., 2012, MNRAS, 427, 146

Heymans C., et al., 2021, Astronomy & Astrophysics, 646, A140

Hildebrandt, H. et al., 2020, A&A, 633, A69

Hildebrandt H., et al., 2016, Monthly Notices of the Royal Astronomical Society, 465, 1454–1498

Hirata C. M., Seljak U., 2004, Phys. Rev. D, 70, 063526

Hou J., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 2521–2534

Hou J., et al., 2021, MNRAS, 500, 1201

Huchra J., Davis M., Latham D., Tonry J., 1983, ApJS, 52, 89

Ivanov M. M., Sibiryakov S., 2018, J. Cosmology Astropart. Phys., 2018, 053

Ivanov M. M., Simonović M., Zaldarriaga M., 2020a, Phys. Rev. D, 101, 083504

Ivanov M. M., Simonović M., Zaldarriaga M., 2020b, J. Cosmology Astropart. Phys., 2020, 042

Ivezić Ž., et al., 2019, The Astrophysical Journal, 873, 111

Joudaki S., et al., 2016, Monthly Notices of the Royal Astronomical Society, 465, 2033–2052

Kaiser N., 1992, The Astrophysical Journal, 388, 272

Kaufman G. M., 1967, Astron. & Astrophys.

Kazin E. A., Sánchez A. G., Blanton M. R., 2012, MNRAS, 419, 3223

Kilbinger M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 472, 2126

Kitaura F.-S., et al., 2016, MNRAS, 456, 4156

Kitching T. D., et al., 2014, MNRAS, 442, 1326

Kokron N., DeRose J., Chen S.-F., White M., Wechsler R. H., 2021, MNRAS, 505, 1422

Kourkchi E., et al., 2020, The Astrophysical Journal, 902, 145

Kovac J. M., Leitch E. M., Pryke C., Carlstrom J. E., Halverson N. W., Holzapfel W. L., 2002, , 420, 772

Kuijken K., et al., 2015, MNRAS, 454, 3500

Kumar S., Nunes R. C., Yadav P., 2022, J. Cosmology Astropart. Phys., 2022, 060

Lahav O., et al., 2002, MNRAS, 333, 961

Laureijs R., et al., 2011, arXiv e-prints, p. arXiv:1110.3193

Lazeyras T., Schmidt F., 2019, J. Cosmology Astropart. Phys., 2019, 041

Lazeyras T., Wagner C., Baldauf T., Schmidt F., 2016, J. Cosmology Astropart. Phys., 2016, 018

Lemos P., et al., 2021, MNRAS, 505, 6179

Lewis A., Bridle S., 2002, Physical Review D, 66

Lewis A., Challinor A., Lasenby A., 2000, The Astrophysical Journal, 538, 473

Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118

Limber D. N., 1954, The Astrophysical Journal, 119, 655

Linder E. V., 2003, Phys. Rev. Lett., 90, 091301

Lippich M., et al., 2019, MNRAS, 482, 1786

LoVerde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506

Lyke B. W., et al., 2020, ApJS, 250, 8

MacCrann N., Zuntz J., Bridle S., Jain B., Becker M. R., 2015, MNRAS, 451, 2877

Mather J. C., et al., 1994, The Astrophysical Journal, 420, 439

Maus M., Chen S.-F., White M., 2023, J. Cosmology Astropart. Phys., 2023, 005

McDonald P., Roy A., 2009, J. Cosmology Astropart. Phys., 2009, 020

Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, MNRAS, 454, 1958

Miller L., et al., 2013, MNRAS, 429, 2858

Neveux R., et al., 2020, MNRAS, 499, 210

Nishimichi T., Bernardeau F., Taruya A., 2016, Physics Letters B, 762, 247

Nishimichi T., Bernardeau F., Taruya A., 2017, Phys. Rev. D, 96, 123515

Nishimichi T., et al., 2019, The Astrophysical Journal, 884, 29

Osato K., Nishimichi T., Bernardeau F., Taruya A., 2019, Phys. Rev. D, 99, 063530

Otten E. W., Weinheimer C., 2008, Reports on Progress in Physics, 71, 086201

Pellejero Ibañez M., Stücker J., Angulo R. E., Zennaro M., Contreras S., Aricò G., 2022, MNRAS, 514, 3993

Percival W. J., et al., 2014, MNRAS, 439, 2531

Perlmutter S., et al., 1999, The Astrophysical Journal, 517, 565

Planck Collaboration et al., 2016, Astron. & Astrophys., 594, A13

Planck Collaboration et al., 2020, Astron. & Astrophys., 641, A6

Pueblas S., Scoccimarro R., 2009, Phys. Rev. D, 80, 043504

Ramachandra N., Valogiannis G., Ishak M., Heitmann K., LSST Dark Energy Science Collaboration 2021, Phys. Rev. D, 103, 123525

Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning

Rassat A., et al., 2008, arXiv e-prints, p. arXiv:0810.0003

Raveri M., Hu W., 2019, Phys. Rev. D, 99, 043506

Reid M. J., Braatz J. A., Condon J. J., Lo K. Y., Kuo C. Y., Impellizzeri C. M. V., Henkel C., 2013, The Astrophysical Journal, 767, 154

Reid B., et al., 2016, MNRAS, 455, 1553

Riess A. G., et al., 1998, A. J., 116, 1009

Riess A. G., et al., 2021, arXiv e-prints, p. arXiv:2112.04510

Ross A. J., et al., 2020, MNRAS, 498, 2354

Sánchez A. G., 2020, Phys. Rev. D, 102, 123511

Sánchez A. G., Baugh C. M., Percival W. J., Peacock J. A., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 189

Sánchez A. G., et al., 2017, MNRAS, 464, 1640

Sanchez A. G., Ruiz A. N., Gonzalez Jara J., Padilla N. D., 2021, arXiv e-prints, p. arXiv:2108.12710

Sánchez A. G., Ruiz A. N., Jara J. G., Padilla N. D., 2022, , 514, 5673

Saunders W., et al., 1999, arXiv e-prints, pp astro–ph/9909191

Schneider A., Stoira N., Refregier A., Weiss A. J., Knabenhans M., Stadel J., Teyssier R., 2020, J. Cosmology Astropart. Phys., 2020, 019

Scoccimarro R., 2004, Phys. Rev. D, 70, 083007

Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, The Astrophysical Journal, 496, 586

Scoccimarro R., Couchman H. M. P., Frieman J. A., 1999, The Astrophysical Journal, 517, 531

Semenaite A., et al., 2022, MNRAS, 512, 5657

Semenaite A., et al., 2023, MNRAS, 521, 5013

Senatore L., Zaldarriaga M., 2014, arXiv e-prints, p. arXiv:1409.1225

Sheth R. K., Chan K. C., Scoccimarro R., 2013, Phys. Rev. D, 87, 083002

Simon T., Zhang P., Poulin V., Smith T. L., 2022, arXiv e-prints, p. arXiv:2208.05929

Slipher V. M., 1917, Proceedings of the American Philosophical Society, 56, 403

Smee S. A., et al., 2013, A. J., 146, 32

Smith A., et al., 2020, MNRAS, 499, 269

Spergel D. N., et al., 2003, The Astrophysical Journal Supplement Series, 148, 175

Tadros H., et al., 1999, MNRAS, 305, 527

Takahashi R., 2008, Progress of Theoretical Physics, 120, 549

Tamone A., et al., 2020, MNRAS, 499, 5527

Tanseri I., Hagstotz S., Vagnozzi S., Giusarma E., Freese K., 2022, arXiv e-prints, p. arXiv:2207.01913

Taruya A., 2016, Phys. Rev. D, 94, 023504

Taruya A., Nishimichi T., Saito S., 2010, Physical Review D, 82

Taruya A., Bernardeau F., Nishimichi T., Codis S., 2012, Phys. Rev. D, 86, 103528

Torrado J., Lewis A., 2021, J. Cosmology Astropart. Phys., 2021, 057

Tröster T., et al., 2021a, Astron. & Astrophys., 649, A88

Tröster T., et al., 2021b, Astron. & Astrophys., 649, A88

Tröster T., et al., 2020, Astronomy & Astrophysics, 633, L10

Vagnozzi S., Di Valentino E., Gariazzo S., Melchiorri A., Mena O., Silk J., 2021, Physics of the Dark Universe, 33, 100851

Vlah Z., Seljak U., Yat Chu M., Feng Y., 2016, J. Cosmology Astropart. Phys., 2016, 057

York D. G., et al., 2000, A. J., 120, 1579

Yuan S., Garrison L. H., Eisenstein D. J., Wechsler R. H., 2022, Mon. Not. Roy. Astron. Soc., 515, 871

Zennaro M., Angulo R. E., Pellejero-Ibáñez M., Stücker J., Contreras S., Aricò G., 2023, MNRAS, 524, 2407

Zhai Z., et al., 2019, The Astrophysical Journal, 874, 95

Zhai Z., et al., 2023, The Astrophysical Journal, 948, 99

Zhao C., et al., 2021, MNRAS, 503, 1149

Zheng Z., Coil A. L., Zehavi I., 2007, The Astrophysical Journal, 667, 760

d'Amico G., Gleyzes J., Kokron N., Markovic K., Senatore L., Zhang P., Beutler F., Gil-Marín H., 2020, J. Cosmology Astropart. Phys., 2020, 005

de Jong J. T. A., et al., 2017, Astron. & Astrophys., 604, A134

de Mattia A., et al., 2021, MNRAS, 501, 5616

di Porto C., Amendola L., Branchini E., 2012, MNRAS, 419, 985

Acknowledgments

First of all, I would like to thank my supervisor Ariel G. Sánchez. Thank you for your passion and your patience. I truly appreciate having always been able to count on your guidance and support and your unwavering determination to see all of the glasses (more than) halfway full.

Similarly, I would also like to thank my Master's supervisor Catherine Heymans. Without your encouragement and faith in me, this thesis would have never been started in the first place, thank you for cheering me on ever since with the enthusiasm I could feel through all the distance.

I am forever grateful to the Morning Coffee Crew: Shola Maria Linz-Wylie, Jonny Clarke and Claudia Pulsoni, for being the first people in Munich I could call friends and for all of the coffee (and tea) consumed together and through the video calls. Thank you for all the invaluable advice and sorry for never listening to it!

My most sincere thanks go to all of the wonderful people I was so lucky to share the office with. Thank you to Jiamin Hou for the fascinating discussions and the general wisdom, and the whiskey evenings. Thank you also to Martha Lippich - everything wonderful that I saw in Munich, I saw thanks to you. My biggest thank you hug goes to Matteo Esposito for all of the much-needed unbeatable hugs that I have been so grateful to receive. I couldn't thank Sara Maleubre Molinero enthusiastically enough for all of her endless energy and inspiration and encouragement, especially during the dark days spent together in the trenches of postdoc applications. Finally, thank you to Andrea Fiorilli who in such a short time became such a crucial part of the Dream Office and life in Munich in general. Thank you for all of the coffee and, with it, all of the company, compassion and care that you have shown me. An extra thanks also go to the honorary office-mate Fabian Balzer for all of his German expertise and werewolf sessions and to Brenda for literally always being there.

Of course, I can't continue without thanking the rest of the cosmolss group as well. I can't ever thank enough to Andrea Pezzotta (The Great), who more than earned his name with his endless patience answering my questions and proofreading whenever I asked for help (including parts of Chapter 5 in this thesis). Thank you also for always making me smile with your mere presence and the incredible acting talent. Thank you to Carlos Mauricio Correa for his inexhaustible supplies of encouragement and warmth and thanks to Danny Farrow for all of the 4pm tea which remained a habit long after he moved away but was never quite the same. Thank you to the honorary group member Alex Eggemeier for the wonderful week in Bonn and for always having all of the answers.

I would also like to thank all of the staff at MPE for having shown me much kindness and going above and beyond their job description. Thanks to Peter for all the jokes every time he found me staying too late in the office and most special thanks to Christa Ingram for being an actual magician and for all of her endless smiles brightening up every single day.

Thank you also to everyone I met outside of the Institute: first and foremost, thank you to Laura Herold for being the best Journal Club co-organiser I could have ever asked for (and for editing the German version of the abstract of this thesis). I am so grateful that we were brought together to create the most fantastic of JCs, even though the real JC was the friendship we created along the way. Thank you also to Marta Monelli and Vyoma Muralidhara in MPA whose company I have had the great pleasure to share. Thank you as well to the many people in USM who welcomed me, advised me and generously allowed me to steal their offices (almost) without questioning: Laurence Gong, Steffen Hagstotz, Anik Halder, Nico Hamaus, Giacomo Queirolo, Barbara Sartoris, Stella Seitz, Tamás Norbert Varga and Raphael Zöller.

I would like to thank Maria Paulus for all of her friendship and genuine care shown to me. Thank you also to Linda Baronchelli and Vlas Sokolov for the many dinners and for always showing up with such inspirational enthusiasm. A very special thanks go to Steffen Hagstotz (again) and Raffaella Capasso for all of our cinema outings and beyond.

Thank you to the friends far away. Thank you to Bea Moser who was there at the very beginning of the PhD journey and remained close throughout - thank you for always having a home for me in Zürich, friend. Equally, thank you to Dominik Zürcher, who is wonderful always, and Pascale Berner for all of the best of time spent in schools and conferences and for always being the most marvellous of hosts in Zürich. Thank you also to all of the friends at home, especially my second family, Taškento, and in Edinburgh, and wherever you might be - the list is far too long but I remember and I thank you all.

A separate thanks to the person who has suffered the longest of phone calls, whenever I lacked faith in myself and the world around me - thank you to Eliisabet Hein for never losing patience and for all of the crossword Sundays and beyond.

Galiausiai noriu padėkoti savo šeimai - tik jūsų besąlygiškos paramos dėka galiu nepasiklysti visose nesibaigiančiose kelionėse.

Everything good that I find in me and around me is thanks to all of the people mentioned (and thought about) here and I am extremely grateful for all of that. Mysteriously, even with the Universe expanding, my love for you all never gets diluted.