UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS FÍSICAS
FÍSICA DE ALTAS ENERGÍAS, FÍSICA NUCLEAR, GRAVITACIÓN Y FÍSICA MATEMÁTICA

REDSHIFT SPACE DISTORTIONS IN THE ERA OF STAGE IV DARK ENERGY EXPERIMENTS

MODALIDAD DE GRADUACIÓN

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS (FÍSICA)

PRESENTA:
SADI RAMÍREZ SOLANO

TUTORA O TUTORES PRINCIPALES
DRA. MARIANA VARGAS MAGAÑA, IF
MIEMBROS DEL COMITÉ TUTOR
DR. SÉBASTIEN MICKAËL MARC FROMENTEAU, ICF
DR. ALEJANDRO AVILÉS CERVANTES, ICF

CIUDAD DE MÉXICO, MÉXICO, MAYO DEL 2025

UNAM
Universidad Nacional
Autónoma de México

Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

PROTESTA UNIVERSITARIA DE INTEGRIDAD Y HONESTIDAD ACADÉMICA Y PROFESIONAL
(Graduación con trabajo escrito)

De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la Institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado:

Redshift Space Distortions in the era of stage IV dark energy experiments.

que presenté para obtener el grado de ----Doctorado---- es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi programa de posgrado, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de graduación.

Atentamente

Sadi Ramírez Solano 519020555

Sadi RS

(Nombre, firma y Número de cuenta de la persona alumna)

COORDINACIÓN GENERAL DE ESTUDIOS DE POSGRADO

CARTA AVAL PARA DAR INICIO A LOS TRÁMITES DE GRADUACIÓN

Universidad Nacional Autónoma de México
Secretaría General
Coordinación General de Estudios de Posgrado

Dr. Alberto Güijosa Hidalgo
Programa de Posgrado en Ciencias Físicas
Presente

Quien suscribe, **Dra. Mariana Vargas Magaña**, tutor(a) principal de **Sadi Ramírez Solano**, con número de cuenta **519020555**, integrante del alumnado de **Doctorado en Ciencias (Física)** de ese programa, manifiesto bajo protesta de decir verdad que conozco el trabajo escrito de graduación elaborado por dicha persona, cuyo título es:

Redshift Space Distortions in the Era of Stage IV Dark Energy Experiments

, así como el reporte que contiene el resultado emitido por la herramienta tecnológica de identificación de coincidencias y similitudes con la que se analizó ese trabajo, para la prevención de faltas de integridad académica.

De esta manera, con fundamento en lo previsto por los artículos 96, fracción III del Estatuto General de la UNAM; 21, primero y segundo párrafos, 32, 33 y 34 del Reglamento General de Exámenes y; 22, 49, primer párrafo y 52, fracción II del Reglamento General de Estudios de Posgrado, **AVALO** que el trabajo de graduación presentado se envíe al jurado para su revisión y emisión de votos, por considerar que cumple con las exigencias de rigurosidad académica previstas en la legislación universitaria.

Protesto lo necesario,

Ciudad Universitaria, Cd. Mx., a 7 de mayo de 2025

Dra. Mariana Vargas Magaña

Tutor(a) principal

Universidad Nacional Autónoma de México

Instituto de Física

Redshift Space Distortions in the era of stage IV dark energy experiments.

THESIS

that in order to obtain the title of:

Doctorate in Science (Physics)

presents:

Sadi Ramírez Solano

Advisor:
Dr. Mariana Vargas Magaña

Advisory Committee:
Dr. Sébastien Fromenteau
Dr. Alejandro Aviles Cervantes

Ciudad de México, México, 2025

Disclaimer

The work written here is based on research carried out by Sadi Ramírez Solano under the supervision of Dr. Mariana Vargas Magaña at the Instituto de Física of Universidad Nacional Autónoma de México.

I declare that this work has been composed solely by myself and, except where stated otherwise by reference or acknowledgment, the work presented is entirely my own. The figures and tables in this thesis were created by the author except where explicitly stated otherwise in the figures and tables' captions.

Acknowledgments

I would like to express my deepest gratitude to my family for their unwavering support throughout the years and for standing by every decision I have made. Their encouragement has been essential in allowing me to pursue a path that brings me joy and fulfillment. I am also grateful to those who guided me from a young age, helping me discover and follow a direction that has allowed me to grow both personally and professionally.

Special thanks go to my advisor and my advisory committee for their continuous guidance, support, and trust throughout this journey.

Resumen en español

La actual década de la cosmología está entrando en una nueva era de datos masivos, donde el estudio del agrupamiento de galaxias a través de la estructura a gran escala (LSS) juega un papel crucial, ya que puede proporcionarnos restricciones poderosas sobre parámetros cosmológicos nunca antes alcanzadas. Los altos volúmenes de datos nos permiten mejorar la confiabilidad, y por primera vez, podemos evaluar teorías cosmológicas como el modelo ΛCDM y sus variaciones con alta precisión.

Una de las principales características del modelo estándar de la cosmología es la expansión acelerada del universo, y los experimentos de levantamientos espectroscópicos pueden ayudarnos a comprender su naturaleza al explorar modelos de energía oscura y algunas teorías alternativas a la relatividad general.

En las últimas décadas, se han llevado a cabo experimentos espectroscópicos para analizar la estructura del universo a gran escala. Al mapear y crear un conjunto de datos de millones de galaxias, levantamientos como el Kilo-Degree Survey (KiDS) [1], el LOFAR Two-metre Sky Survey (LoTSS) [2], el 2dF Galaxy Redshift Survey (2dFGRS) [3, 4], el Baryon Oscillation Spectroscopic Survey (BOSS) [5, 6], el extended Baryon Oscillation Spectroscopic Survey (eBOSS) [7, 8] y más recientemente el Dark Energy Survey (DES) [9, 10], Euclid [11, 12] o el Rubin Observatory Large Synoptic Survey Telescope (LSST) [13], podemos investigar la naturaleza de la energía oscura.

En este trabajo, nos enfocamos en DESI, un experimento espectroscópico que fue instalado en el telescopio Mayall de 4 metros ubicado en el Observatorio Nacional de Kitt Peak, y que actualmente está mapeando el universo observando en una área de $14,000, \text{deg}^2$ alrededor de más de 40 millones de galaxias y cuásares en una ardua campaña de cinco años. El instrumento cuenta con un corrector de gran campo en el foco primario con un diámetro de $3.2°$, que dirige la luz hacia 5020 posicionadores de fibra robóticos situados sobre una superficie focal asférica de 0.812 metros [14].

El diseño de DESI fue creado para cumplir con los requisitos de un levantamiento de energía oscura de Etapa IV. El Dark Energy Task Force DETF [15] estableció la definición de Etapa IV para cuantificar las incertidumbres en los parámetros de la ecuación de estado de la energía oscura, w_0 y su evolución w_a. La Figura de Mérito del DETF se define como el inverso del área de la elipse de error en el plano w_0 vs w_a.

La encuesta principal de DESI comenzó el 14 de mayo de 2021, y desde entonces, ha observado de forma constante cientos de millas de galaxias y cuásares por noche, con interrupciones técnicas mínimas. En el primer año, DESI registró más de 15 millones de corrimientos al rojo extragalácticos, convirtiéndose en el levantamiento espectroscópico más grande jamás realizado en la historia de la cosmología. Los datos recolectados por DESI mejorarán las mediciones de la aceleración cósmica, permitirán sondas más sensibles a la evolución temporal de la energía oscura, ajustarán las restricciones sobre modelos de gravedad modificada y proporcionarán límites superiores significativamente mejorados sobre la masa total de los neutrinos. Para el final de DESI, se espera medir la escala de distancia cósmica con una precisión agregada de al menos 0.22% para $0 < z < 1.1$, 0.31% para $1.1 < z < 1.9$, y el parámetro de Hubble con una precisión del 0.84% para $1.9 < z < 3.7$. DESI podría distinguir potencialmente entre las jerarquías de masa normal e invertida de los neutrinos, proporcionar descubrimientos fundamentales sobre la naturaleza de la gravedad mediante mediciones anisotrópicas del agrupamiento, y explorar la inflación a través de la amplitud de la no-Gaussianidad primordial en el espectro de potencia de materia.

La encuesta DESI utilizará el agrupamiento de galaxias para extraer información cosmológica a partir de dos observables clave: las oscilaciones acústicas de bariones (BAO) y las distorsiones en el espacio de corrimiento al rojo (RSD). Las BAO implican fluctuaciones en la densidad de materia bariónica generadas por ondas de densidad acústica en el plasma primordial del universo temprano, cuya firma quedó impresa en tiempos posteriores dentro de la función de correlación a escalas comóviles de $\approx 100\ \mathtt{h^{-1}Mpc}$. Por otro lado, las RSD se refieren a distorsiones en el agrupamiento de galaxias causadas por velocidades peculiares, originadas por la evolución gravitacional, y cuyos efectos introducen dificultades al interpretar la posición real de las galaxias como un mapa tridimensional; sin embargo, nos permiten extraer información valiosa mediante teorías de perturbaciones.

Un aspecto clave de la próxima era de la cosmología de precisión es que los nuevos datos se enfocarán principalmente en la formación de estructura en el régimen no lineal. Esto abrirá una ventana para poner a prueba teorías a nivel de sub-porcentaje, donde probar modelos más complejos que pueden introducir parámetros adicionales, así como algunos modelos de gravedad, puede brindarnos ideas sobre la naturaleza de la energía oscura.

Este trabajo se centra en cómo extraemos la información, partiendo de las bases de la cosmología, donde los modelos teóricos se basan en la teoría de perturbaciones, lo cual nos permite describir y comparar las propiedades estadísticas observadas en los levantamientos espectroscópicos. Primero, introducimos los fundamentos de la Teoría de Perturbaciones Euleriana y la Teoría de Perturbaciones Lagrangiana. Más adelante, exploramos cómo diferentes códigos, construidos a partir de estos dos enfoques, pueden extenderse a órdenes perturbativos superiores e incluir efectos físicos adicionales. Entre estos códigos, nos enfocamos particularmente en la Teoría de Perturbaciones Lagrangiana con Convolución combinada con el Modelo de Transmisión Gaussiano. Luego describimos cómo utilizar estos códigos para extraer información mediante diferentes metodologías: ajuste estándar, ajuste de forma y modelado completo. En este trabajo extraemos información por ajuste de forma en espacio de configuración por primera vez. Para lograrlo, incorporamos un muestreador y entrenamos una red neuronal para acelerar el proceso. Finalmente, evaluamos nuestros modelos usando simulaciones y los aplicamos a datos reales. Adicionalmente, probamos las diferentes metodologías para la comparación de modelos y métodos de DESI 2024.

Contents

Chapter 1

Introduction

The current decade of cosmology is undergoing a new era of massive data, where the study of galaxy clustering through large-scale structure (LSS) plays a crucial role, as it can provide us powerful constraints on cosmological parameters never before achieved. High volumes enable us to enhance reliability, and for the first time, we can evaluate cosmology theories such as the ΛCDM model and its variations with high precision.

One of the main characteristics of the standard model of cosmology is the accelerating expansion of the universe, and spectroscopy survey experiments can help us provide insights about its nature by exploring dark energy models and some alternative theories to general relativity.

In recent decades, spectroscopy experiments have been carried out to analyze the universe's structure at large scales. By mapping and creating a set of data of millions of galaxies, surveys like, the Kilo-Degree Survey (KiDS) [1], LOFAR Two-metre Sky Survey (LoTSS) [2], the 2dF Galaxy Redshift Survey (2dFGRS) [3, 4], the Baryon Oscillation Spectroscopic Survey (BOSS) [5, 6], the extended Baryon Oscillation Spectroscopic Survey (eBOSS) [7, 8] and more recently, the Dark Energy Survey (DES) [9, 10], Euclid [11, 12] or The Rubin Observatory Large Synoptic Survey Telescope (LSST) [13], we can investigate the nature of dark energy.

In this work, we focus on DESI, a spectroscopic experiment that was installed on the 4-meter Mayall Telescope located at the Kitt Peak National Observatory, and is currently mapping the universe by observing in a $14,000\,\mathrm{deg}^2$ footprint around more than 40 million galaxies and quasars in an arduous five-year survey. The instrument features a wide-field prime-focus corrector with a 3.2° diameter, which directs light onto 5020 robotic fiber positioners situated on a 0.812 meters aspheric focal surface [14].

DESI design was created to fulfill the requirements of a Stage IV dark energy survey. The Dark Energy Task Force DETF [15] established the Stage IV definition to quantify uncertainties in the dark energy equation of state parameters, w_0 and its evolution w_a. The DETF Figure of Merit is defined as the inverse of the area of the error ellipse in the w_0 vs w_a plane.

The main DESI survey started on May 14th, 2021, and since then, it has consistently observed hundreds of miles of galaxies and quasars per night, with minimal technical interruptions. In the first year, DESI recorded over 15 million extragalactic redshifts, making it the largest spectroscopic survey ever conducted in cosmology history. The data collected by DESI will enhance measurements of cosmic acceleration, enable more sensitive probes into the time evolution of dark energy, tighten constraints on modified gravity models, and provide significantly improved upper limits on the total neutrino mass. By the end of DESI, we expect to measure the cosmic distance scale with at least 0.22% aggregate precision from $0 < z < 1.1$, 0.31% from $1.1 < z < 1.9$, and the Hubble parameter with 0.84% precision from $1.9 < z < 3.7$. DESI could potentially distinguish between the normal and inverted neutrino mass hierarchies, provide groundbreaking insights into the nature of gravity through anisotropic clustering measurements, and explore inflation through the amplitude of primordial non-Gaussianity in the matter power spectrum.

The DESI survey will use Galaxy clustering to extracts cosmological information from two key observables, baryonic acoustic oscillations (BAO), and Redshift Space Distortions (RSD). BAO involves fluctuations in the baryonic matter density generated by acoustic density waves in the primordial plasma of the early universe, and whose signature was imprinted at later times within the correlation function at comoving scales of $\approx 100\ \mathtt{h^{-1}Mpc}$. while RSD refers to distortions in the clustering of galaxies caused by peculiar velocities, those originated by gravitational evolution, and whose effects introduce difficulties in interpreting the real position of the galaxies as a 3-D map, it allows us to extract valuable information through perturbation theories.

One key aspect of the next era of precision cosmology is that the new data will mostly focus on structure formation in the nonlinear regime. It will open a window to test theories at sub percent, where testing more complex models that can introduce extra parameters, and some gravity models can give us insights about the nature of the dark energy.

This work focuses on how we extract information, starting from the basis of cosmology, where the theoretical models are based on perturbation theory, which allows us to describe and compare the statistical properties observed in spectroscopic surveys. We first introduce the basics of Eulerian Perturbation Theory and Lagrangian Perturbation Theory. Later, we explore how different codes, built on these two approaches, can extend to higher perturbative orders and include additional physical effects. Among these codes, we focus particularly on Convolution Lagrangian Perturbation Theory combined with the Gaussian Streaming Model. We then describe how to use these codes to extract information through different methodologies: standard fitting, shape fitting, and full modeling. In this work we extract shape-fit information in configuration space for the first time. To achieve this, we incorporate a sampler and train a neural network to accelerate the process. Finally, we evaluate our models using simulations and apply them to real data. Additionally, we test the different methodologies for the DESI 2024 comparison of models and methods.

1.1 The Standard Model of Cosmology

In the 1940s, George Gamow and his collaborators, Ralph Alpher and Robert Herman, predicted that the universe began in a very dense, extremely hot state, an idea later called the “Big Bang.” Their work was the foundation for what would become the Standard Model of Cosmology. They proposed that, as the universe expanded, it was cooling, allowing for the formation of atomic nuclei in a process known as Big Bang Nucleosynthesis (BBN). A key prediction of their model was the existence of radiation from this hot early state, which was later confirmed in 1965 with the discovery of the Cosmic Microwave Background (CMB) by Arno Penzias and Robert Wilson. Over time, the Standard Model of Cosmology, also known as the Lambda Cold Dark Matter (ΛCDM) model, emerged as the main framework, incorporating dark matter and dark energy to explain the large-scale structure and accelerated expansion of the universe. Today, cosmology is experiencing a revolutionary period, mainly due to recent high-precision observations that have dramatically reduced theoretical speculation and opened new perspectives on the universe. New measurements from cosmic surveys of stage IV are refining our understanding of fundamental parameters, testing the limits of ΛCDM, and looking for possible new physics beyond the standard model. These advances are transforming our ability to probe the early universe, the nature of dark energy, and the fundamental laws governing cosmic evolution.

Latter, in 1992, the COBE satellite experiment, led by G. Smoot and J. Mather, confirmed that the universe is in thermal equilibrium and practically homogeneous and isotropic with tiny temperature anisotropies ($\Delta T \simeq 10^{-5} K$), which were the seeds for cosmic structure formation. These fluctuations resulted compatible with the ones predicted by inflation models and reinforced the cosmological principle, stating that the universe is homogeneous and isotropic, i.e., no matter which direction you look, the universe exhibits the same properties, and that any given point in the universe appears identical in its properties to any other point.

Einstein's theory of general relativity provided, for the first time in history, a rigorous and testable framework for understanding the universe. The Standard Model of Cosmology, the ΛCDM model, assumes the cosmological principle that states that on large scales, the universe is homogeneous, isotropic, in addition to spatially flat. Within this framework, the total energy density is predominantly composed of two unseen components: a cosmological constant (Λ), associated with dark energy and accounting for approximately 70% of the universe's energy content today, and cold non-relativistic matter, which constitutes the remaining portion. This includes contributions from baryonic matter, non-massive neutrinos, and photons. The ΛCDM model is formulated within the context of a flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which serves as the foundation for modern cosmological descriptions of the universe's large-scale structure and evolution.

$$ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \tag{1.1}$$

where $a(t)$ is the scale factor, which describes how the distances in the universe change over time, and the parameter k represents the spatial curvature of the universe. To ensure consistency with the assumptions of homogeneity and isotropy, the energy-momentum tensor is modeled as that of a perfect fluid,

$$T_{\mu\nu} = \rho u_\mu u_\nu + P\left(g_{\mu\nu} + u_\mu u_\nu\right). \tag{1.2}$$

Starting from the metric in Equation 1.12, we can straightforwardly derive the following key equations:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \tag{1.3}$$

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{4\pi G}{3}(\rho + 3p), \tag{1.4}$$

with H being the Hubble parameter, and dots being derivatives with respect to cosmic time. These equations, known as the Friedman equations, describe the expansion dynamics of the universe, incorporating the effects of matter, curvature, the cosmological constant, and also radiation or any other energetic components. The last one is due to the principle of generalized covariance within the Einstein equations.

In addition to the expansion equations, we can derive an extra equation that governs the evolution of the energy density, given by

$$\dot{\rho} + 3H(\rho + P) = 0. \tag{1.5}$$

This equation arises from the conservation of the energy-momentum tensor, expressed as $\nabla_\mu T^{\mu\nu} = 0$. It describes how the energy density ρ evolves over time as the universe expands, accounting for the pressure P of the cosmic fluid.

To solve the system of equations 1.3, 1.4, and 1.5, which involve three unknown variables (a, ρ, and p), it is necessary to introduce an additional assumption, a barotropic equation of state to characterize different cosmic fluids. This assumption allows us to express the pressure-to-density ratio, $w_X = \frac{P_X}{\rho_X}$, for various components of the universe, included $\rho_\Lambda = \frac{\Lambda c^2}{8\pi G}$ as follows:

$$w_X = \begin{cases} \frac{1}{3} & \text{for} \quad X = \text{radiation} \\ 0 & \text{for} \quad X = \text{dust} \;\; \text{(Non relativistic matter)} \\ -1 & \text{for} \quad X = \text{cosmological constant} \end{cases} \tag{1.6}$$

The system can be used to obtain

$$\Omega \equiv \sum_X \Omega_X = 1 + \frac{K}{a^2 H^2}. \tag{1.7}$$

Here we define the density parameter as $\Omega_X \equiv \frac{8\pi G}{3H^2}\rho_X$. The last terms in equation 1.7 is defined as $\Omega_K = -K/(a^2 H^2)$.

From the previous result, we have the following scenarios for the value of the curvature,

$$K = \begin{cases} > 0 & \text{for} \quad \Omega > 1 \quad \text{(closed universe, positive curvature)} \\ < 0 & \text{for} \quad \Omega < 1 \quad \text{(open universe, negative curvature)} \\ = 0 & \text{for} \quad \Omega = 1 \quad \text{(Euclidean universe, zero curvature)} \end{cases} \tag{1.8}$$

When we set $K = 0$, corresponding to a spatially flat universe, and the total energy density equals the critical density, defined as $\rho_c = \frac{3H^2}{8\pi G}$, the Friedmann equation reduces to a constraint on the density parameters, $\sum_i \Omega_i = 1$. Although this scenario might appear improbable, observational data tend to indicate that the universe is Euclidean.

Equation 1.7 is often expressed in terms of redshift, z, which accounts for the expansion of the universe by relating the observed wavelength of light to the scale factor:

$$z = (\lambda_0 - \lambda)/\lambda = a_0/a - 1 \tag{1.9}$$

This leads to an evolution equation for the Hubble parameter,

$$H(z) = H_0 \sqrt{\Omega_{\Lambda 0} + \Omega_{K0}(1+z)^2 + \Omega_{M0}(1+z)^3 + \Omega_{R0}(1+z)^4}, \tag{1.10}$$

with $\Omega_{\Lambda 0}$, Ω_{K0}, Ω_{M0} and Ω_{R0}, the present-day cosmological components. This equation describes how the Hubble parameter H evolves with redshift and how each component influences its evolution.

One of the most relevant quantities that we use to compare models with observations is the *comoving distance*, which quantifies the proper distance between two points in the universe, measured along a path defined at constant cosmological time. In a FLRW universe, the comoving distance to an object at redshift z is given by:

$$\chi(z) = c \int_0^z \frac{dz'}{H(z')}, \tag{1.11}$$

where the function $H(z)$, determined by the matter and energy content of the universe, governs the redshift dependence of $\chi(z)$.

1.2 Dynamics of Structure Formation of the Universe

The evolution of our universe from its nearly uniform state after the Big Bang to the complex cosmic web we observe today represents one of the most difficult questions in cosmology. This process, driven primarily by gravitational forces acting on initial density fluctuations, has created the formation of structures, from individual galaxies to massive galaxy clusters and vast cosmic filaments. In cosmology, the formation of large-scale structures can be studied by treating dark matter and baryons as a unified dark matter fluid. This approach is justified because, on large scales, both densities and velocities are small, and baryons behave as a non-relativistic fluid. As a result, baryons and dark matter can be treated the same way in their contribution to the gravitational potential. During the late-time large-scale structure formation, particularly in the matter-dominated era, this fluid is considered non-relativistic under the weak-field approximation. Therefore, in this approach, We can work with the scalar perturbations by employing the FLRW metric in the conformal Newtonian gauge,

$$ds^2 = -(1+2\Phi)dt^2 + a^2(t)(1-2\Psi)d\mathbf{x}^2, \quad \Phi = \Psi, \tag{1.12}$$

where potentials Φ and Ψ represent scalar perturbations to the metric: Φ modifies the time-time component, while Ψ modifies the spatial part, related to spatial curvature perturbations. The equality $\Phi = \Psi$ holds in the absence of anisotropic stress, which is valid when matter is composed of perfect fluids or non-relativistic matter (like cold dark matter).

In a collisionless framework, the gravitational dynamics of the dark matter fluid are dictated by the Vlasov–Poisson equations. These equations provide a comprehensive description of the evolution of the dark matter distribution function, denoted as $f(\mathbf{x}, \mathbf{p})$,

$$\frac{\mathrm{d}f(\mathbf{x}, \mathbf{p})}{\mathrm{d}\tau} = \frac{\partial f}{\partial \tau} + \frac{\partial f}{\partial x_i}\frac{dx_i}{d\tau} + \frac{\partial f}{\partial p_i}\frac{dp_i}{d\tau} = 0, \tag{1.13}$$

where the conformal time $\tau = dt/a$ follows the definition $ds^2 = a^2\left(-d\tau^2 + d\mathbf{x}^2\right)$. These relations help us to model how countless particles of dark matter behave under the influence of gravity, which is related to the matter distribution in the universe through the following integral,

$$ma^{-3}\int d^3\mathbf{p} f(\mathbf{x}, \mathbf{p}) \equiv \bar{\rho}(t)\left(1 + \delta_m(\mathbf{x}, \tau)\right), \tag{1.14}$$

where, the density contrast δ is defined in terms of the total matter density ρ as

$$\rho(\mathbf{x}, t) = \bar{\rho}(t)[1 + \delta(\mathbf{x}, t)]. \tag{1.15}$$

On the other hand, the gravitational potential Φ is determined by Poisson's equation given by $\nabla_{\mathrm{x}}^2\Phi = 4\pi G a^2 \bar{\rho}\delta_m$. In this context, the canonical momentum is given by $p_i = ma\mathbf{v}$, where $\mathbf{v}$ represents the physical peculiar velocity, defined as $\mathbf{v} = a\, d\mathbf{x}/dt = d\mathbf{x}/d\tau$. So, the dynamics of particles are characterized by the equations of motion $\frac{d\mathbf{x}}{d\tau} = \frac{\mathbf{p}}{ma}$ and $\frac{d\mathbf{p}}{d\tau} = -ma\nabla_{\mathbf{x}}\Phi$. These equations capture the dynamics of particles under the influence of the gravitational potential, which models the evolution of structures in the universe.

The Vlasov and Poisson equations provide a comprehensive framework for describing structure formation in a universe dominated by dark matter. However, solving these equations analytically is practically impossible. As a result, it is common to work instead with the velocity moments of the distribution function, which include the mean number density n, the mean velocity $\mathbf{v}$, and the velocity dispersion σ_{ij}. These moments are defined as follows:

$$
\begin{aligned}
n(\mathbf{x}) &= \int d^3\mathbf{p} f(\mathbf{x}, \mathbf{p}) \equiv m a^{-3} \rho(\mathbf{x}), \\
n(\mathbf{x})\mathbf{v}(\mathbf{x}) &= \int d^3\mathbf{p} \frac{\mathbf{p}}{ma} f(\mathbf{x}, \mathbf{p}), \\
n(\mathbf{x}) \left(v_i v_j + \sigma_{ij}(\mathbf{x})\right) &= \int d^3\mathbf{p} \frac{p_i p_j}{m^2 a^2} f(\mathbf{x}, \mathbf{p}).
\end{aligned}
\tag{1.16}
$$

Alternatively, from these equations 1.12, we can derive the continuity and Euler equations of fluid mechanics,

$$
\begin{aligned}
\frac{\partial n}{\partial \tau} + \nabla_{\mathbf{x}} \cdot (n\mathbf{v}) &= 0 \\
\frac{dv_i}{d\tau} + \mathcal{H} v_i + v_j \nabla_{\mathbf{x},j} v_i &= -\nabla_{\mathbf{x},i} \Phi - \frac{1}{n} \nabla_{\mathbf{x},j} \left(n \sigma_{ij}\right).
\end{aligned}
\tag{1.17}
$$

If we work in the framework of single-stream limit,

$$
f(\mathbf{x}, \mathbf{p}) = \frac{\rho(\mathbf{x})}{m a^{-3}} \delta_D(\mathbf{p} - m a v(\mathbf{x})),
\tag{1.18}
$$

with δ_D the Dirac-delta function. We can obtain the equations governing the perturbation theory. In this limit, the initial conditions at early times follow linear dynamics, and particle velocities at a single point $\mathbf{x}$ are coherent and the velocity dispersion vanishes, so we have, swapping in δ for n,

$$
\begin{aligned}
&\frac{\partial \delta}{\partial \tau} + \nabla_{\mathbf{x}} \cdot ((1+\delta)\mathbf{v}) = 0, \\
&\frac{dv_i}{d\tau} + \mathcal{H}_i + v_j \nabla_{\mathbf{x},j} v_i = -\nabla_{\mathbf{x},i} \Phi,
\end{aligned}
\tag{1.19}
$$

where we have defined $\mathcal{H} = Ha$. These equations form the foundation of Eulerian perturbation theory (EPT), which is also commonly referred to as standard perturbation theory (SPT).

1.3 Linear theory

If we analyze the evolution of small perturbations, at linear order, and we only keep first-order deviations in density contrast δ, velocity field $\mathbf{v}$, and gravitational potential Φ, neglecting higher-order interactions, we can derive a set of linear equations governing the dynamics of structure,

$$\begin{aligned} &\frac{\partial \delta}{\partial \tau}+\theta=0, \\ &\frac{d v_i}{d \tau}+\mathcal{H} v_i=-\nabla_{\mathbf{x}, i} \Phi, \quad \nabla_{\mathbf{x}}^2 \Phi=\frac{3}{2} \mathcal{H}^2 \Omega_m(\tau) \delta, \end{aligned} \tag{1.20}$$

with $\Omega_m(\tau)=\rho_m(\tau) / \rho_{\text{crit}}(\tau)$ the matter density fraction and $\theta=\nabla_{\mathbf{x}} \cdot \mathbf{v}$ the velocity. In an expanding universe, the vorticity $\mathbf{w}=\nabla_{\mathbf{x}} \times \mathbf{v}$ decays over time. To isolate the longitudinal component, we take the divergence of the Euler equation, which removes the transverse contribution. Combining this with the continuity equation, which relates velocity divergence to density fluctuations, we obtain,

$$\frac{d^2 \delta}{d \tau^2}+\mathcal{H} \frac{d \delta}{d \tau}=\frac{3}{2} \mathcal{H}^2 \Omega_m(\tau) \delta. \tag{1.21}$$

This second-order differential equation admits both a growing and a decaying solutions, expressed as $\delta_{\text{lin}}(\tau)=D(\tau) \delta_{+}+D_{-}(\tau) \delta_{-}$. The growing mode is given by

$$D(\tau)=H(a) \int_0^a \frac{d a^{\prime}}{a^3 H\left(a^{\prime}\right)}, \tag{1.22}$$

while the decaying mode follows $D_{-}(\tau)=H(a)$, which we can assume negligible at the relevant time. A common cosmological convention normalizes the growth factor to unity today, giving the velocity divergence:

$$\theta_{\text{lin}}(\tau)=-f(a) \mathcal{H} \delta_{\text{lin}}(\tau), \tag{1.23}$$

with the linear growth rate $f(a)=d \ln D / d \ln a$. During matter domination, the growth factor scales as $D(a)=a$, and the growth rate remains constant $f(a)=1$.

Alternatively, we can work in the Lagrangian approach, where we label each element of fluid by its Lagrangian position $\mathbf{x}$, which is determined at subsequent times by the evolution of the Lagrangian displacement field Ψ and its eulerian position $\mathbf{q}$,

$$\mathbf{x}(\mathbf{q},\tau) = \mathbf{q} + \Psi(\mathbf{q},\tau). \tag{1.24}$$

By taking the second derivative, we can obtain an alternative form of the equation of motion, expressed as $\frac{d^2\mathbf{x}}{d\tau^2} = \frac{d^2\Psi(\mathbf{q},\tau)}{d\tau^2}$. This formulation leads to the equation $\frac{d^2\Psi}{d\tau^2} + \mathcal{H}\frac{d\Psi}{d\tau} = -\nabla_{\mathbf{x}}\Phi$.

And, from the conservation of the number of particles, we know that

$$1 + \delta(\mathbf{x},\tau) = \int d^3\mathbf{q}\delta_D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q},\tau)) = \frac{1}{|\det\left[\delta_{ij} + \Psi_{i,j}\right]|}, \tag{1.25}$$

furthermore, at leading order in Lagrangian space, the gradients in Lagrangian and Eulerian coordinates are nearly the same, and the density contrast is simply given by $\delta(\mathbf{x},\tau) = -\nabla\Psi(\mathbf{q},\tau)$. And, as any rotational (curl) component of $\Psi(\mathbf{q},\tau)$ decays with time, we can focus only on its divergence.

This allows us to write $\Psi(\mathbf{q},\tau)$ in terms of $\delta(\mathbf{q},\tau)$, leading to the relation

$$\Psi(\mathbf{q},\tau) = -\nabla^{-1}\delta(\mathbf{x},\tau). \tag{1.26}$$

In Fourier space, this becomes

$$\Psi(\mathbf{k},\tau) = \frac{i\mathbf{k}}{k^2}D(\tau)\delta_0(\mathbf{k}), \tag{1.27}$$

where $\delta_0(\mathbf{k})$ is the initial density contrast, and $D(\tau)$ describes how structures grow over time. This result is known as the Zeldovich approximation.

1.4 Clustering statistics

From Inflation, we know that the early universe was characterized by quantum fluctuations that seeded a Gaussian random field, providing the foundation for the large-scale structure we observe today. To study the distribution of galaxies, we use the two-point correlation function,

$$\xi(\mathbf{r}) = \langle \delta\left(\mathbf{r}_1\right) \cdot \delta\left(\mathbf{r}_2\right) \rangle_{\mathbf{r}=\mathbf{r_2}-\mathbf{r_1}} \quad , \tag{1.28}$$

which measures the probability of finding galaxy pairs separated by a distance $\mathbf{r}$. On large, quasilinear scales, structure formation can be effectively analyzed because deviations from Gaussianity are minimal, and the cosmological signal is predominantly captured in the correlation function. The dependence of ξ on the separation $\mathbf{r}$ arises from translation invariance, and in the absence of observational effects, cosmological correlations also exhibit rotational invariance, leading to $\xi(\mathbf{r}) = \xi(r)$.

However, this invariance is disrupted by redshift-space distortions, where line-of-sight velocities break isotropy. Alternatively, the power spectrum, defined as the Fourier transform of the correlation function, is employed:

$$\xi(\mathbf{r}) = \int d^3\mathbf{k}\, P(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{1.29}$$

In the absence of observational effects, the power spectrum can be expressed as $\langle \delta\left(\mathbf{k}_1\right) \delta\left(\mathbf{k}_2\right) \rangle = P\left(\mathbf{k}_1\right)(2\pi)^3 \delta_D\left(\mathbf{k}_1 + \mathbf{k}_2\right)$, where translational invariance is represented by the Dirac delta function, and rotational invariance implies that $P(\mathbf{k}) = P(k)$.

1.5 One loop power spectrum

According to Wick's theorem, after lineal order, the next leading order contribution to the power spectrum, arises from interactions between two pairs of linear fields $\delta_{\text{lin}}(\mathbf{p})$. Specifically, there are two such contributions: $2\left\langle \delta^{(1)}\delta^{(3)} \right\rangle$ and $\left\langle \delta^{(2)}\delta^{(2)} \right\rangle$. Consequently, the matter power spectrum up to 1-loop order in Eulerian Perturbation Theory (EPT) is expressed as:

$$P(k) = P_{\text{lin}}(k) + 2P^{(13)}(k) + P^{(22)}(k) \tag{1.30}$$

where the terms are defined as follows:

$$
\begin{aligned}
P^{(13)}(k) &= 3P_{\text{lin}} \int \frac{d^3\mathbf{p}}{(2\pi)^3} F_3(\mathbf{p}, -\mathbf{p}, \mathbf{k}) P_{\text{lin}}(\mathbf{p}), \\
P^{(22)}(k) &= 2 \int \frac{d^3\mathbf{p}}{(2\pi)^3} F_2(\mathbf{p}, \mathbf{k}-\mathbf{p})^2 P_{\text{lin}}(\mathbf{p}) P_{\text{lin}}(\mathbf{k}-\mathbf{p}).
\end{aligned}
\tag{1.31}
$$

Here, $P_{\text{lin}}(k)$ represents the linear power spectrum, while $P^{(13)}(k)$ and $P^{(22)}(k)$ account for the non-linear corrections at 1-loop order, incorporating the effects of higher-order interactions through the kernels F_3 and F_2.

In Lagrangian perturbation theory, the equivalent expression arises from the conservation of number, leading to the formulation:

$$
\delta(\mathbf{k}) = \int d^3\mathbf{q}\, e^{i\mathbf{k}\cdot\mathbf{q}} \left(e^{i\mathbf{k}\cdot\Psi(\mathbf{q})} - 1 \right). \tag{1.32}
$$

This expression implies that the power spectrum can be represented as:

$$
P(k) = \int d^3\mathbf{q}\, e^{i\mathbf{k}\cdot\mathbf{q}} \left(\left\langle e^{i\mathbf{k}\cdot\Delta} \right\rangle_{\mathbf{q}=\mathbf{q}_1-\mathbf{q}_2} - 1 \right), \tag{1.33}
$$

where $\Delta = \Psi(\mathbf{q}_1) - \Psi(\mathbf{q}_2)$.

By applying the cumulant theorem, we can expand the expression for the power spectrum,

$$
\left\langle e^{i\mathbf{k}\cdot\Delta} \right\rangle = \exp\left[\ \tfrac{1}{2} k_i k_j A_{ij}(\mathrm{q}) \quad \tfrac{i}{6} k_i k_j k_k W_{ijk}(\mathrm{q}) + \dots \ \right] \tag{1.34}
$$

where we have defined $\mathbf{A} = \langle \Delta\Delta \rangle$ and $\mathbf{W} = \langle \Delta\Delta\Delta \rangle$. These moments of the pairwise displacement Δ receive contributions at one-loop order,

$$
\begin{aligned}
&A_{ij}(\mathrm{q}) = A_{ij}^{\text{lin}}(\mathrm{q}) + A_{ij}^{1-100p}(\mathrm{q}), A_{ij}^{1-1\infty0p}(\mathrm{q}) = 2A_{ij}^{(13)}(\mathrm{q}) + A_{ij}^{(22)}(\mathrm{q}), \\
&W_{ijk}(\mathrm{q}) = W_{ijk}^{(112)} + W_{ijk}^{(121)} + W_{ijk}^{(211)},
\end{aligned}
\tag{1.35}
$$

and up to this order, the power spectrum is given by

$$
P(k) = \int d^3\mathrm{q} e^{i\mathbf{k}\cdot\mathrm{q} - \frac{1}{2}k_i k_j A_{ij}^{\text{Lio}}(\mathrm{q})} \left(\ 1 - \tfrac{1}{2} k_i k_j A_{ij}^{1-\text{lop}} \ (\mathrm{q}) - \frac{i}{6} k_i k_j k_k W_{ijk}(\mathrm{q}) \right). \tag{1.36}
$$

1.6 Higher order Perturbation Theory

The two conventional frameworks for perturbatively modeling cosmological structure formation are Eulerian and Lagrangian perturbation theory.

Lagrangian perturbation theory models cosmological structure formation by tracking the displacement field $\Psi(\mathbf{q}, t)$ of fluid elements from their initial Lagrangian coordinates $\mathbf{q}$ to their final Eulerian positions $\mathbf{x}(\mathbf{q}, t)$, where the relationship is expressed as $\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \Psi(\mathbf{q}, t)$. These fluid elements cluster under the influence of gravity and their displacements obey the equation of motion Ψ. In what follows, derivatives are with respect to conformal time τ and we define $\mathcal{H} = aH$ as the conformal Hubble parameter, while Φ is the gravitational potential, which we solve for order-by-order in terms of the linear density contrast δ_{lin} as

$$\Psi = \Psi^{(1)} + \Psi^{(2)} + \cdots \tag{1.37}$$

with

$$\Psi_i^{(n)}(\mathbf{q}) = \frac{i}{n!} \int_{\text{k},\text{P}_1 \ldots \text{P}_n} e^{i\text{k}\cdot\mathbf{q}} \delta^D_{\text{k}-\text{p}} L_i^{(n)} \left(\mathbf{p}_1, \ldots, \mathbf{p}_n\right) \delta_{\text{lin}} \left(\mathbf{p}_1\right) \ldots \delta_{\text{lin}} \left(\mathbf{p}_n\right), \tag{1.38}$$

were we use the shorthands $\text{p} - \sum_i \text{p}_i$ and $\delta^D_{\text{k}-\text{p}} - (2\pi)^3 \delta^{(D)}(\text{k} - \text{p})$.

By contrast, Eulerian perturbation theory solves perturbatively for the density and velocity at the observed Eulerian position.

$$\begin{aligned} \delta(\mathbf{k}) &= \sum_n \int_{\text{P}_1 \ldots \text{p}_n} \delta^D_{\text{k}-\text{p}_{1n}} F'_n \left(\mathbf{p}_1, \ldots, \mathbf{p}_n\right) \delta_{\text{lin}} \left(\mathbf{p}_1\right) \ldots \delta_{\text{lin}} \left(\mathbf{p}_n\right), \\ v_i(\mathbf{k}) &= -if\mathcal{H} \frac{k_i}{k^2} \sum_n \int_{\text{P}_1 \ldots \text{P}_n} \delta^D_{\text{k}-\text{p}_{1n}} G_n \left(\mathbf{p}_1, \ldots, \mathbf{p}_n\right) \delta_{\text{lin}} \left(\mathbf{p}_1\right) \ldots \delta_{\text{lin}} \left(\mathbf{p}_n\right). \end{aligned} \tag{1.39}$$

Nevertheless, despite their differences, LPT and EPT are formally equivalent and, as we will show later, yield consistent results.

1.7 Modeling biased tracers

Cosmological surveys face an inherent limitation, they cannot directly observe the underlying total matter distribution of the universe. Instead, these surveys detect tracers like galaxies and halos that form in the nonlinear regime of structure formation. This introduces complexities when connecting theoretical predictions to observations, as these tracers do not perfectly follow the matter density field. To address this problem, we follow two approaches.

In the Lagrangian approach, the initial positions of these tracers follow a distribution determined by local initial conditions,

$$\begin{aligned} F\left[\delta_{\rm lin}(\mathbf{q}), s_{{\rm lin},ij}(\mathbf{q}), \ldots, \nabla\delta_{\rm lin}(\mathbf{q})\right] =& 1 + \delta_g\left(\mathbf{q}, \tau_0\right) \\ =& 1 + b_1\delta_{\rm lin}(\mathbf{q}) + \frac{1}{2}b_2\left(\delta^2_{\rm lin}(\mathbf{q}) - \left\langle\delta^2_{\rm lin}\right\rangle\right) + b_5\left(s^2_{\rm lin}(\mathbf{q}) - \left\langle s^2_{\rm lin}\right\rangle\right) \\ &+ b_3 O_3(\mathbf{q}) + \cdots + b_\nabla\nabla^2\delta_{\rm lin}(\mathbf{q}) + \epsilon(\mathbf{q}) \end{aligned} \tag{1.40}$$

The function $F[\ldots]$ represents a general bias expansion in terms of the underlying matter fields. The quantity $\delta_g(\mathbf{q}, \tau_0)$ denotes the galaxy overdensity at present time τ_0. The parameters b_1, b_2, and b_5 are bias coefficients: The terms $\langle\delta^2_{\rm lin}\rangle$ and $\langle s^2_{\rm lin}\rangle$ are ensemble averages of the respective quantities, subtracted to ensure a mean-zero contribution for the bias. The parameter b_∇ is a higher-derivative bias coefficient, linked to the Laplacian of the linear density field, $\nabla^2\delta_{\rm lin}(\mathbf{q})$. Finally, $\epsilon(\mathbf{q})$ accounts for stochastic contributions, representing random fluctuations in the galaxy distribution not captured by deterministic bias terms.

Given this biased functional, these initial overdensities can then be mapped to the evolved overdensities of biased tracers via number conservation,

$$\begin{aligned} 1 + \delta_g(\mathbf{x}, \tau) &= \int d^3\mathbf{q} F(\mathbf{q})\delta_D(\mathbf{x} - \mathbf{q} - \Psi(\mathbf{q}, \tau)) \\ (2\pi)^3\delta_D(\mathbf{k}) + \delta_g(\mathbf{k}) &= \int d^3\mathbf{q} e^{i\mathbf{k}\cdot(\mathbf{q}+\Psi(\mathbf{q}))}F(\mathbf{q}) \end{aligned} \tag{1.41}$$

In LPT clustering arises from two distinct factors: the initial bias encoded in F(q), which reflects how tracers are distributed in their initial (Lagrangian) positions, and the nonlinear dynamics captured by the displacement field Ψ, which governs how those tracers move to their final (Eulerian) positions through $\mathbf{x} = \mathbf{q} + \Psi$.

In the Eulerian approach, the galaxy overdensity is expressed in terms of a bias expansion based on present-day operators. The tracer field is expanded up to third order in terms of the nonlinear Eulerian fields as,

$$\delta_h = c_1\delta + \frac{c_2}{2}\delta^2 + c_5 s^2 + \frac{c_3}{6}\delta^3 + c_{15}\delta s^2 + c_{st} st + c_{52} s^3 + c_\psi \psi \tag{1.42}$$

In this expression, the bias expansion is written in terms of the nonlinear Eulerian matter density field δ and other composite operators. The coefficient c_1 represents the linear Eulerian bias, while c_2 captures the second-order (quadratic) bias in δ. The parameter c_5 corresponds to the tidal bias, with s^2 being the square of the tidal tensor, defined as $s_{ij}s^{ij}$, where s_{ij} is the traceless part of the second derivative of the gravitational potential. The third-order contributions are described by several terms: c_3 multiplies δ^3, accounting for cubic bias in the density field; c_{15} multiplies the mixed term δs^2, representing interactions between density and tidal fields; and c_{52} multiplies s^3, a cubic tidal field contribution. The term $c_{st} st$ involves t, another tidal-related operator that appears at third order.

At one loop, LPT and EPT are not independent frameworks but can be connected through the following transformations of the bias parameters.

$$\begin{aligned} c_1 &= 1 + b_1 \\ c_2 &= b_2 + \frac{8}{21}b_1, \quad c_5 = b_5 - \frac{2}{7}b_1 \\ c_3 &= b_3 + ab_1 \end{aligned} \tag{1.43}$$

1.8 The Redshift-Space Power Spectrum

The redshift space power spectrum is a fundamental statistical measure of galaxy clustering that exhibits distortions when compared to its real space counterpart. These distortions arise from the peculiar velocities of galaxies, their deviations from the uniform Hubble flow due to gravitational interactions. Known as redshift-space distortions (RSD), these effects modify the observed spatial distribution of galaxies by altering their apparent positions along the line of sight.

$$z_{obs} = z_{\text{Hubble Flow}} + z_{\text{peculiar velocity}}, \tag{1.44}$$

with $z_{\text{peculiar velocity}} = \mathbf{v} \cdot \mathbf{n}/aH$. The resulting anisotropic power spectrum provides valuable cosmological information, as it encodes both the growth rate of cosmic structure and the expansion history of the universe.

To obtain the power spectrum in redshift space, we begin with the conservation of the number of particles, which gives us the following equation,

$$1 + \delta_s(s, \tau) = \int d^3x (1 + \delta_g(x, \tau)) \delta_D(s - x - u), \tag{1.45}$$

where s is the position in redshift space, x is the real-space coordinate, u is the displacement due to peculiar velocities, and δ_D is the Dirac delta. Taking the Fourier transform leads to

$$(2\pi)^3 \delta_D(\mathbf{k}) + \delta_s(\mathbf{k}) = \int d^3x (1 + \delta_g(x, \tau)) e^{i\mathbf{k} \cdot (\mathbf{x} + \mathbf{u}(x))} \tag{1.46}$$

We can define the moment-generating function $\tilde{M}(\mathbf{J}, \mathbf{k})$ as,

$$\tilde{M}(\mathbf{J}, \mathbf{k}) = \frac{k^3}{2\pi^2} \int d^3r \, e^{i\mathbf{k} \cdot \mathbf{r}} \langle (1 + \delta_g(\mathbf{x}_1))(1 + \delta_g(\mathbf{x}_2)) e^{i\mathbf{J} \cdot \Delta\mathbf{u}} \rangle. \tag{1.47}$$

where the term $\Delta\mathbf{u} = \mathbf{u}(\mathbf{x}_1) - \mathbf{u}(\mathbf{x}_2)$ represents the difference in the peculiar velocity field $\mathbf{u} = \hat{n}(\hat{n} \cdot \mathbf{v})/\mathcal{H}$. between the two points and $\mathbf{J}$ acts as a Fourier conjugate variable to $\Delta\mathbf{u}$. The redshift-space power spectrum $P_s(k)$ is obtained as a special case of the moment-generating function

$$\frac{k^3}{2\pi^2} P_s(k) = \tilde{M}(\mathbf{J} = \mathbf{k}, \mathbf{k}) = \frac{k^3}{2\pi^2} \int d^3r \, e^{i\mathbf{k} \cdot \mathbf{r}} \langle (1 + \delta_g(\mathbf{x}_1))(1 + \delta_g(\mathbf{x}_2)) e^{i\mathbf{k} \cdot \Delta\mathbf{u}} \rangle. \tag{1.48}$$

If we apply the cumulant theorem to the logarithm,

$$\ln[1+\Delta(k)] = \ln[1+\widetilde{M}(\mathbf{J}=0,\mathbf{k})] + iJ_i\widetilde{C}_i^{(1)}(\mathbf{k}) \quad \frac{1}{2}J_iJ_j\widetilde{O}_{ij}^{(2)} + \ldots \tag{1.49}$$

with J_i the i-component of $\mathbf{J}$ and $\tilde{C}^{(n)}_{i_1\ldots i_n}$ the cumulants. The first few cumulants are related to the Fourier pairwise velocity moments by

$$\begin{aligned}
\tilde{C}_i^{(1)}(\mathbf{k}) &= \frac{k^3}{2\pi^2}\frac{\tilde{\Xi}_i(\mathbf{k})}{1+\Delta^2},\\
\widetilde{C}_{ij}^{(2)}(\mathbf{k}) &= \frac{k^3}{2\pi^2}\frac{\tilde{\Xi}_{ij}(\mathbf{k})}{1+\Delta^2} - \tilde{C}_i^{(1)}\tilde{C}_j^{(1)},\\
\tilde{C}_{ijk}^{(3)}(\mathbf{k}) &= \frac{k^3}{2\pi^2}\frac{\tilde{\Xi}_{ijk}(\mathbf{k})}{1+\Delta^2} - \tilde{C}_{\{ij}^{(2)}\tilde{C}_{k\}}^{(1)} - \tilde{C}_i^{(1)}\tilde{C}_j^{(1)}\tilde{C}_k^{(1)},\\
\tilde{C}_{ijkl}^{(4)}(\mathbf{k}) &= \frac{k^3}{2\pi^2}\frac{\tilde{\Xi}_{ijkl}(\mathbf{k})}{1+\Delta^2} - \tilde{C}_{\{ijk}^{(3)}\tilde{C}_{l\}}^{(1)} - \tilde{C}_{\{ij}^{(2)}\tilde{C}_{kl\}}^{(2)} - \tilde{C}_i^{(1)}\tilde{C}_j^{(1)}\widetilde{C}_k^{(1)}\tilde{C}_l^{(1)},
\end{aligned} \tag{1.50}$$

where $\tilde{\Xi}^{(n)}_{i_1\ldots i_n}$ is defined as the Fourier transforms of

$$\Xi^{(n)}_{i_1\cdots i_n} = \langle (1+\delta_1)\,(1+\delta_2)\,\Delta\mathbf{u}_{i_1}\cdots\Delta\mathbf{u}_{i_n}\rangle\,, \tag{1.51}$$

with $\Delta\mathbf{u}_i = \mathbf{u}_1 - \mathbf{u}_2$. Then, the redshift-space power spectrum can be expressed as,

$$\mathbf{1} + \frac{k^3}{2\pi^2}P_s(\mathbf{k}) = \left(\mathbf{1} + \triangle^2(k)\right)\exp\left[\sum_{n-1}^{\infty}\frac{i^n}{n!}k_{i_1}\ldots k_{i_n}\tilde{C}^{(n)}_{i_1\ldots i_n}(\mathbf{k})\right]. \tag{1.52}$$

We have looked at the derivation and essential features of the redshift-space power spectrum, which we use to model galaxy clustering in the. We began by describing the basic equation governing the transition from real space to redshift space, where peculiar velocities add anisotropies. The moment-generating function was presented as a practical tool to obtain the redshift-space power spectrum. Finally, we calculate the higher-order moments using the cumulant expansion. In further chapters, we will look at different higher-order models based on the EPT or LPT basis, offering a potent approach for deriving cosmological data from spectroscopic surveys.

Chapter 2

Modeling Redshift Space Distortions

Perturbation theory (PT), in its two approaches—Lagrangian Perturbation Theory (LPT) and Eulerian Perturbation Theory (EPT)—provides a robust theoretical framework for modeling the correlation function and its Fourier-space counterpart, the power spectrum. Spectroscopic surveys yield catalogs of galaxies and quasars, which are then analyzed using statistical techniques to extract valuable information. By combining Bayesian inference with PT, we can derive essential cosmological insights from these statistical measures. These methods allow us to quantify galaxy clustering and place constraints on key cosmological parameters. In this chapter, we introduce more advanced models that incorporate various effects, beginning with our primary configuration-space model, the Gaussian Streaming Model [16]. We also discuss three Fourier-space models: PyBird, Velocileptors, and FOLPS.

2.1 Gaussian Streaming Model

The Gaussian Streaming Model (GSM) is a theoretical framework used to model redshift-space distortions (RSD) in the large-scale structure (LSS) of the universe [17]. It provides a way to predict the redshift-space correlation function by assuming that the probability distribution of galaxy velocities along the line of sight follows a Gaussian distribution. This model follows the Lagrangian approach to describe the evolution of cold dark matter (CDM) structures, following the trajectories of individual particles from their initial positions $\boldsymbol{q}$. These trajectories are determined by the displacement field $\boldsymbol{\Psi}(\boldsymbol{q},t)$.

In this approach, the peculiar velocity of a particle is given by $\boldsymbol{v} = a\dot{\boldsymbol{x}} = a\dot{\boldsymbol{\Psi}}$. Assuming the distant observer approximation, the maping between real and aparent position will be given by a shift of $\boldsymbol{u} \equiv \hat{\boldsymbol{n}}\frac{\boldsymbol{v}\cdot\hat{\boldsymbol{n}}}{aH} = \hat{\boldsymbol{n}}\frac{\dot{\boldsymbol{\Psi}}\cdot\hat{\boldsymbol{n}}}{H}$, where $\hat{\boldsymbol{n}}$ denotes the line-of-sight direction. Then, an object's apparent position, $\boldsymbol{s}$, will be displaced along the line of sight due to its peculiar velocity,

$$\boldsymbol{s} = \boldsymbol{q} + \boldsymbol{\Psi} + \hat{\boldsymbol{n}}\frac{\dot{\boldsymbol{\Psi}}\cdot\hat{\boldsymbol{n}}}{H}. \tag{2.1}$$

The starting point is the conservation of tracer counts X,

$$[1+\delta_s(s)]\, d^3s = [1+\delta_X(\boldsymbol{x})]\, d^3x. \tag{2.2}$$

From equations 2.2 and 1.45, we can obtain an expression for the correlation function of tracers in redshift space,

$$1+\xi_s(\boldsymbol{s}) = \int \frac{d^3k}{(2\pi)^3} d^3r e^{i\boldsymbol{k}\cdot(\boldsymbol{s}-\boldsymbol{r})}[1+\mathcal{M}(\boldsymbol{k},\boldsymbol{r})] \tag{2.3}$$

with $r = x_2 - x_1$ and $s = s_2 - s_1$.

As demonstrated in Chapter 1, the density-weighted pairwise velocity generating function is a useful tool. Here, we redefine it in configuration space.

$$1+\mathcal{M}(\boldsymbol{J},\boldsymbol{r}) = \left\langle (1+\delta_X(\boldsymbol{x}_1))\,(1+\delta_X(\boldsymbol{x}_2))\, e^{-i\boldsymbol{J}\cdot\Delta\boldsymbol{u}} \right\rangle \tag{2.4}$$

where $\Delta\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}_2) - \boldsymbol{u}(\boldsymbol{x}_1)$.

Now, we expand in cumulants $\mathcal{C}$ the logarithm of the generating function as,

$$\mathcal{Z}(\boldsymbol{J},\boldsymbol{r}) \equiv \log\Big[1+\mathcal{M}(\boldsymbol{J},\boldsymbol{r})\Big] = \sum_{n=0}^{\infty}\frac{(-i)^n}{n!}J_{i_1}\cdots J_{i_n}\mathcal{C}^{(n)}_{i_1\cdots i_n}(\boldsymbol{r}) \tag{2.5}$$

with

$$\mathcal{C}^{(n)}_{i_1\cdots i_n}(\boldsymbol{r}) = i^n \frac{\partial^n \mathcal{Z}(\boldsymbol{J},)}{\partial J_{i_1}\cdots\partial J_{i_n}}\bigg|_{\vec{J}=0}. \tag{2.6}$$

Then, the correlation function in redshift space can be expressed as

$$1+\xi_s(\boldsymbol{s}) = \int \frac{d^3k}{(2\pi)^3} d^3x\, e^{i\boldsymbol{k}\cdot(\boldsymbol{s}-\boldsymbol{r})} \exp\left[\sum_{n=0}^{\infty} \frac{(-i)^n}{n!} k_{i_1}\cdots k_{i_n} \mathcal{C}^{(n)}_{i_1\cdots i_n}(\boldsymbol{r})\right], \tag{2.7}$$

The generating function can also be expanded using:

$$\mathcal{C}^{(0)}(\boldsymbol{r}) = \log[1+\xi(r)], \tag{2.8}$$

$$\mathcal{C}^{(1)}_i(\boldsymbol{r}) = \frac{\Xi^{(1)}_i(\boldsymbol{r})}{1+\xi(r)} \equiv v^{\hat{n}}_{12,i}, \tag{2.9}$$

$$\mathcal{C}^{(2)}_{ij}(\boldsymbol{r}) = \frac{\Xi^{(2)}_{ij}(\boldsymbol{r})}{1+\xi(r)} - \mathcal{C}^{(1)}_i(\boldsymbol{r})\mathcal{C}^{(1)}_j(\boldsymbol{r}) \equiv \hat{\sigma}^{2\hat{n}}_{12,ij} - v^{\hat{n}}_{12,i}v^{\hat{n}}_{12,j} = \sigma^2_{12,ij}, \tag{2.10}$$

In this modeling, the quantities $v^{\hat{n}}_{12,i}$, $\sigma^{2\hat{n}}_{12,ij}$, and $\sigma^2_{12,ij}$, $\hat{\sigma}^{2\hat{n}}_{12,ij}$, represent the pairwise velocity and velocity dispersion along the line of sight, expressed in terms of moments and cumulants. The expressions for these terms are given in the appendix.

Now, we can reexpress equation (2.7), in terms of the moments and the correlation function in redshift space as

$$1+\xi_s(\boldsymbol{s}) = \int d^3r\big[1+\xi(r)\big] \int \frac{d^3k}{(2\pi)^3} \exp\Big[\boldsymbol{i}\boldsymbol{k}\cdot(\boldsymbol{s}-\boldsymbol{r}-\boldsymbol{v}^{\boldsymbol{n}}_{12}) - \frac{1}{2}\boldsymbol{k}^T\boldsymbol{\sigma}^{2\boldsymbol{n}}_{12}\,\boldsymbol{k} + \cdots\Big]. \tag{2.11}$$

If we stop at the second order cumulant $\boldsymbol{\sigma}^{2\boldsymbol{n}}_{12}$, the k-integral can be formally performed analytically, giving

$$1+\xi_s(\boldsymbol{s}) = \int \frac{d^3r}{(2\pi)^{3/2}|\sigma^2_{12}|^{1/2}}\big[1+\xi(r)\big] \exp\left[-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{r}-\vec{v}^{\boldsymbol{n}}_{12})^{\mathbf{T}}[\boldsymbol{\sigma}^{2\boldsymbol{n}}_{12}]^{-1}(\boldsymbol{s}-\boldsymbol{r}-\vec{v}^{\boldsymbol{n}}_{12})\right], \tag{2.12}$$

which is known as the *Gaussian Streaming Model.* Although the real expected velocity is not Gaussian, that approximation works well for scales above $s = 20h^{-1}$Mpc as demonstrated in the comparison with simulations of [18].

2.1.1 CLPT

To compute the moments, we use Convolution Lagrangian Perturbation Theory (CLPT) following [18–20]. Our model incorporates a Lagrangian biasing function F, which defines the connection between galaxy distributions and matter fluctuations, as described in [21, 22]:

$$1+\delta_X(\boldsymbol{q}) = F(\delta, \nabla^2\delta) = \int \frac{d^2\boldsymbol{\Lambda}}{(2\pi)^2}\tilde{F}(\boldsymbol{\Lambda})e^{i\mathbf{D}\cdot\boldsymbol{\Lambda}}. \tag{2.13}$$

Then, using the conservation of tracer number, which not necessarily hold for halos, but aligns with other widely used methods that impose biasing through the symmetries of the theory [23]. Here, $\tilde{F}(\boldsymbol{\Lambda})$ is the Fourier transforming of $F(\boldsymbol{D})$, with arguments $\mathbf{D} = (\delta, \nabla^2\delta)$ and spectral parameters $\boldsymbol{\Lambda} = (\lambda, \eta)$, dual to $\mathbf{D}$, we can obtain,

$$1+\delta_X(\boldsymbol{x}) = \int \frac{d^3q}{(2\pi)^3}\int d^3q e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{q})}\int \tilde{F}(\boldsymbol{\Lambda})e^{i\mathbf{D}\cdot\boldsymbol{\Lambda}-i\boldsymbol{k}\cdot\boldsymbol{\Psi}}, \tag{2.14}$$

The renormalized bias parameters are obtained following [21, 22]

$$b_{nm} = \int \frac{d\boldsymbol{\Lambda}}{(2\pi)^2}\tilde{F}(\boldsymbol{\Lambda})e^{-\frac{1}{2}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}}(i\lambda)^n(i\eta)^m, \tag{2.15}$$

However, in this work we consider only b_1 and b_2 following [19]. Aditionally we introduce the tidal Lagrangian bias, b_{s^2}, following [24].

To obtain the cumulants we need to find its relation with the moments. Starting from,

$$1+\delta_X(\boldsymbol{x}_1) = \int d^3q_1\frac{d^3k_1}{(2\pi)^3}\frac{d\lambda_1}{2\pi}\tilde{F}(\lambda_1)e^{i\lambda\delta_1+i\boldsymbol{k}_1\cdot(\boldsymbol{x}_1-q_1-\Psi_1)},$$

we can expand in cumulants to obtain,

$$\begin{aligned}\Xi^{(n)}_{i_1\cdots i_n}(\boldsymbol{r}) = \hat{n}_{i_i}\cdots\hat{n}_{i_n}\int d^3q\int\frac{d^3k}{(2\pi)^3}e^{i\boldsymbol{k}\cdot(\boldsymbol{q}-\boldsymbol{r})}\int\frac{d\lambda_1}{2\pi}\frac{d\lambda_2}{2\pi}\tilde{F}(\lambda_1)\tilde{F}(\lambda_2)\\ \times\hat{n}_{j_i}\cdots\hat{n}_{j_n}\left\langle\frac{\dot{\Delta}_{j_1}}{H}\cdots\frac{\dot{\Delta}_{j_n}}{H}e^{i[\lambda_1\delta_1+\lambda_2\delta_2+\boldsymbol{k}\cdot\Delta]}\right\rangle,\end{aligned}$$

with $\Delta_i \equiv \Psi_i(\boldsymbol{q}_2) - \Psi_i(\boldsymbol{q}_1)$ and used $\Delta u_i = H^{-1}(\hat{n}_j \dot{\Delta}_j)\hat{n}_i$.

In the context of CLPT, the real-space correlation function $\xi_X(r)$, is given by [18, 22, 25–28],

$$
\begin{aligned}
1 + \xi_X(r) = \int \frac{d^3q}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{e}r-\boldsymbol{q})} \Bigg\{ & 1 - \frac{1}{2}A_{ij}^{loop}G_{ij} + \frac{1}{6}\Gamma_{ijk}W_{ijk} \\
& + b_1(-2U_i g_i - A_{ij}^{10}G_{ij}) + b_1^2(\xi_L - U_iU_jG_{ij} - U_i^{11}g_i) + b_2(\frac{1}{2}\xi_L^2 - U_i^{20}g_i - U_iU_jG_{ij}) \\
& - 2b_1b_2\xi_L U_i g_i + 2(1+b_1)b_{\nabla^2\delta}\nabla^2\xi_L + b_{\nabla^2\delta}^2\nabla^4\xi_L \Bigg\},
\end{aligned}
\tag{2.16}
$$

the first moment of the generating function is given by the pairwise velocity,

$$
\begin{aligned}
v_{12,i}(\boldsymbol{r}) = \frac{f}{1+\xi_X(r)} \int \frac{d^3q\, e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})}}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} \Bigg\{ & - g_r\dot{A}_{ri} - \frac{1}{2}G_{rs}\dot{W}_{rsi} \\
& + b_1\left(2\dot{U}_i - 2g_r\dot{A}_{ri}^{10} - 2G_{rs}U_r\dot{A}_{si}\right) + b_1^2\left(\dot{U}_i^{11} - 2g_rU_r\dot{U}_i - g_r\dot{A}_{ri}\xi_L\right) \\
& + b_2\left(\dot{U}_i^{20} - 2g_rU_r\dot{U}_i\right) + 2b_1b_2\xi_L\dot{U}_i + 2b_{\nabla^2\delta}\nabla_i\xi_L \Bigg\},
\end{aligned}
\tag{2.17}
$$

and the second moment is the pairwise velocity dispersion,

$$
\begin{aligned}
\sigma_{12,ij}^2(\boldsymbol{r}) = \frac{f^2}{1+\xi_X(r)} \int \frac{d^3q\, e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})}}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} \Bigg\{ & \ddot{A}_{ij} - g_r\ddot{W}_{rij} - G_{rs}\dot{A}_{ri}\dot{A}_{sj} \\
& + 2b_1\left(\ddot{A}_{ij}^{10} - g_r\dot{A}_{r\{i}\dot{U}_{j\}} - g_rU_r\ddot{A}_{ij}\right) + b_1^2\left(\xi_L\ddot{A}_{ij} + 2\dot{U}_i\dot{U}_j\right) + 2b_2\dot{U}_i\dot{U}_j \Bigg\},
\end{aligned}
\tag{2.18}
$$

Here, the matrix A_{ij} is given by $A_{ij}(\boldsymbol{q}) = \langle \Delta_i^{(1)}\Delta_j^{(1)}\rangle_c = 2\int \frac{d^3p}{(2\pi)^3}\left(1 - e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\right)\frac{p_ip_j}{p^4}P_L(p) = A_{ij}^L(\boldsymbol{q}) + A_{ij}^{loop}(\boldsymbol{q})$, and $\Delta_i = \Psi_i(\boldsymbol{q}_2) - \Psi_i(\boldsymbol{q}_1)$. The linear power spectrum is the one obtained in Standard perturbation Theory $\xi_L(q) = \int \frac{d^3p}{(2\pi)^3}e^{i\boldsymbol{p}\cdot\boldsymbol{q}}P_L(p)$.

Now, if we decompose the vector $\boldsymbol{k}$, $\boldsymbol{s}$ and $\boldsymbol{r}$ in components parallel and perpendicular to the line of sight $\boldsymbol{n}$, Then, we can split the $\boldsymbol{k}$ integral in parallel and perpendicular to the line-of-sight integrations, leading to the correlation function within the GSM as

$$1+\xi_s(s_\parallel, s_\perp) = \int_{-\infty}^{\infty} \frac{dr_\parallel}{\left[2\pi\sigma_{12}^2(r,\mu)\right]^{1/2}} \left[1+\xi(r)\right] \exp\left[-\frac{\left(s_\parallel - r_\parallel - \mu v_{12}(r)\right)^2}{2\sigma_{12}^2(r,\mu)}\right], \tag{2.19}$$

with $r^2 = r_\parallel^2 + s_\perp^2$.

2.1.2 EFT terms

At large scales ($\boldsymbol{q} \to \infty$), the terms inside the curly brackets in Eq. (3.20) lead to a nonzero zero-lag correlator, expressed as an integral over the linear power spectrum and loop corrections. However, since standard perturbation theory struggles to accurately describe correlations at zero separation, an effective field theory (EFT) counterterm must be introduced to account for these effects. This additional contribution modifies the pairwise velocity dispersion by shifting $\hat{\sigma}^2_{12,mn}$ according to:

$$\hat{\sigma}^2_{12,mn} \to \hat{\sigma}^2_{12,mn} + \sigma^2_{\rm EFT}\delta_{mn}\frac{1+\xi^{\rm ZA}(r)}{1+\xi_X^{\rm CLPT}(r)}. \tag{2.20}$$

We also include the $c_1^{\rm EFT}$ parameter in the , despite it is degenerated with curvature bias [19]), where its effects are encoded in the second cumulant of the pairwise velocity generation function and has a small impact in the monopole correlation function and only the quadrupole is quite sensitive to it at scales around $r < 40\ h^{-1}$Mpc.

2.1.3 `gsm-eft` code

The `gsm-eft` code, written in `C` and available on https://github.com/alejandroaviles/gsm, computes the one-loop GSM two-point correlation function efficiently.

The CLPT integrals involve a $\boldsymbol{q}$-integration with a Gaussian kernel centered at $\boldsymbol{q} = \boldsymbol{r}$, which becomes challenging for large r due to the need for a finer angular grid. To address this, we re-center the integrals at $\boldsymbol{q} = \boldsymbol{r}$, allowing for accurate angular integration using the Gauss-Legendre method with just `gsm_NGL` $= 16$ weights.

Aditionally, When exploring the parameter space by MCMC methods, the cumulant σ_{12}^2 can become negative, leading to unphysical results. To prevent this, we decompose σ_{12}^2

into linear and loop components and introduce a transformation when the loop term falls below a threshold $-c_{tol}\sigma^2_{12,L}$, with $c_{tol} \approx 0.999$. This transformation smooths σ^2_{12}, ensuring it remains strictly positive while preserving continuity,

$$\sigma^2_{12} = \sigma^2_{12,L} + \sigma^2_{12,\mathrm{loop}} \quad \longrightarrow \quad \sigma^2_{12} = \sigma^2_{12,L} + f(\sigma^2_{12,L}, \sigma^2_{12,\mathrm{loop}}) \tag{2.21}$$

where

$$f(\sigma^2_{12,L}, \sigma^2_{12,\mathrm{loop}}) = \frac{A\,\sigma^2_{12,\mathrm{loop}}}{\sigma^2_{12,\mathrm{loop}} + B} - A - \sigma^2_{12,L}, \tag{2.22}$$

with constants A and B are given by

$$A = \frac{(-1 + c_{tol})(B - c_{tol}\,\sigma^2_{12,L})}{B}\sigma^2_{12,L}, \quad B = (-1 + 2c_{tol})\sigma^2_{12,L} \tag{2.23}$$

This ensures that the one-loop cumulant σ^2_{12} remains smooth and positive. GSM requires the following input parameters:

Cosmological Parameters
H_0
ω_b
ω_{cdm}
$\log(10^{10}A_s)$
w_0
w_a
n_s

Lagrangian Bias Parameters
$b_1, b_2, b_{s^2}, b_{\nabla^2\delta}$

EFT parameters
σ^2_{EFT} and c_1^{EFT}

Table 2.1: List of parameters needed for GSM labeled as cosmological, Lagrangian Bias Parameters, and EFT parameters. Table extracted from [29].

2.2 Fourier Space Models

2.2.1 Pybird

PyBird is a computational model based on 1-loop Eulerian perturbation theory, developed for analyzing biased tracers in redshift space. It employs an effective field theory (EFT) approach, with a parameterization that allows for a straightforward transformation between the two models [30]. The Python implementation of PyBird can be found publicly at: `https://github.com/pierrexyz/pybird.`

The EFT formalism in PyBird leads to a power spectrum model for biased tracers in redshift space, which includes loop corrections and counterterms. The redshift-space galaxy power spectrum up to one-loop order is expressed as:

$$
\begin{aligned}
P_g(k,\mu) = {} & Z_1(\mu)^2 P_{\text{lin}}(k) + 2\int \frac{d^3q}{(2\pi)^3} Z_2(\mathbf{q},\mathbf{k}-\mathbf{q},\mu)^2 P_{\text{lin}}(|\mathbf{k}-\mathbf{q}|) P_{\text{lin}}(q) \\
& + 6 Z_1(\mu) P_{\text{lin}}(k) \int \frac{d^3q}{(2\pi)^3} Z_3(\mathbf{q},-\mathbf{q},k,\mu) P_{\text{lin}}(q) \\
& + 2 Z_1(\mu) P_{\text{lin}}(k) \left(c_{ct}\frac{k^2}{k_M^2} + c_{r,1}\mu^2\frac{k^2}{k_M^2} + c_{r,2}\mu^4\frac{k^2}{k_M^2} \right. \\
& \left. + \frac{1}{\pi_g}\left(c_{s,1} + c_{s,2}\frac{k^2}{k_M^2} + c_{s,3} f\mu^2 \frac{k^2}{k_M^2} \right) \right)
\end{aligned}
\tag{2.24}
$$

where P_{lin} is the linear power spectrum of Cold Dark Matter and baryons. The SPT kernels Z_n help to model the effect of RSD and galaxy bias at different perturbative orders. It includes EFT counter terms $c_{ct}, c_{r,1}, c_{r,2}$ to model short-wavelength effects into long-wavelength modes. The stochastic terms are$c_{\epsilon,1}, c_{\epsilon,2}, c_{\epsilon,3}$.The 4 bias parameters: b_1, b_2, b_3, b_4 are encoded in the redshift kernel Z_n calculations. These parameters are listed in Table 2.2.

2.2.2 Velocileptors

`velocileptors` is a fully resummed, one-loop, Lagrangian Perturbation Theory model that includes stochastic contributions and counterterms [32]. These terms contribute to

Cosmological Parameters
H_0
ω_b
ω_{cdm}
$\log(10^{10}A_s)$
w_0
w_a
n_s
Eulerian Bias Parameters
b_1, b_2, b_3, b_4
Counterterms
$c_{ct}, c_{r,1}, c_{r,2}$
Stochastic Parameters
$c_{s,1}, c_{s,2}, c_{s,3}$

Table 2.2: List of input parameters used by pybird labeled as cosmological, bias, counterterms, and stochastic parameters. Table extracted from [31].

the final LPT redshift-space power spectrum:

$$P_s(k) = P_s^{\mathrm{LPT}}(k) + k^2(\alpha_0 + \alpha_2\mu^2 + \alpha_4\mu^4)P_s^{\mathrm{Zel}}(k) + R_3^3(1 + \sigma_2 k^2\mu^2 + \sigma_4 k^4\mu^4), \qquad (2.25)$$

The expression includes the nonlinear LPT power spectrum $P_s^{\mathrm{LPT}}(k)$, the Zeldovich power spectrum $P_s^{\mathrm{Zel}}(k)$ which correct for small-scale effects using EFT counterterms $\alpha_0, \alpha_2, \alpha_4$. Additionally, it includes stochastic contributions parameterized by R_3^3 and higher-order angular corrections involving σ_2 and σ_4. In this way, `velocileptors` incorporates an Effective Perturbation Theory (EPT) model that consistently accounts for biases while employing a distinct resummation approach. This model is built using Lagrangian Perturbation Theory (LPT) expressions with an infrared (IR) resummation scheme that suppresses the Baryon Acoustic Oscillation (BAO) signal. The parameters needed as input are listed in Table 2.3.

Cosmological Parameters
H_0
ω_b
ω_{cdm}
$\log(10^{10}A_s)$
w_0
w_a
n_s
LPT Bias Parameters
$(1+b_1)\sigma_8, b_2, b_s, b_3$
Stochastic/Counterterms
$\alpha_0, \alpha_2, SN_0, SN_2$

Table 2.3: List of input parameters used by velocileptors labeled as cosmological, bias, counterterms, and stochastic parameters. Table extracted from [31].

2.2.3 FOLPS

`Folps_nu` is an Effective Field Theory (EFT) model implemented in Python, designed to incorporate massive neutrinos [33]. This model is based in the EFT up to one-loop order and computes the redshift-space power spectrum using the following formulation:

$$P^s_{\mathrm{EFT}}(k,\mu) = P_{\delta\delta}(k) + 2f_0\mu^2 P_{\delta\theta}(k) + f_0^2\mu^4 P_{\theta\theta}(k) + A^{\mathrm{TNS}}(k,\mu) + D(k,\mu) \tag{2.26}$$

$$+(\alpha_0 + \alpha_2\mu^2 + \alpha_4\mu^4)k^2 P_{\mathrm{lin}}(k) + P_{\mathrm{shot}}\left[\alpha_0^{\mathrm{shot}} + \alpha_2^{\mathrm{shot}}(k\mu)^2\right], \tag{2.27}$$

This expression includes contributions from the tracers one loop real power spectrum for the fields of velocity and density: $P_{\delta\delta}(k)$, $P_{\delta\theta}(k)$, and $P_{\theta\theta}(k)$. Additional terms $A^{\mathrm{TNS}}(k,\mu)$ and $D(k,\mu)$ account for nonlinear corrections from mode coupling and higher-order effects defined in [34]. The EFT counterterms $(\alpha_0+\alpha_2\mu^2+\alpha_4\mu^4)k^2P_{\mathrm{lin}}(k)$ capture the backreaction of small-scale nonlinearities on large-scale modes, as well as the nonlinear mapping from real to redshift space, while the stochastic (shot noise) component $P_{\mathrm{shot}}[\alpha_0^{\mathrm{shot}} + \alpha_2^{\mathrm{shot}}(k\mu)^2]$ models noise from the discreteness of tracers like galaxies.

A key feature of `Folps_nu` is that it employs EdS kernels (called `fk`-kernels) to properly account for the effects of the free-streaming scale introduced by massive neutrinos. The source code for `Folps_nu` is available at: `https://github.com/henoriega`. The parameters needed as input are listed in Table 2.4.

Cosmological Parameters
H_0
ω_b
ω_{cdm}
$\log(10^{10} A_s)$
w_0
w_a
n_s
Eulerian Bias Parameters
$b_1, b_2, b_{s^2}, b_{3nl}$
Stochastic/Counterterms
$\alpha_0, \alpha_2, \alpha_0^{shot}, \alpha_2^{shot}$

Table 2.4: List of input parameters used by FOLPS labeled as cosmological, bias, counterterms, and stochastic parameters. Table extracted from [31].

Chapter 3

Extracting Cosmological Information from Redshift-Space Galaxy Clustering

To extract cosmological information from redshift-space galaxy clustering, we can follow the framework of perturbation theory. Several models have introduced additional terms to account for various effects, enhancing our understanding of the large-scale structure of the universe, among them, the four models introduced in the previous chapter, the Gaussian Streaming Model, and the Fourier models PyBird, Velocileptors, and FOLPS. By analyzing the spatial distribution of galaxies in redshift space and comparing it with the theoretical model, we can infer key cosmological parameters, probe the growth of cosmic structures, or test different theories of gravity. In this chapter, we incorporate these additional ingredients into three distinct methodologies, each coupled with different approaches to explore the parameter space using Markov chain Monte Carlo (MCMC) following Bayesian inference.

The standard approach compressed this information into the Alcock-Paczyński parameters and $f\sigma_8$. However, more recent methodologies have emerged, one of which is the full modeling approach, which involves analyzing the complete shape of the power spectrum or the correlation function, including the linear power spectrum. Another approach is shape fit, which extends the standard method by allowing variations in the slope of the linear power spectrum, providing additional flexibility in cosmological analyses.

3.1 Standard approach

The standard approach is a methodology based on a fixed template, where the fiducial cosmology is held constant while allowing variations through the parameterization of three key variables associated with late-time dynamics, the Alcock-Paczyński parameters $\{\alpha_\parallel, \alpha_\perp\}$ and the growth factor multiplied by the amplitude of matter density fluctuations on large scales in the universe $f\sigma_8$. The parameters $\alpha_\perp$ and $\alpha_\parallel$ represent the scaling factors for the dilations perpendicular and parallel to the line of sight, and are defined as:

$$\alpha_\perp(z) = \frac{D_A(z)\, r_\mathrm{d}^\mathrm{ref}}{D_A^\mathrm{ref}(z)\, r_\mathrm{d}}, \quad \alpha_\parallel(z) = \frac{H^\mathrm{ref}(z)\, r_\mathrm{d}^\mathrm{ref}}{H(z)\, r_\mathrm{d}}, \tag{3.1}$$

The quantities involved include D_A, which corresponds to the angular diameter distance, H, referring to the Hubble parameter, and r_d denotes the sound horizon at the drag epoch. The notation `ref` in the superscript indicates that the estimation is performed under the reference cosmology. [1].

On the other hand, The parameter $f\sigma_8$ modulates the relative amplitude of the power spectrum. The linear growth rate of structures, f, was previously introduced in 1.23, meanwhile, the quantity σ_8, which characterizes the amplitude of matter power spectrum fluctuations on scales of up to $8\,h^{-1}\,\mathrm{Mpc}$ is defined as

$$\sigma_8^2(z) = \int_0^\infty \mathrm{d}(\ln k)\, k^3 P(k,z) W_{TH}^2(k\, 8h^{-1}\mathrm{Mpc}) \tag{3.2}$$

with W_{TH} the spherical top hat filter. In the standard methodology, σ_8 is determined by the reference cosmology and is fully degenerate with f, which is why we consider their product.

In this work, we employ the alternative parametrization:

---

[1] In this context, ”reference cosmology” is used interchangeably with ”fiducial cosmology,” without differentiating between grid-based or template cosmologies

$$\alpha = \alpha_{||}^{1/3}\alpha_{\perp}^{2/3}, \quad \epsilon = \left(\frac{\alpha_{||}}{\alpha_{\perp}}\right)^{1/3}. \tag{3.3}$$

We introduce distortions in the clustering analysis by replacing s with $s'(s_{\text{ref}}, \mu_{\text{ref}})$ and μ with $\mu'(\mu_{\text{ref}})$, where these transformations are derived using the re-scaling parameters α and ϵ, following:

$$s'(s_{\text{ref}}, \mu_{\text{ref}}) = s_{\text{ref}}\ \alpha\sqrt{(1+\epsilon)^4\mu_{\text{ref}}^2 + (1-\mu_{\text{ref}}^2)(1+\epsilon)^{-2}}, \tag{3.4}$$

$$\mu'^2(\mu_{\text{ref}}) = \left[1 + \left(\frac{1}{\mu_{\text{ref}}^2} - 1\right)(1+\epsilon)^{-6}\right]^{-1}. \tag{3.5}$$

The multipoles $\xi_\ell(s_{\text{ref}})$ are estimated within the reference cosmology using the transformed coordinates $s'(s_{\text{ref}}, \mu_{\text{ref}})$ and $\mu'(\mu_{\text{ref}})$. To incorporate the dilation parameters into our framework, we interpolate each modeled multipole ξ_ℓ^{model} based on these transformed coordinates. Additionally, the observed Legendre polynomials, $\mathcal{L}^{\text{obs}}(\mu')$, are computed using $\mu'(\mu_{\text{ref}})$. The observed correlation function $\xi^{\text{obs}}(s', \mu')$ is then constructed by summing the product of each multipole with its corresponding Legendre polynomial. The final observed multipoles are expressed as $\xi_\ell(s') = \sum_{\ell'} a_{\ell\ell'}\xi_{\ell'}(s)$, where the coefficients $a_{\ell\ell'}$ govern the transformation. As demonstrated in [35], this methodology achieves a good approximation taking to the order $\ell' = 4$.

This methodology enables us to fit the data multipoles to the observed multipoles, allowing us to derive constraints on the compressed parameters $\Theta_{\texttt{obs}}$ and their covariance C_Θ. We can translate this information into cosmological parameters by using a cosmological model, this procedure is carried out in §3.2.1.

3.2 Shape Fit

The *ShapeFit* methodology introduced in [36–39] refines the *Standard* approach by introducing two additional parameters that modify the shape of the power spectrum. These parameters, m and n, adjust the spectral slope—n accounts for deviations from the fiducial spectral index n_s, while m introduces a scale-dependent correction that impacts wave numbers near and beyond the turnaround scale. By compressing the data vector into five key parameters, $\alpha_{||}, \alpha_{\perp}, f\sigma_{s8}, m, n$, *ShapeFit* enhances constraints while preserving the model-independent nature of the standard methodology, making it a strong alternative to *Full Modeling* approach. These two additional parameters modify the reference power spectrum $P_{\text{ref}}(k)$, introducing variations in the fiducial cosmology and producing a new trial power spectrum $P'_{\text{ref}}(k)$ defined as:

$$\ln\left(\frac{P'_{\text{ref}}(k)}{P_{\text{ref}}(k)}\right) = \frac{m}{a}\tanh\left[a\ln\left(\frac{k}{k_p}\right)\right] + n\ln\left(\frac{k}{k_p}\right), \tag{3.6}$$

here, the pivot scale defined as $k_p = \pi/r_d$, where the baryon suppression reaches its peak. Using the methodology outlined in [40], we relate the slope parameter m to the cosmology by taking the derivative with respect to $\ln k$ of a ratio that combines both the de-wiggled linear power spectrum $P^{\text{lin}}_{\text{no-wiggle}}$ and the primordial power spectrum $\mathcal{P}_{\mathcal{R}}$ via:

$$m = \frac{d}{d\ln k}\left(\ln\left[\frac{P^{\text{lin}}_{\text{no-wiggle}}\ (k_p/s, \boldsymbol{\Theta})}{P^{\text{lin}}_{\text{no-wiggle}}\ \left(k_p, \boldsymbol{\Theta}^{\textbf{ref}}\right)}\frac{\mathcal{P}_{\mathcal{R}}(k, \Theta^{\text{ref}})}{\mathcal{P}_{\mathcal{R}}(k/s, \Theta)}\right]\right)\Bigg|_{k=k_p}, \tag{3.7}$$

where we use the Eisenstein-Hu formalism to obtain the *de-wiggled* linear power spectrum linear power spectrum, $P^{\text{lin}}_{\text{no-wiggle}}$.

If n is treated as a free parameter, we can use the following relationships to determine the spectral index n_s.

$$n = n_s - n_s^{\text{ref}}, \tag{3.8}$$

where the subscript ref indicates the reference cosmology.

The last modification introduced by *ShapeFit* to the *Standard* methodology is the redefinition of the clustering amplitude, which ensures that all compressed variables are consistently expressed in sound horizon units via σ_{s8}:

$$\sigma^2_{s8}(z, \Omega) = \int_0^\infty \mathrm{d}(\ln k)\,(ks)^3 P(ks, z) W^2_{\text{TH}}(ks\, 8\, h^{-1}\text{Mpc}), \tag{3.9}$$

where $s = r_d/r_d^{\text{ref}}$.

One advantage of this methodology is that we avoid recalculating all the loop corrections. Instead, we utilize the following approximations within Eulerian Perturbation Theory, saving computing time in the process without sacrificing accuracy.

$$
\begin{aligned}
P_{22}(k) &= \int \frac{dp^3}{(2\pi)^3} P'_{\rm ref}(p) P'_{\rm ref}(|\boldsymbol{k}-\boldsymbol{p}|) F_2(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{p}) \qquad (3.10)\\
&\approx \left(\frac{P'_{\rm ref}(k)}{P_{\rm ref}(k)}\right)^2 \int \frac{dp^3}{(2\pi)^3} P_{\rm ref}(p) P_{\rm ref}(|\boldsymbol{k}-\boldsymbol{p}|) F_2(\boldsymbol{k}, \boldsymbol{k}-\boldsymbol{p}), \qquad (3.11)
\end{aligned}
$$

These approximations are essential for efficient MCMC computations in the ShapeFit approach and have been validated in [41].

ShapeFit in Configuration Space

ShapeFit can also be implemented in configuration space, offering a complementary approach to the Fourier space analysis. In this framework, we focus on measuring the correlation function rather than the power spectrum. As illustrated in Figure 3.1, we compute the correlation function by employing the Gaussian Streaming Model (GSM) detailed in Section 2.1, together with the Convolution Lagrangian Perturbation Theory framework [26]. Following this approach, we extract cosmological information via compressed parameters. Similar to the Fourier space analysis, we employ various techniques to accelerate computation and obtain eq. (2.19), namely the real-space correlation function $\xi(r)$, the pairwise velocity $v_{12}(r)$, and the pairwise velocity dispersion $\sigma_{12}(r)$.

For instance, taking the Fourier transform of Equation 1.36 produces the one-loop, unbiased dark matter real-space correlation function, as demonstrated in [26], which is a simplification of the method used here, but serves as an illustration:

$$
1+\xi(r) = \int \frac{d^3q}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})} \left\{1 - \frac{1}{2}A_{ij}^{\rm loop}G_{ij} + \frac{1}{6}\Gamma_{ijk}W_{ijk}\right\}, \qquad (3.12)
$$

We can take advantage of the structure of the equation, which can be separated into linear and loop pieces,

$$\begin{aligned} A_{ij}(\boldsymbol{q}) &= 2\int \frac{d^3q}{(2\pi)^3}\big(1-e^{i\boldsymbol{p}\cdot\boldsymbol{q}}\big)\frac{k_ik_j}{k^4}\left(P_L(k)+\frac{3}{7}Q_1(k)+\frac{10}{21}R_1(k)\right) \\ &= A^L_{ij}(\boldsymbol{q})+A^{\text{loop}}_{ij}(\boldsymbol{q}), \end{aligned} \tag{3.13}$$

Here, the k-functions *k-functions* $Q_1(k)$ and $R_1(k)$ are derived from four-point correlations of linear density fields, which have a similar structure to eq. 3.10. They are defined in [25] as,

$$Q_1(k) = \int \frac{d^3p}{(2\pi)^3} q_1(\boldsymbol{k},\boldsymbol{p})P_L(p)P_L(|\boldsymbol{k}-\boldsymbol{p}|), \tag{3.14}$$

$$R_1(k) = \int \frac{d^3p}{(2\pi)^3} r_1(\boldsymbol{k},\boldsymbol{p})P_L(k)P_L(p), \tag{3.15}$$

Using the kernels q_1 and r_1, we generate the trial matrix $A\prime_{ij}$ by substituting the trial reference power spectrum $P\prime_{ref}$ from eq. (3.6) into eq. (3.13). To make the computations faster—similar to the approach used for the non-linear power spectrum, then, we calculate the functions Q_1 and R_1 as follows.

$$Q_1(k;m,n) = \left(\frac{P'_{\text{ref}}(k;m,n)}{P_{\text{ref}}(k)}\right)^2 \int \frac{d^3p}{(2\pi)^3} q_1(\boldsymbol{k},\boldsymbol{p})P_{\text{ref}}(p)P_{\text{ref}}(|\boldsymbol{k}-\boldsymbol{p}|), \tag{3.16}$$

$$R_1(k;m,n) = \left(\frac{P'_{\text{ref}}(k;m,n)}{P_{\text{ref}}(k)}\right)^2 \int \frac{d^3p}{(2\pi)^3} r_1(\boldsymbol{k},\boldsymbol{p})P_{\text{ref}}(k)P_{\text{ref}}(p), \tag{3.17}$$

Here, we indicate how the trial functions depend on the ShapeFit parameters.

We extend the method to higher orders by aplying the same trick to all k-functions, from which the q-functions are derived. These q-functions are then used to compute the real-space correlation function, $\xi(r)$, the pairwise velocity $v_{12}(r)$, and velocity dispersion $\sigma_{12}(r)$ as presented in [18, 22, 25–28]. These components are essential for constructing the redshift-space correlation function within the Gaussian Streaming Model. For biased tracers, we employ Convolution Lagrangian Perturbation Theory (CLPT) to compute the necessary ingredients. In this framework, the real-space correlation function for tracer X, is obtained within CLPT eq. 3.12, eq. 3.19 and eq. 3.20, which expressed in terms of the bias parameters b_1, b_2 and b_{s^2} and EFT counterterm $c_{1,\mathrm{EFT}}$ take the form for the real-space correlation function:

$$
\begin{aligned}
1+\xi_X(r) = \int \frac{d^3q}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})} \Bigg\{ & 1 - \frac{1}{2}A_{ij}^{loop}G_{ij} + \frac{1}{6}\Gamma_{ijk}W_{ijk} \\
& + b_1(-2U_i g_i - A_{ij}^{10}G_{ij}) + b_1^2(\xi_L - U_iU_jG_{ij} - U_i^{11}g_i) + b_2(\frac{1}{2}\xi_L^2 - U_i^{20}g_i - U_iU_jG_{ij}) \\
& - 2b_1b_2\xi_L U_i g_i - 2b_{s^2}g_iV_i^{10} + b_{s^2}^2\zeta - 2b_1b_{s^2}g_iV_i^{12} + b_2b_{s^2}\chi^{12} - \frac{1}{2}c_{1,\mathrm{EFT}}\,\mathrm{tr}\,G \Bigg\},
\end{aligned} \tag{3.18}
$$

The pairwise velocity:

$$
\begin{aligned}
v_{12}(\boldsymbol{r}) = \frac{f\,\hat{r}_i}{1+\xi_X(r)} \int \frac{d^3q\, e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})}}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} \Bigg\{ & -g_r\dot{A}_{ri} - \frac{1}{2}G_{rs}\dot{W}_{rsi} \\
& + b_1\left(2\dot{U}_i - 2g_r\dot{A}_{ri}^{10} - 2G_{rs}U_r\dot{A}_{si}\right) + b_1^2\left(\dot{U}_i^{11} - 2g_rU_r\dot{U}_i - g_r\dot{A}_{ri}\xi_L\right) \\
& + b_2\left(\dot{U}_i^{20} - 2g_rU_r\dot{U}_i\right) + 2b_1b_2\xi_L\dot{U}_i \Bigg\},
\end{aligned} \tag{3.19}
$$

and the pairwise velocity dispersion:

$$
\begin{aligned}
\sigma_{12}^2(\boldsymbol{r}) = \frac{f^2\,\hat{r}_i\hat{r}_j}{1+\xi_X(r)} \int \frac{d^3q\, e^{-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{q})^{\mathbf{T}}\mathbf{A}_L^{-1}(\boldsymbol{r}-\boldsymbol{q})}}{(2\pi)^{3/2}|\mathbf{A}_L|^{1/2}} \Bigg\{ & \ddot{A}_{ij} - g_r\ddot{W}_{rij} - G_{rs}\dot{A}_{ri}\dot{A}_{sj} \\
& + 2b_1\left(\ddot{A}_{ij}^{10} - g_r\dot{A}_{r\{i}\dot{U}_{j\}} - g_rU_r\ddot{A}_{ij}\right) + b_1^2\left(\xi_L\ddot{A}_{ij} + 2\dot{U}_i\dot{U}_j\right) + 2b_2\dot{U}_i\dot{U}_j \Bigg\},
\end{aligned} \tag{3.20}
$$

Although our formalism does not incorporate every perturbation theory (PT) correction and counterterm up to the one-loop level, it has been demostrated its robustness [16] and confirmed that it meets the requirements of the DESI Year one data. For those interested in a more comprehensive treatment, complete Lagrangian Perturbation Theory/Effective Field Theory frameworks are available in the literature [42, 43].

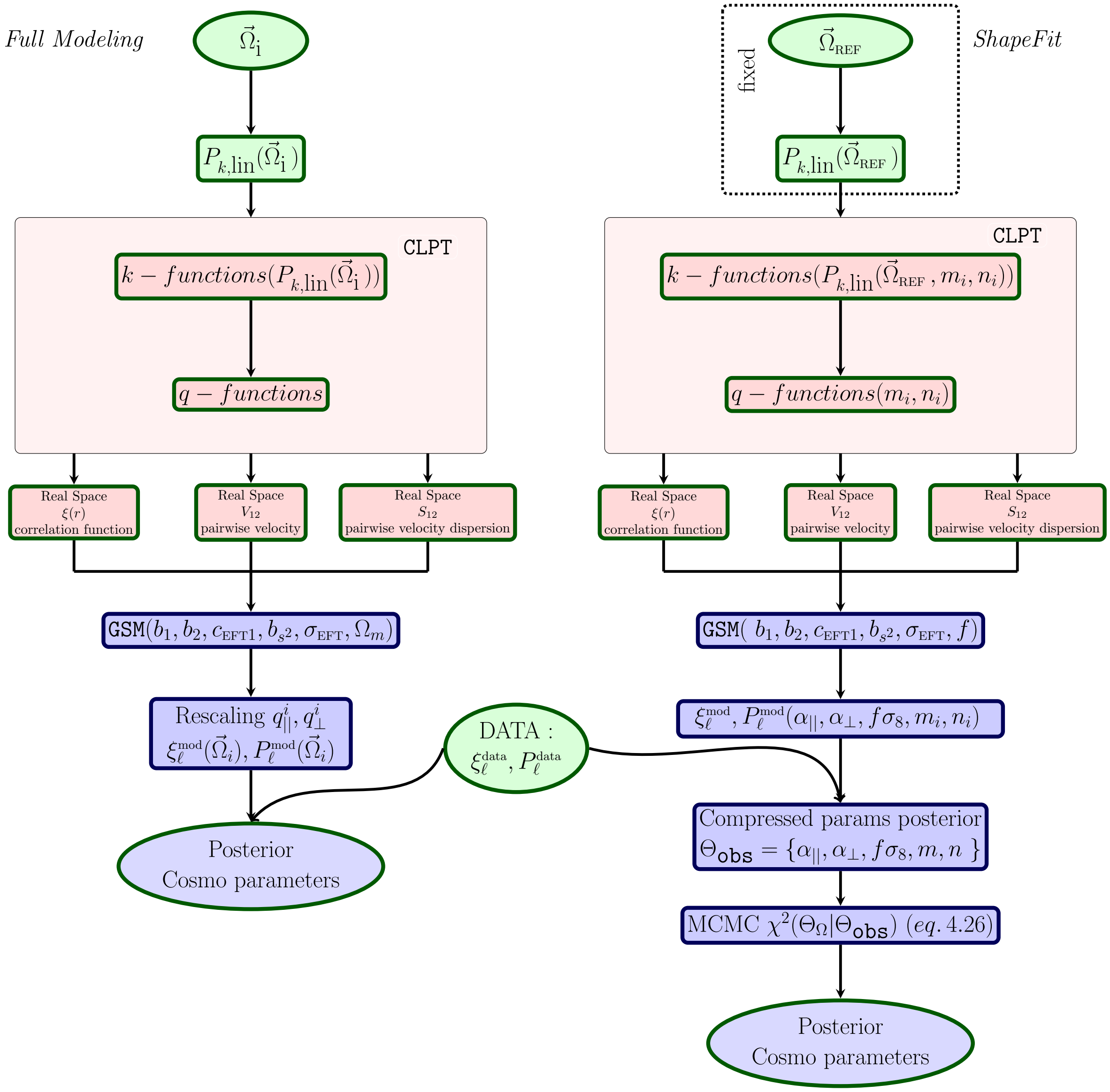

Figure 3.1: Schemes for the *Full Modeling* (left panel) and the *ShapeFit* (right panel) analysis, allowing to show the differences in the modeling and the cosmological parameters extraction of the information. Figure extracted from [44].

3.2.1 Extracting Cosmological Parameters from Compressed Observables

Both the Standard and ShapeFit methodologies compress cosmological information into the Alcock-Paczynski parameters $\alpha_\perp$, $\alpha_\parallel$ and $f\sigma_8$. As illustrated in Figure 3.1, an additional step is required to translate these compressed parameters into the underlying cosmological parameters.

In our analysis, we begin by compressing the cosmological information into a set of observables, denoted by $\Theta_{\texttt{obs}}$, which can include $\alpha_\perp$, $\alpha_\parallel$ and $f\sigma_8$, n and m. To convert these compressed parameters and recover the cosmological parameters, $\Omega = \{\omega_m, A_s, h, ...\}$. For any given cosmological model defined by Ω, we can compute the corresponding set of compressed parameters, Θ_Ω, using the theoretical relations of Sections 3.1 and 3.2. These relations allow us to predict the compressed observables from our chosen cosmology.

To identify the best-fit region in the cosmological parameter space, we run a Markov Chain Monte Carlo (MCMC). In this analysis, we assume that the observed compressed parameters follow a Gaussian distribution. Let C_Θ be the covariance matrix of these parameters, then, the likelihood is quantified via the chi-square statistic

$$\chi^2(\Theta_\Omega - \Theta_{\texttt{obs}}) = (\Theta_\Omega - \Theta_{\texttt{obs}})^T C_\Theta^{-1} (\Theta_\Omega - \Theta_{\texttt{obs}}). \tag{3.21}$$

where the value of this χ^2 guides the MCMC to explore the regions of Ω that yield Θ_Ω in good agreement with the observations $\Theta_{\texttt{obs}}$.

We employ two compression methodologies: the *Standard* and the *Shapefit* approaches. In the *Standard* methodology, the compressed parameter set is given by,

$$\Theta_\Omega = \{D_A(z,\Omega)/r_d(\Omega), D_H(z)/r_d(\Omega), f\sigma_8(z,\Omega)\}, \tag{3.22}$$

where $D_A(z,\Omega)$ is the angular diameter distance, $D_H(z,\Omega)$ is the Hubble distance, $r_d(\Omega)$ is the sound horizon, and $f\sigma_8(z,\Omega)$ encapsulates the growth rate and amplitude of the clustering. The corresponding equations for these variables are described by Equations 3.3, 1.23 and 3.2.

In the ShapeFit methodology, we extend the Standard approach by incorporating additional parameters, m and n, which capture modifications to the power spectrum. In this case, Θ_Ω is expanded to

$$\Theta_\Omega = \{D_A(z,\Omega)/r_d(\Omega), D_H(z)/r_d(\Omega), f\sigma_{s8}(z,\Omega), m(\Omega), n(\Omega)\}, \tag{3.23}$$

where m, n, and σ_{s8} are respectively given by equations 3.7, 3.8, and 3.9.

where $f\sigma_{s8}$ is the rescaled amplitude and m and n, account for changes in the tilt of the power spectrum. It is important to note that the m parameter, in particular, modifies both, the tilt and the amplitude of the power spectrum, thereby influencing the value of $f\sigma_{s8}$. This deformation is captured by [40]:

$$\begin{aligned} f\sigma_{s8} &= f\frac{(f\sigma_{s8})^{\text{ref}}}{\left(f\left(P_{\text{lin}}\left(k_p\right)\right)^{1/2}\right)^{\text{ref}}}\left(\frac{1}{s^3}P_{\text{lin}}\ \left(k_p/s\right)\right)^{1/2}, \\ &= \frac{(f\sigma_{s8})^{\text{ref}}}{\left(fA_{sp}^{1/2}\right)^{\text{ref}}}(fA_{sp}^{1/2}) \times \exp\left(\frac{m}{2a}\tanh\left(a\ln\left(\frac{r_d^{\text{ref}}}{8}\right)\right)\right), \\ &\approx \frac{(fA_{sp}^{1/2})}{\left(fA_{sp}^{1/2}\right)^{\text{ref}}}\left(f\sigma_{s8}\right)^{\text{ref}}, \end{aligned} \tag{3.24}$$

whith $A_{sp} = s^{-3}P^{\text{lin}}_{\text{no-wiggle}}\ (k_p/s, \boldsymbol{\Theta})$. The final step in eq 3.24 holds because we restrict our MCMC evaluation to the amplitude at the pivot scale.

3.3 Full Shape

The full-shape methodology for constraining cosmological parameters directly compares the theoretical and observed statistics without compressing the information. In practice, we vary the theoretical model for the correlation function, $\xi_\ell^{\text{Model}}(s)$ (or the power spectrum), at different points in the parameter space following a perturbation theory model. Then, we compare these computed models with the measured clustering, $\xi_\ell^{\text{Data}}(s)$. In our analysis, we use the GSM-EFT model 2.1 to generate the multipoles of the galaxy correlation function in redshift space. Note that the following approach is general and can be implemented using any perturbation theory correlation function code—for instance, Velocileptors, folps or pybird can also be used.

To compute the correlation function with GSM-EFT, we require some input parameters. These parameters can be split into two distinct groups. The first group consists of the cosmological parameters: h, ω_b, ω_{cdm}, A_s, n_s, $N_{\rm eff}$, $\Omega_{\rm ncdm}$, these parameters are needed to generate the linear power spectrum using any Boltzmann solver code, here, we use CAMB. The second group comprises nuisance parameters that model the relationship between galaxies and matter as well as the EFT counterterms, including parameters like b_1, b_2, $b_{\nabla^2\delta}$ and b_{s^2}, and $\sigma^2_{\rm EFT}$ and $c_{1,\rm EFT}$.

Clustering measurements are performed using a reference cosmology to convert redshift to distance, which can introduce Alcock-Paczyński distortions. To correct for these distortions when comparing our model multipoles to the data, we introduce late-time re-scaling parameters, $q_{||}$ and $q_\perp$, which correct the calculated clustering parallel and perpendicular to the line of sight. In this way, the separation components in the true cosmology ($s'_{||}$,$s'_\perp$) are expressed in terms of those in the fiducial cosmology ($s_{||}$,$s_\perp$) as follow:

$$s'_{||} = s_{||}q_{||}, \quad s'_\perp = s_\perp q_\perp. \tag{3.25}$$

In this approach, the geometric distortion parameters along and perpendicular to the line of sight differ from those used in the standard methodology, eq. 3.1, because they do not depend on the value of the sound horizon, r_d. They are defined as follows:

$$q_\perp(z) = \frac{D_A(z)}{D_A^{\rm ref}(z)}, \quad q_{||}(z) = \frac{H^{\rm ref}(z)}{H(z)}, \tag{3.26}$$

where the ref superscript denotes the fiducial cosmology where the data multipoles where calculated. As in the compressed parameters, we adopt an alternative parametrization for these distortions, defining,

$$q_\alpha = q_{||}^{1/3}q_\perp^{2/3} \quad q_\epsilon = \left(\frac{q_{||}}{q_\perp}\right)^{1/3} \tag{3.27}$$

and we follow a similar structure as in the compressed parameters, eq. 3.4. We implement the distortions by transforming the separation and angular coordinates: $s \to s'(s_{\rm ref}, \mu_{\rm ref})$ and $\mu \to \mu'(\mu_{\rm ref})$, where the new coordinates are calculated using the re-scaling parameters $\{q_\alpha, q_\epsilon\}$ as:

$$s'(s_{\rm ref}, \mu_{\rm ref}) = s_{\rm ref}\ q_\alpha \sqrt{(1+q_\epsilon)^4 \mu_{\rm ref}^2 + (1-\mu_{\rm ref}^2)(1+q_\epsilon)^{-2}}, \tag{3.28}$$

$$\mu'^2(\mu_{\rm ref}) = \left[1 + \left(\frac{1}{\mu_{\rm ref}^2} - 1\right)(1+q_\epsilon)^{-6}\right]^{-1} \tag{3.29}$$

To apply the dilation parameters, we interpolate each model multipole $\xi_\ell^{\rm model}$ using $s'(s_{\rm ref}, \mu_{\rm ref})$ and $\mu'(\mu_{\rm ref})$. Then, we compute the corresponding observed Legendre polynomials $\mathcal{L}^{\rm obs}(\mu')$.

The observed correlation function $\xi^{\rm obs}(s'(s_{\rm ref}, \mu_{\rm ref}), \mu'(\mu_{\rm ref}))$ is then reconstructed as the sum of multipoles weighted by their respective Legendre polynomials, so that the observed multipoles in the reference cosmology can be expressed as

$$\xi_\ell(s') = \sum_{\ell'} a_{\ell\ell'} \xi_{\ell'}(s), \tag{3.30}$$

When we incorporate the distortion parameters, it causes mixing between multipoles, resulting in a non-diagonal transformation matrix $a_{\ell\ell'}$. Although the sum in principle extends up to $\ell = 8$ for one-loop contributions, the dominant terms are the diagonal ones $a_{\ell\ell}$, and contributions for $\ell = 6$ and $\ell = 8$ are neglegibles compared to those for $\ell = 0, 2, 4$; hence, truncating the sum $\ell = 4$ is an excellent approximation.

As we have stated, our primary objective is to determine the posterior distributions of the cosmological parameters given our data multipoles.

Finally, we determine the posterior distributions of the cosmological parameters using the measured data multipoles by exploring the parameter space with MCMC chains.

3.4 Bayesian Inference

A fundamental aspect of the methodology used here is incorporating a Bayesian framework to constrain model parameters $\boldsymbol{\theta}$. In this framework, the likelihood function $\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta})$ represents the probability of observing the data $\boldsymbol{D}$ given the parameters $\boldsymbol{\theta}$, while $\pi(\boldsymbol{\theta})$ denotes the prior distribution, encoding previous knowledge about the parameters. Assuming a normal distribution for the covariance matrix, the likelihood function takes the form

$$\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta}) \propto \left(\chi^2\right)^{\frac{\nu-2}{2}} \exp\left(-\frac{\chi^2}{2}\right) \tag{3.31}$$

where ν represents the degrees of freedom, and χ^2 quantifies the discrepancy between observed and expected values. The χ^2 statistic, defined as

$$\chi^2 = (\boldsymbol{m} - \boldsymbol{d})^T C^{-1} (\boldsymbol{m} - \boldsymbol{d}), \tag{3.32}$$

measures the deviation between the model predictions $\boldsymbol{m}$ and the observational data $\boldsymbol{d}$, with C representing the covariance matrix of the data.

In this work, the two-point correlation function is estimated from the data using the Landy-Szalay estimator, defined as:

$$\hat{\xi}_{LS}(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \tag{3.33}$$

where:

$$\begin{aligned}
DD(r) &= \frac{dd(r)}{\frac{N_d(N_d-1)}{2}}, && \text{(Normalized data-data pairs)} \\
RR(r) &= \frac{rr(r)}{\frac{N_r(N_r-1)}{2}}, && \text{(Normalized random-random pairs)} \\
DR(r) &= \frac{dr(r)}{N_d \cdot N_r}, && \text{(Normalized data-random pairs)}
\end{aligned} \tag{3.34}$$

Here, $dd(r)$, $rr(r)$, and $dr(r)$ represent the counts of galaxy pairs, random pairs, and data-random pairs at a given separation r, respectively. The normalization factors incorporate the total number of observed galaxies, N_d, and the number of synthetic random points, N_r, which are used to model an unclustered distribution. This is the quantity that we will compare with our theoretical modeling.

3.5 Posterior and Samplers.

We use the bayes theorem to obtain the posterior probability $\mathcal{P}(\boldsymbol{\theta} \mid \boldsymbol{D})$, which depends proportionally on both the prior $\pi(\theta)$ distribution and the likelihood function $\mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta})$. This formulation allows Bayesian inference to systematically incorporate uncertainties and correlations within the data, providing a rigorous probabilistic framework for parameter estimation.

$$\mathcal{P}(\boldsymbol{\theta} \mid \boldsymbol{D}) \propto \mathcal{L}(\boldsymbol{D} \mid \boldsymbol{\theta}) \times \pi(\boldsymbol{\theta}) \tag{3.35}$$

Markov Chain Monte Carlo methods are widely used in Bayesian inference to explore the posterior distribution of model parameters given observed data. Here, I introduce two sampling techniques to efficiently obtain the posterior distribution. These samplers explore the high-dimensional parameter space while ensuring proper convergence and representativity of the posterior.

3.5.1 Zeus.

Zeus is a sampler implemented in Python for the Ensemble Slice Sampling method (ESS), a novel Markov Chain Monte Carlo (MCMC) algorithm intended for efficient Bayesian parameter inference, particularly in cosmological and astronomical applications[45]. This method minimizes the need for hyperparameter tuning and is robust against linear correlations, additionally, it scales effectively to thousands of CPUs without additional overhead and maintains high efficiency even in highly multimodal distributions.

The principal idea of ESS is to introduce an auxiliary variable to facilitate sampling from complex posterior distributions [46]. Traditional slice sampling operates by first drawing a height variable uniformly from the function value at the current sample and then constructing an interval within which a new sample is drawn. The interval is adjusted iteratively until a valid sample is obtained. This way of doing things makes it easier to explore different options and avoid getting stuck in one place.

To extend ESS to multiple walkers and explore the posterior simultaneously, this ensemble structure allows sharing information between chains, improving convergence and sampling efficiency. The walkers' movements are influenced by different updates derived from other walkers' pairs, ensuring that the most efficient movements are informed by the underlying distribution. Some of these movements are available, including Gaussian moves, which sample direction vectors from a normal distribution; global moves, which use Dirichlet Process Gaussian Mixtures to explore multimodal structures; and Kernel Density Estimation moves, which approximate the distribution using density estimation techniques.

A significant advantage of ESS over traditional MCMC methods is its insensitivity to parameter tuning. While standard Metropolis-Hastings algorithms require a meticulous selection of proposal distributions and step sizes, ESS dynamically adapts based on the sampled ensemble. This adaptivity improves performance in scenarios with strong correlations and complex posterior geometries. Moreover, the ability to parallelize ESS across multiple CPUs makes it particularly suitable for large-scale cosmological analyses.

3.5.2 EMCEE.

Traditional MCMC techniques, such as the Metropolis-Hastings (M-H) algorithm, often require careful tuning of proposal distributions and can struggle with high-dimensional or highly anisotropic distributions. emcee is a Python implementation of the affine-invariant ensemble sampler proposed by Goodman and Weare in 2010 that can address these challenges[47].

This algorithm uses an ensemble-based approach, where multiple walkers explore the parameter space simultaneously and communicate between them to adapt their step sizes over the parameter space exploration [48]. It also includes affine invariance, which helps the algorithm to adapt automatically to the geometry of the posterior distribution, improving convergence speed and efficiency. Some advantages include only one or two tuning parameters, parallelization, and faster convergence when using the ensemble sampling method in multiple CPU cores.

By taking advantage of these features, emcee allows us to have an efficient and user-friendly framework for MCMC sampling, making Bayesian inference more accessible and easy to implement.

Chapter 4

Accelerating Full Shape Analysis with Neural Networks

Over the past decade, mapping the large-scale structure (LSS) of the Universe using galaxy surveys has become a key tool in understanding the cosmos. Current projects, such as the Dark Energy Spectroscopic Instrument[1] (DESI), are producing detailed three-dimensional maps by measuring the positions and distances of millions of galaxies. However, this new era of big data demands the development of more advanced methodologies capable of efficiently processing and analyzing such vast amounts of information.

At the same time, sophisticated models have been developed to predict these galaxy clustering patterns. The methods Effective Field Theory [27, 43, 49–52] (EFT) of LSS built on the Perturbation Theory [53] allows us to extract the encoded information. More recently, full-shape or full-modeling analyses have emerged as a real option, and the observed clustering, often converted into statistical measures like the correlation function, has been compared directly with theoretical predictions. However, as the models are built to capture finer details, particularly modeling the smaller scales where nonlinear effects become significant, the computational time required to produce these theoretical models is still very expensive.

Here is where neural networks can help us by learning from a set of pre-computed templates across the parameter space, neural networks can quickly and accurately repro-

---

[1] www.desi.lbl.gov

duce the predictions of complex theoretical models. This machine learning approach has some advantages like, costly calculations and reducing the time needed for each evaluation changing from tens of hours to just minutes when run in parallel. The key advantage is that we can train neural networks to reach the level of precision that meets the high demands of modern LSS surveys while being much less affected by the need for dense sampling of the parameter space compared to other methods like Gaussian process emulators. Nowadays, this approach, where emulators are used to speeding up the computational time are incorporated in the full-shape analysis to constrain cosmological parameters [54–57], where including higher-order statistics [58, 59] and even beyond ΛCDM models [60–64] does not mean a problem in computational time.

Consequently, previous surveys have been reanalized, e.g., the BOSS and eBOSS data using the full-shape methodology. The methods used in these studies can be splited into two main categories: one involves efficient theoretical templates of the power spectrum [e.g. 54–56, 65–67] and the other employs emulator techniques designed to reproduce the theretical models [e.g. 68]. Across all these approaches, there is a clear agreement that these refined methods yield much tighter constraints on cosmological parameters such as the Hubble constant H_0 [69].

Unlike to Fourier space, in configuration space, the number of analyses performed in the literature is still small. It is because some researchers state that Fourier space models are more competitive as they tend to yield tighter constraints. Nevertheless, Configuration space has some advantages, and that is why it should still be used to analyze survey data. Particularly when dealing with the well-localized BAO peak, which in Fourier space becomes distributed across a wide range of wave numbers. Adding to that, some of the observational systematics have different effects in Fourier and configuration space.

We can explore some of these analyses, for example, Chen et al. carried out their main analysis in Fourier space while also computing the correlation function as a consistency check. Conversely, Zhang et al. focused solely on fitting the correlation function using the PyBird code. The key difference with our method is that we start from a Lagrangian framework, allowing us to directly obtain the correlation function without first computing the power spectrum and applying infrared resummations to then Fourier transform the results.

We can find some works like in [66], which primarily worked in Fourier space, later calculating the correlation function to verify the consistency of their results, while [70] focused entirely on fitting the correlation function using the PyBird code where the whole analysis was carried out in fourier space and the, converted into configuration space performing a Fourier transform. In contrast, our methodology works within a Lagrangian framework that enables us to obtain the correlation function directly.

Recent advances in machine learning algorithms, particularly with neural networks, have significantly cut down the time required to evaluate complicated models [e.g. 71]. These algorithms are trained on large datasets of pre-calculated templates spanning various points in the parameter space, allowing them to replicate the behavior of these models accurately. Once properly trained, neural networks can mimic even highly PT models with errors well below the precision thresholds needed by large-scale structure surveys like DESI. Additionally, unlike local interpolation methods such as Gaussian process emulators, neural networks do not rely solely on nearby training points, which means that their prediction errors are less affected by the distance to the nearest data point.

In this chapter we model the redshift space correlation function using a Gaussian Streaming approach up to one-loop perturbation theory. Although our code has been improved to efficiently reduce the time required in each evaluation, when exploring the parameter space using MCMC, the required time to reach the convergence of a chain is still large, and can reach two days. To address this, we developed a neural network emulator that accelerates the production of individual templates, cutting the MCMC runtime from several tens of hours to about 60 minutes, and with a reduction to under 20 minutes when run in parallel on our laptop. All findings in this section were published in [16]

4.1 Neural Network Accelerators

Machine learning has opened up new ways to accelerate complex theoretical models. In particular, neural network accelerators can be trained on large datasets of pre-computed theoretical templates to learn the mapping between cosmological parameters and observables. Given a vector of cosmological parameters, $\boldsymbol{\theta}$, the network approximates the target function $f_{\rm model}(\boldsymbol{\theta})$ by constructing a surrogate model of the form

$$f_{\rm NN}(\boldsymbol{\theta}) = \sum_{i=1}^{N} w_i \, \sigma\left(\mathbf{a}_i \cdot \boldsymbol{\theta} + b_i\right), \tag{4.1}$$

where σ is an activation function, and w_i, $\mathbf{a}_i$, and b_i are the learnable parameters of the network. Once trained, these neural network models generate predictions in a fraction of a second, significantly reducing computational time while maintaining the precision required for full-shape analysis.

In the context of full shape analysis, the parameter space of our model is relatively large, with 7 parameters as the configuration with the fewest parameters. Here, each model evaluation takes approximately 1.5 seconds on our computer. Given that the process requires on the order of 10^5 evaluations, and considering that we are interested in running different settings for MCMC chains, there is a strong motivation to optimize the model evaluation process. This is where neural network accelerators come into play, as they can dramatically reduce the computation time while maintaining the accuracy needed for our analysis.

Various strategies exist in the literature to accelerate the estimation of clustering statistics, and the optimal choice often depends on the available training set size. In our work, the Gaussian Streaming Model described in Section 2.1 offers a cost-effective, nonetheless, it is still needed to speed up the computational time by creating an emulator.

In this chapter, we introduce our emulation methodology, which is an adaptation for the configuration space of the approach presented in [71], which was designed to work efficiently in Fourier space. This approach allows us to emulate the multipoles of the correlation function, reducing the evaluation time of the likelihood to approximately 0.015 seconds, an improvement of two orders of magnitude compared to without an emulator. Our neural network is based on the public code from [71][2], but with some modification to reproduce the multipoles of the correlation funciton.

Here, the objective is to emulate the multipoles of the correlation function by obtaining

---

[2] Which is available at https://github.com/sfschen/EmulateLSS/tree/main

the learnable parameters of the network. The model can be viewed in two step, the first consists of generating a power spectrum template using the publicly available CAMB code [72, 73], alternativelly, we could use CLASS as well, a process that takes roughly one second. The second step, uses the power spectrum as input for our `gsm-eft` code, which computes the correlation function multipoles in an additional half-second. This efficient pipeline allows us to generate training datasets containing tens of thousands of points within a reasonable computational time.

Since our goal is to emulate the first multipoles of the correlation function, we considered two options: The first is building a single neural network that concatenates the multipoles, and the second, constructing separate emulators for each one. The concatenated approach would expand the output layer by a factor of three without adding any extra information, so we opted to implement individual emulators. In this study, we utilize only the first two non-zero multipoles, as we do not include the hexadecapole in our standard pipeline. On our personal laptops, each emulator takes roughly 30 minutes to be trained. Although it might be possible to shorten this training time by fine-tuning certain parameters, we have chosen to keep the current configuration since the training time is not a concern.

Unlike other approaches including Principal components are used to reduce the dimensionality of high-dimensional datasets. Where the input power spectra matrix is decomposed into eigenvectors (which depend on the wave number) and their corresponding eigenvalues (which rely solely on cosmology); this allows for an approximation of the power spectra using a linear combination of only the most significant eigenvalues while discarding the rest, thereby reducing the number of required predictions. Retaining only the most significant components captures the essential information, thereby reducing the number of predictions required. In some works operating in Fourier space [e.g. 71, 74], principal component analysis is employed to predict values for all wave numbers, typically involving hundreds of points, to avoid making hundreds of predictions that would otherwise increase the network's complexity and training time.

In contrast, our approach models the correlation function over a range from 20 to 130 h^{-1} Mpc in 22 bins (each with a width of 5 h^{-1} Mpc) in redshift space, meaning that each emulator is only tasked with predicting 22 numbers. Given that the number of required principal components in Fourier space is comparable to the number of bins we employ, we have found no necessity to incorporate principal component analysis into our methodology.

To build our emulator, we begin by generating 60,000 multipoles using GSM, where we separate them as follows:

- **Training Set:** We utilize 50,000 models to train the neural network, allowing the model to obtain values for the learnable parameters.
- **Validation Set:** 5,000 models are set aside as a validation set. This set is used during training to monitor performance and decide when to adjust the learning rate, avoiding overfitting.
- **Test Set:** 5,000 models are reserved for evaluating the accuracy of the neural network, providing a way to obtain the precision of the methodology's performance.

To distribute points in the parameter space for the three data sets, we employ Korobov sequences [75]. A Korobov sequence is a lattice rule that generates uniformly distributed points within a high-dimensional unit cube. For each distinct redshift we model, we run three separate sequences to form the training, test, and validation sets. We ensure that these sequences are independent and contain no overlapping points by initializing each one with a different seed.

To help the neural network learn more efficiently, we want the output values to be in a consistent range. To achieve this, we apply a hyperbolic sine transformation to each multipole in the training set. This not only scales the values to be more comparable, but also emphasizes important scales of the multipoles, such as the BAO peak.

Our emulator is built using a Multi-Layer Perceptron architecture with four hidden layers and 128 neurons each one. Its predictive accuracy meets our precision needs in a manageable time, so no further architectural tuning was necessary. During training, we progressively reduce the learning rate from 10^{-2} to 10^{-6} in steps of one order of magnitude while doubling the training batch size at each step. Additionally, we employ a custom activation function [e.g. 71, 76],

$$a(X) = \left[\gamma + (1 + e^{-\beta \odot X})^{-1}(1 - \gamma)\right] \odot X, \tag{4.2}$$

which performs better than other more common functions like the Rectified Linear Unit, by introducing free parameters γ and β that are optimized with each hidden layer. This combined approach ensures that our emulator learns effectively and robustly.

During the training, the algorithm reduces the learning rate if no significant improvement in model accuracy, as measured by the mean squared error on the validation set, is observed for a set number of epochs. This number is called patience, for example, in our work we use a patience of 1000 epochs (about 30 minutes of training), which allows the model enough time to potentially escape from local minima and improve its convergence, however, using a shorter patience value, such as 100 epochs (roughly 5 minutes), can speed up training but may compromise accuracy. We carefully monitor the validation set during this process to ensure the model is not overfitting. As drawn in picture 4.1, our neural network implementation consists of three identical networks, each dedicated to modeling one of the correlation function multipoles. Each network follows a consistent architecture beginning with an input layer of 7 nodes that process the initial parameters. The structure then progresses through three hidden layers, each containing 128 nodes and utilizing ReLU activation functions for non-linear transformations. The final output layer employs a linear activation function to produce the correlation function values for its respective multipole. This parallel architecture, replicated three times, allows for independent processing of each multipole while maintaining structural consistency across all three networks, ensuring uniform treatment of the correlation function components.

4.2 Validating our Methodology with High Precision Mocks

In this section, we present our methodology for generating a full-shape EFT-GSM emulator of the correlation function multipoles with the aim of re-analyzing the BOSS data

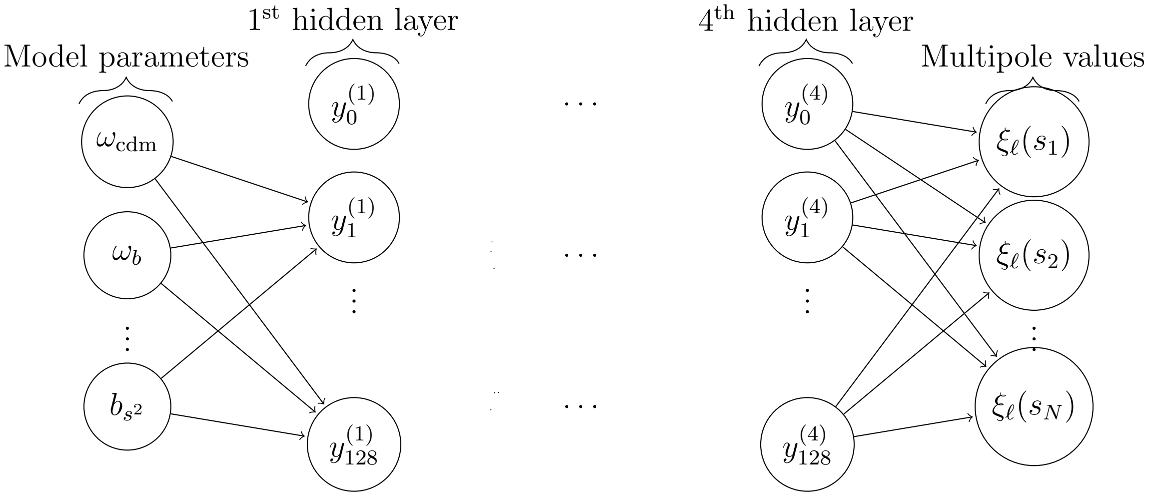

Figure 4.1: Neural network architecture for correlation function multipoles ($\ell = 0, 2, 4$), showing the 7-node input layer, three 128-node hidden layers with ReLU activation, and linear output layer. An identical structure is replicated for each multipole. Figure extracted from [44].

sets. Here, we perform a series of tests using the NSERIES simulations to evaluate the performance of our approach. Specifically, we first assess how accurately and precisely our methodology recovers the simulation parameters using the surrogate model built with neural networks and then, we compare the results against the original EFT-GSM model to validate our approach.

4.2.1 NSERIES Simulations

To test our methodology, we use two complementary simulations to carry out the full shape analysis: one set is dedicated to constructing the covariance matrices for the construction of the likelihood estimations, while the other serves as high-precision mocks for validating our methodology. Here, we describe this set of simulations, which allows us to test our methodology by ensuring we get the optimal statistical errors of our results.

- **NSERIES Mocks:** The NSERIES mocks [6] are a collection of high-resolution N-body simulation, that have been employed in both the BOSS DR12 and eBOSS DR16 analyses. Designed primarily to assess various fitting techniques for theoretical systematics, this suite comprises 84 distinct mock catalogues. These mocks originate from seven independent simulations run over a volume of $(2.6h^{-1}\text{Gpc})^3$ using the `GADGET2` N-body code [77]. Each simulation is further divided by projecting it in seven different orientations and slices, yielding a total of 84 unique datasets. Galaxies are inserted into these simulations following a Halo Occupation Distribution (HOD) prescription that replicates the properties of the CMASS galaxy sample. The cosmology adopted for the NSERIES simulations are: $\Omega_m = 0.286$, $h = 0.7$, $\Omega_b = 0.047$, $\sigma_8 = 0.820$, $A_s \times 10^9 = 2.146$, and $n_s = 0.96$. For our analysis, we utilized the cutsky version of these mocks, which mimic the footprint and number density characteristic of the CMASS north galactic cap at a redshift of $z = 0.55$.

- **MD-Patchy Mocks:** The MultiDark Patchy BOSS DR12 mocks (MD-Patchy mocks) [78, 79] comprise a suite of 1000 simulations designed to construct the covariance matrix for the BOSS data analysis. These mocks are generated based on second-order Lagrangian perturbation theory and incorporate a stochastic halo biasing model, calibrated against high-resolution N-body simulations. Each mock is constructed within a cubic volume of $(2.5h^{-1}\text{Gpc})^3$ and populated with halos using an HOD approach with the goal of reproducing the BOSS samples. The adopted cosmology for MD-Patchy is: $\Omega_m = 0.307115$, $\Omega_\Lambda = 0.692885$, $\Omega_b = 0.048$, $\sigma_8 = 0.8288$, and $h = 0.6777$. These mocks have been designed to replicate both the number density and the survey footprint of the CMASS and LOWZ samples from Data Release 12 and are then divided into three redshift bins.

4.2.2 Testing Neural Networks

We now demonstrate a series of tests designed to evaluate the precision of the neural network approach detailed in section 4.1. First, we assess how well the model predicts the multipoles of the GSM templates from section 2.1 by comparing the neural network's outputs against GSM predictions for 5000 test samples at a redshift of $z = 0.55$. For each test case j, we calculate the percent error using the formula $P_j^{err}(s) = \left|\frac{\xi_j^{GSM}(s) - \xi_j^{NN}(s)}{\xi_j^{M}(s)}\right|$, where $\xi_j^{GSM}(s)$ represents the GSM-predicted multipole, and $\xi_j^{NN}(s)$ is the corresponding value predicted by the neural network. This error metric effectively measures the relative discrepancy between the neural network emulator and the original statistical predictions.

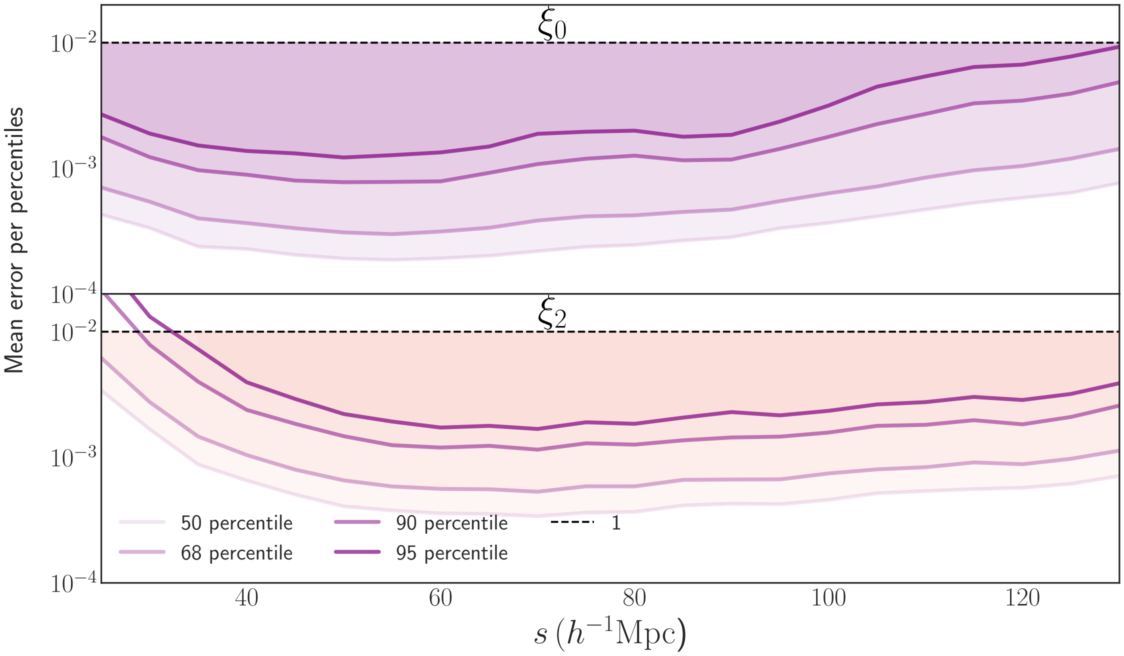

Figure 4.2: Neural network prediction accuracy versus GSM model, showing error percentiles for multipoles. Black line marks 1% threshold, with most of the predictions achieving 0.1% accuracy for the 68% Confidence level. Figure extracted from [44].

To begin with, the data in Figure 4.2 presents percentile plots corresponding to the 50%, 68%, 90%, and 95% error levels, indicating the error threshold below which a given percentage of the 5000 measured errors fall for each separation. Note that almost all of these percentile curves are below the 1% accuracy benchmark. A small deviation is observed for the quadrupole's 90% error at smaller scales. This outcome underscores the neural network's ability to closely replicate the GSM model's multipoles, generally maintaining errors under 1%. Furthermore, the placement of the 68% percentile around a 0.1% error level reveals that most multipole predictions are remarkably precise, with errors around one-tenth of a percent, making larger discrepancies near 1% quite uncommon.

To further validate our methodology, we performed two distinct MCMC fitting procedures on the mean mock of the NSERIES simulations. One approach relies on the GSM model detailed in section 2.1, while the other employs a neural network-based surrogate trained to reproduce the GSM model's behavior at the NSERIES mock redshift. Both

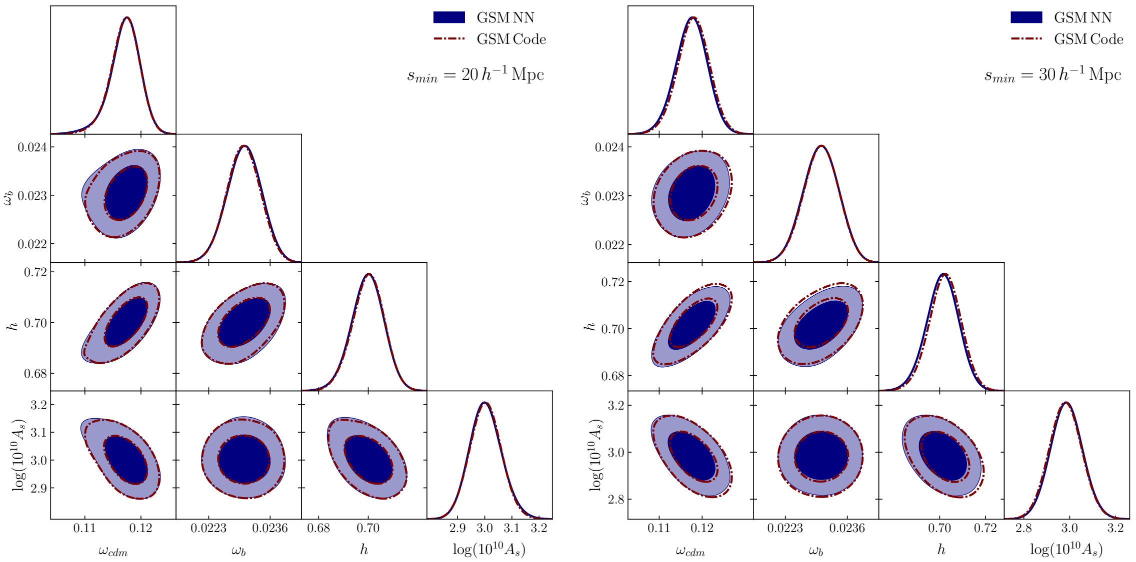

Figure 4.3: Comparison of 1σ (solid lines) and 2σ (shaded regions) posterior contours from the mean of the 84 NSERIES simulations, showing neural network predictions (blue) versus direct GSM calculations (red). Figure extracted from [16].

methods were executed under two different configurations in which the minimum scales is changed: the first follows the standard range from 20 h^{-1} Mpc to 130 h^{-1} Mpc, whereas the second modifies the minimum range, extending it to 30 h^{-1} Mpc.

The triangular plots in Figure 4.3 illustrate a comparison of the likelihood contours for both models. A good similarity is evident in the 1-D histograms, indicating practically identical parameter predictions for the EFT-GSM model and the Neural Network model. We find that for the standard configuration of 20 h^{-1} Mpc, the differences observed in the 2D contours are negligible as expected due to the errors presented in 4.2 and have no significant impact on the best-fit values.

Since our neural network surrogate models can efficiently replicate the data while drastically reducing the convergence time for MCMC chains, all subsequent fits in this chapter are obtained using the Neural Network model.

4.2.3 Testing EFT-GSM Emulator with NSERIES

In this section, we evaluate the effectiveness of our methodology by trying to recover the known free parameters from the N-series simulations described in 4.2.1. Here our modeling based in the GSM consist of free cosmological parameters: h, ω_b, ω_{cdm}, A_s, and the bias and EFT counterterms: b_1, b_2, $b_{\nabla^2\delta}$, and $\sigma^2_{\rm EFT}$. Note that we let $b_{s^2} = 0$ and $c_{1,\rm EFT} = 0$ for our standard configuration, as the effect they model is negligible for the volume we use for the NSERIES simulations.

We present the accuracy and precision of our results where, rather than fitting a single mock, we fit the mean multipole computed from all 84 mocks. This strategy effectively reduces shot noise and variations caused by sampling fluctuations in individual mocks. Therefore, it yields a more robust and reliable estimation of the multipole models.

In our analysis, we derive the error estimates from the one thousand MD-Patchy z_3 simulations, also described in Section 4.2.1. To compute the sample covariance, we use the multipoles from the combined North and South Galactic Cap samples, which corresponds to a volume of $V_{\rm eff} = 4.1\,h^{-3}{\rm Mpc}^3$. To match an effective volume comparable to the expected volumes of next-generation surveys such as DESI, we rescale the covariance matrix by a factor of 1/10 (i.e., $10 \times 4.1\,h^{-3}{\rm Mpc}^3$). This rescaling allows us to assess whether the accuracy of our methodology will be adequate for the demands of upcoming large-scale surveys.

In our analysis, we derive the error estimates from one thousand MD-Patchy z_3 simulations, which are qalso introduced in Section 4.2.1. To compute the sample covariance, we use the multipoles from the combined North and South Galactic Cap sample, which corresponds to an effective volume of $V_{\rm eff} = 4.1\,h^{-3}{\rm Mpc}^3$. To match our methodology against the expected volumes of next-generation surveys such as DESI, we rescale the covariance matrix by a factor of 1/10 (i.e., $10 \times 4.1\,h^{-3}{\rm Mpc}^3$). This rescaling allows us to rigorously assess whether the accuracy of our methodology will be adequate for the demands of upcoming large-scale surveys.

Due to the inherent complexities of modeling small, non-linear scales, it is crucial to determine the redshift-space distance range where our model produces reliable fits. To address this, we simultaneously fit both the monopole and quadrupole of the correlation function over three distinct ranges, each with a different lower limit ($s_{\rm min} = 20$, 30, and $40\,h^{-1}\,{\rm Mpc}$). This strategy allows us to find the scale at which the model's parameter estimates most accurately reflect the true values. Since the information content at larger

Parameter	$s_{\min} = 20\ h^{-1}$ Mpc	$s_{\min} = 30\ h^{-1}$ Mpc	$s_{\min} = 40\ h^{-1}$ Mpc
ω_{cdm}	$0.1172^{+0.0028}_{-0.0023}$	0.1176 ± 0.0036	0.1204 ± 0.0045
ω_b	0.02305 ± 0.00036	0.02305 ± 0.00036	0.02303 ± 0.00036
h	0.6998 ± 0.0064	0.7014 ± 0.0068	0.7036 ± 0.0072
$\ln(10^{10} A_s)$	3.003 ± 0.059	2.986 ± 0.069	2.921 ± 0.089
σ^2_{EFT}	14^{+13}_{-7}	24^{+20}_{-20}	20^{+20}_{-20}
b_1	$1.060^{+0.083}_{-0.070}$	1.093 ± 0.087	1.18 ± 0.12
b_2	$0.17^{+0.97}_{-0.65}$	$0.9^{+1.3}_{-1.0}$	1.6 ± 1.8

Table 4.1: Figure 2: Parameter constraints 68% confidence from MCMC analysis of 84 NSERIES mocks, showing results across different minimum scales for combined monopole and quadrupole fits.

scales is minimal, we fix our upper limit at $s_{\max} = 130\, h^{-1}$ Mpc, a range where our model remains robust, and we consistently apply a distance bin width of $5\, h^{-1}$ Mpc.

Table 4.1 presents our parameter estimates alongside the corresponding uncertainties derived from Markov Chain Monte Carlo (MCMC) analysis. A notable trend is observed: as the minimum scale decreases, the associated errors become progressively smaller. In particular, setting $s_{\min} = 20\, h^{-1}$Mpc yields the tightest constraints on all parameters, while the recovered values remain in agreement with the simulation's cosmology.

To assess the reliability of our model, we compare the mean parameter estimates from our MCMC chains with the true cosmological values from the NSERIES simulations for the four free parameters in our analysis. Figure 4.4 presents a triangular plot summarizing these results. In this plot, the colored histograms represent the 1-D probability distributions of the parameter, whisle the gray lines indicate the reference NSERIES cosmological values.

Our analysis reveals that parameter estimates obtained with a minimum scale of $40\, h^{-1}$Mpc exhibit poorer performance compared to the other two cases. This effect is particularly evident in the histograms, where the distributions for the minimum scale are more closely centered around the reference NSERIES values (gray lines). The reduced

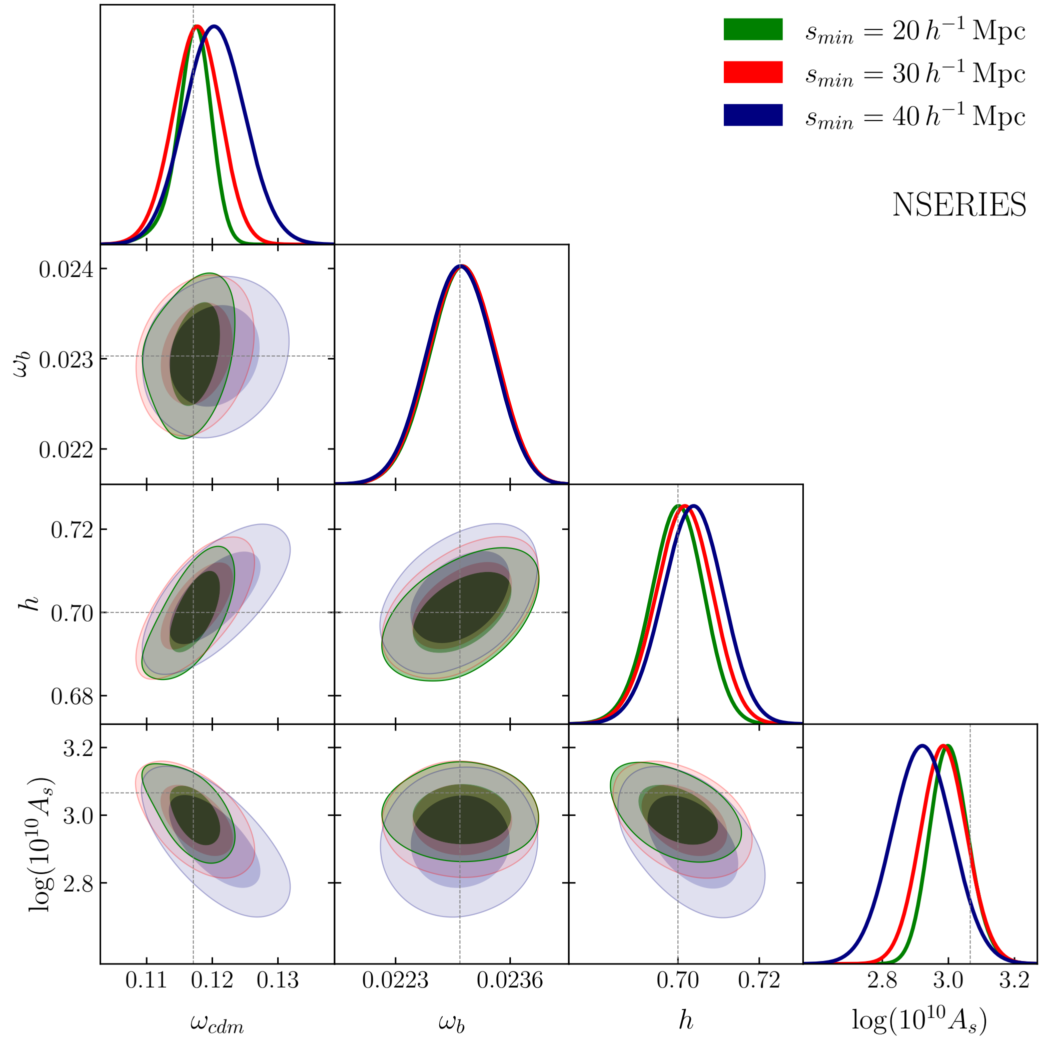

Figure 4.4: Cosmological parameter constraints showing 1σ and 2σ from NSERIES mocks (N=84), showing joint $\ell = 0$ and $\ell = 2$ multipole analysis. Gray lines indicate true simulation values. Figure extracted from [16].

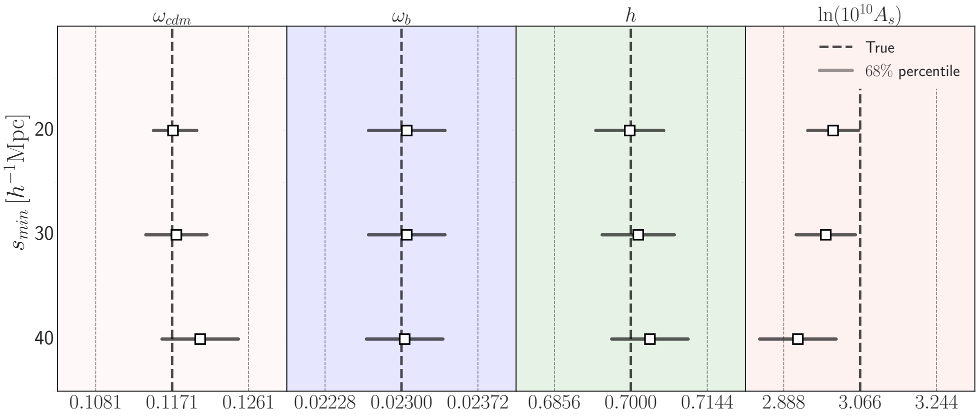

Figure 4.5: In this plot, we compare the mean values obtained from our MCMC chains with the reference cosmology of the NSERIES simulations. The results are shown using square markers, while the true NSERIES values are indicated by dashed lines. Each marker is presented with an error bar representing the 1σ uncertainty from the MCMC chains. The different cases, corresponding to minimum scales of $20\,h^{-1}$Mpc, $30\,h^{-1}$Mpc, and $40\,h^{-1}$Mpc, are are displayed from top to bottom. Figure extracted from [16].

accuracy at larger minimum scales can be explained by the exclusion of smaller-scale bins, which typically have smaller uncertainties. As a result, omitting these bins leads to a decrease in the overall constraining power of the model.

The figure 4.4 displays the 1σ and 2σ confidence regions as colored contours. Notably, for the $s_{\mathrm{min}} = 20\,h^{-1}$Mpc case, the true NSERIES cosmology aligns well within the 1σ region, as indicated by the overlap of the gray reference lines with the solid green contours. To present these findings more clearly, Figure 4.5 summarizes the deviations of the estimated parameters from the true values. The plot highlights that all three models successfully recover ω_{cdm}, ω_b, and h within 1σ. However, the model with $s_{\mathrm{min}} = 40\,h^{-1}$Mpc shows greater deviations from the true values of ω_{cdm} and h. Additionally, only the $s_{\mathrm{min}} = 20\,h^{-1}$Mpc model provides an estimate of A_s that remains within 1σ of the NSERIES cosmology.

Moreover, constraints on ω_{cdm} and A_s are notably tighter in the $s_{\mathrm{min}} = 20\,h^{-1}$Mpc model compared to the $s_{\mathrm{min}} = 30\,h^{-1}$Mpc case. Given its better accuracy and precision, we adopt $s_{\mathrm{min}} = 20\,h^{-1}$Mpc as the fixed minimum scale for the subsequent analysis.

4.3 Results with SDSS-III BOSS Catalogues

The previous section demonstrated that the EFT-GSM model effectively recovers the cosmological parameters from NSERIES simulations while confirming that our surrogate models can reliably replicate the results of the EFT-GSM code. Building on this validation, the next step is to implement our approach on real galaxy data. In the following analysis, we apply our methodology to the BOSS DR12 LRG correlation function to derive constraints on cosmological parameters, assessing the robustness of our surrogate model in practical observational scenarios.

4.3.1 Data

For this analysis, we use the publicly available data from the Sloan Baryon Oscillation Spectroscopic Survey (BOSS) [5], which is a key component of the Sloan Digital Sky Survey III [SDSS-III; 80]. For this study, we specifically use the galaxy catalogs from Data Release 12 [81], which were compiled using observations from the 2.5-meter telescope located at Apache Point Observatory in New Mexico, USA [82]. Spectroscopic measurements were carried out with the multi-object spectrographs [83], ensuring precise data collection. A comprehensive discussion of the data reduction process and discussions of the data are available in [84].

Originally, the BOSS target selection was designed to recollect data from two distinct galaxy samples, with the LOWZ sample focusing on luminous red galaxies at redshifts below 0.4 in redshift and the CMASS sample capturing massive galaxies within the 0.4 to 0.7 redshift range. Subsequently, researchers combined these datasets into three overlapping bins, labeled z_1, z_2, and z_3—to maximize constraints on dark energy parameters as noted in [6, 85]. Furthermore, the construction of the catalogue, which covers essential aspects like masks, completeness, and weights, is elaborated following [6, 17, 86]. Here, table 4.2 descreibes the main characteristics of these samples.

For our analysis, we concentrates on the low and high redshift bins (z_1 and z_3) which remain distinct in redshift and share a comparable effective volume (V_{eff}). Here, the

Name	z-range	z_{eff}	N_{gal}	V_{eff}	V
LOWZ	$0.15 < z < 0.43$	0.32	361,762	2.87	3.7
CMASS	$0.43 < z < 0.70$	0.57	777,202	7.14	10.8
BOSS z_1	$0.20 < z < 0.50$	0.38	604,001	3.7	6.4
BOSS z_2	$0.40 < z < 0.60$	0.51	686,370	4.2	7.3
BOSS z_3	$0.50 < z < 0.75$	0.61	594,003	4.1	12.3

Table 4.2: This study follows the sample characteristics presented in the table, the effective volume incorporates a parameter of $P_0 = 10,000\,h^{-3}\text{Mpc}^3$, ensuring that the statistical framework is well-defined. Additionally, a fiducial cosmology with $\Omega_m = 0.310$ and $h = 0.676$ is adopted to maintain consistency across analyses. Table extracted from [16].

effective volume is calculated using the formula

$$V_{\text{eff}} = \sum_i \left(\frac{\bar{n}(z_i) P_0}{1 + \bar{n}(z_i) P_0} \right)^2 \Delta V(z_i) \tag{4.3}$$

where $\Delta V(z_i)$ denotes the volume of the shell at redshift z_i and the parameter $P_0 = 10,000 h^{-3}\text{Mpc}^3$ is chosen because it aligns with the amplitude of the power spectrum that maximizes the BAO signal [17, 87].

4.3.2 Baseline Analysis

Employing a Neural Network emulator, our baseline methodology fits the free cosmological parameters h, ω_b, ω_{cdm}, and A_s alongside the bias terms and EFT counterterms b_1, b_2, $b_{\nabla^2\delta}$, and σ^2_{EFT}. In this analysis, we set $b_{s^2} = 0$ and $c_{1,\text{EFT}} = 0$ over a scale ranging from $20\,h^{-1}$Mpc to $130\,h^{-1}$Mpc, thereby ensuring a consistent framework. This configuration serves as the foundation for deriving constraints on the cosmological parameters, where three unique fits were computed by employing different combinations of the BOSS samples described in Section 4.3.1. Specifically, the first two fits analyzed the monopole and quadrupole moments of the z_1 and z_3 datasets, respectively. Alternatively, a third fit combined the multipoles from both z_1 and z_3 into a model designated as $z_1 + z_3$. For this analysis, the covariance matrix was derived using the MD-Patchy mocks introduced in Section 4.2.1, which allows us to get the required precision of our statistical analysis.

Parameter	z_1	z_3	$z_1 + z_3$
ω_{cdm}	0.1038 ± 0.0064	0.1238 ± 0.0076	0.1115 ± 0.0050
ω_b	0.02237 ± 0.00037	0.02236 ± 0.00037	0.02237 ± 0.00037
h	0.673 ± 0.017	0.705 ± 0.017	0.688 ± 0.012
$\ln(10^{10} A_s)$	3.29 ± 0.17	$2.69^{+0.18}_{-0.20}$	3.03 ± 0.13
Ω_m	0.280 ± 0.012	0.296 ± 0.015	0.2846 ± 0.0093
$10^9 A_s$	$2.71^{+0.42}_{-0.49}$	$1.49^{+0.23}_{-0.31}$	$2.09^{+0.25}_{-0.29}$

Table 4.3: Our MCMC approach yielded estimated means and corresponding 1σ uncertainties for three distinct BOSS samples: z_1, z_3, and the combined $z_1 + z_3$. Utilizing both the monopole and quadrupole moments of the correlation function, throughout the fitting process, we capture a comprehensive picture of the clustering behavior. In addition, error covariance estimates were obtained using the MD-Patchy mocks presented in this chapter. Table extracted from [16].

In Table 4.3, the limits used for our 4 distinct cosmological parameters are summarized. It shows that the error margins associated with z_1 and z_3 are almost the same, and a better precision is observed for the combined parameter $z_1 + z_3$, while concurrently, the estimates for h and A_s exhibit uncertainties approximately 25% lower than those determined individually for z_1 and z_3. Additionally, the margin of error for ω_{cdm} is reduced by nearly 33% compared to that of z_3.

Figure 4.6 displays a triangular plot of the MCMC fits for three BOSS datasets, where almost every parameter aligns within the 1σ level except A_s, which agrees at the 2σ between z_1 and z_3. It is evident that discrepancies between the redshift bins z_1 and z_3 have been observed before, as documented in previous studies [54, 66], especially in full-shape correlation function analyses. We notice that the dataset corresponding to z_1 predicts lower values for both ω_{cdm} and h, whereas z_3 suggests a diminished estimate for A_s. On the other hand, the combined fit $z_1 + z_3$ yields intermediate parameter values relative to the individual bins as expected. Moreover, the predictions for ω_b remain indistinguishable across all samples, given that its constraints are determined by the prior. Finally, further exploration is conducted in Section 4.3.4, where an expanded prior is employed to evaluate

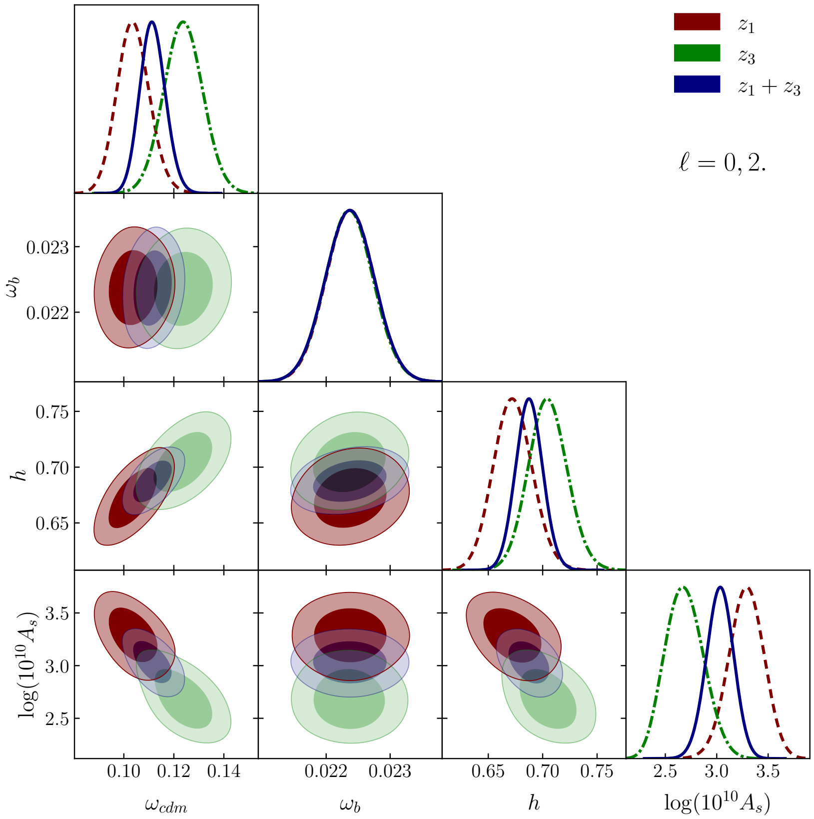

Figure 4.6: A triangular plot that represents the MCMC fits applied to the BOSS data sets for z_1 and z_3 along with an additional chain where both parameters are fitted simultaneously, with distinct colors matching those detailed in the figure labels. Beginning with the contours, the shaded areas illustrate the 1σ and 2σ uncertainty regions. On the other hand, the histograms on the margins reveal the one-dimensional distributions for each parameter. Figure extracted from [16].

the capacity of large-scale structure data to independently constrain cosmological parameters and to test the methodology within a more extended parameter space, adding an original perspective to our analysis.

4.3.3 Comparison to Some Alternative Full-Shape Analyses

In recent years, several research teams have re-analysed the BOSS dataset using a full-shape methodology, although focusing primarily on Fourier space analyses. Our study diverges by working within the configuration space, aiming to obtain consistency between different spaces. Furthermore, parameter estimations derived from our configuration space model are compared against Fourier space results reported by (D'Amico [55], Ivanov [54], Philcox [56], Troster [65], and Chen [66]), while also being contrasted with the configuration space findings from [70]. To establish an equitable comparison, only Zhang constraints that rely solely on BOSS data are taken into account. Additionally, an alternative full-shape strategy involving an extended compression methodology is compared, where the parameter introduces a few extra free parameters that capture the power spectrum slope. Called Shapefit, this approach, introduced by Brieden [37] employs this approach within Fourier space, and our work further compares its parameter estimations with those obtained using our method.

So far, the analyses mentioned before have been performed using the BOSS DR12 dataset, with a predominant focus on the redshift bins z_1 and z_3 samples that we have also employed. Although one study by D'Amico [55] chosed to work with the LOWZ and CMASS samples. Because all these analysis use the same dataset using largely overlapping sample selections, it is expected that the parameter estimates will align with one another.

Figure 4.7 illustrates the parameter estimations from these methods using square markers, while the first column showcases our own estimates denoted by starred markers. Different parameters are analyzed to include H_0, A_s, and the total mass density Ω_{m} quantity that sums the energy density from dark matter, baryons, and all other matter sources. We focus on these three parameters, as they facilitate direct benchmarking with the outcomes of other studies, even though only two of those works report A_s. Accordingly, our derived

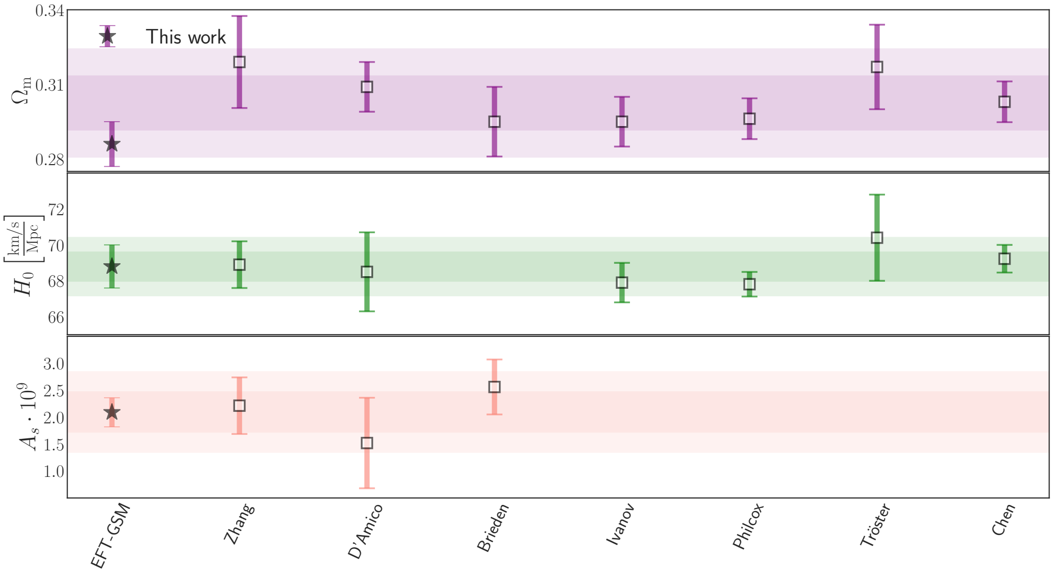

Figure 4.7: Our analysis presents parameter estimates from BOSS $z_1 + z_3$ data using our method, marked by starred dots, and these are contrasted with results from five independent studies that implemented a full shape analysis in both Fourier and configuration space alongside predictions from the Shapefit methodology. These external results are referenced as square dots. The uncertainties are represented by error bars based on 1 σ estimates generated via MCMC analysis in each individual study. Subsequently, visual stripes indicate the regions corresponding to one and two standard deviations from the mean across the different panels. Figure extracted from [16].

results, which encompass these results, are explicitly detailed in Table 4.3.

Figure 4.7 illustrates that our estimates for A_s and H_0 align within 1σ with those obtained in other studies, while our Ω_m predictions are also within 1σ for nearly all cases, except for D'Amico, Zhang, and Tröster, where the agreement is within 2σ. It is important to note that D'Amico employs LOWZ and CMASS samples rather than the z_1 and z_3 samples used in our analysis, which, due to different redshift ranges and BOSS galaxy subsets, is expected to contribute to the observed discrepancies. Additionally, Tröster applies a broad prior on n_s, a parameter that remains fixed in our standard methodology, thereby affecting the fit results for both Ω_m and H_0, and leading to slight disagreements. Moreover, Zhang varies two extra parameters, namely n_s and Σm_ν, modifications that account for the shifts in error bars and mean positions.

In our analysis, the precision of our model is similar to most of the previous studies, yet notable exceptions include Philcox [56] and Chen [66], whose constraints are tighter than ours. Alternatively, both of these studies benefit from incorporating additional geometrical information derived from BAO post-reconstruction in both Fourier and configuration space, which effectively narrows their parameter uncertainties. Conversely, Tröster [65] reports slightly broader constraints due to the adoption of wider priors on ω_b, a choice that naturally leads to less stringent limits.

In summary, our EFT-GSM shows a good level of consistency with other full-shape analyses, as differences are confined to the 1-2σ range, specifically, approximately (1.7σ D'Amico, 1.6σ Troster and Zhang). Alternatively, these discrepancies can largely be explained by variations in the utilized galaxy samples, the number of free cosmological parameters involved, and the specific priors applied in each study, in addition to the PT model employed. Moreover, our findings suggest that the EFT-GSM framework is competitive against the Fourier space models but has the advantage of working in the configuration space, where direct physical interpretation can be undertaken.

4.3.4 Extensions to Baseline analysis

Our EFT-GSM approach performance, as shown in the previous sections, recovers the cosmological parameters from NSERIES simulations under error conditions anticipated by upcoming surveys like DESI. Tests on BOSS data further reinforced our method's accuracy, as our findings closely mirrored those of other full-shape analyses. Subsequently, we are now branching out by exploring various model configurations. Next, our following steps focus on refining constraints on cosmology by adjusting the priors for two critical parameters, n_s, and ω_b, a step that not only tests the robustness of our approach but also provides fresh insights into how prior assumptions shape outcomes.

Clustering analyses of large-scale structures often incorporate external observables to constrain cosmological parameters that remain hard to fix with the perturbation theory model. In our baseline study, we decided to keep n_s fixed at the value detailed in Table 5.1, a value originally established through CMB observations. Additionally, we set tight priors on ω_b by utilizing measurements of the deuterium-to-hydrogen ratio obtained from quasar absorption systems. Using a well-determined reaction cross-section between deuterium and helium-3 allowed us to impose stringent constraints on ω_b. Commonly, these constraints are referred to as Big Bang Nucleosynthesis (BBN) priors throughout our work and in most of the literature.

In this extended analysis, we lower the constraining power of our priors on two crucial cosmological parameters;

$$\omega_b :, \mathcal{N}[0.02237, 0.00037] \tag{4.4}$$

$$n_s :, \mathcal{U}[0.5, 1.5] \tag{4.5}$$

As illustrated in 4.4 and Figure 4.9, the analysis reveals that altering the prior range on baryon density notably affects the uncertainty of several cosmological parameters. Beyond that, the contrast between the conventional BBN priors (depicted in yellow) and the 10× BBN priors versions (shown in green) indicates a substantial increase in error margins, particularly for the Hubble parameter and the cold dark matter density. Furthermore, the data underscore that while the precision declines, errors escalate by roughly 2 and 1.6 times for h and ω_{cdm}, respectively. The central estimates of the parameters remain virtually unchanged.

Research by Tröster [65], which employed priors approximately ten times larger than standard BBN values, uncovered broader posterior distributions compared to other analyses of the BOSS DR12 dataset. Initially, his work suggested that these expanded priors are a key factor leading to lower precision in parameter estimation. Which is consistent with our work, as we can notice in figure 4.7.

Recent investigations, as presented by Ivanov [54] indicate that modifying the priors on ω_b can substantially influence cosmological parameter constraints, where a broader prior weakened the bounds on h while only modestly affecting Ω_m. Interestingly, further analysis by Brieden [39] reveals that in a Full Shape framework, the precision of BAO amplitude measurements is intricately linked to the ω_b/ω_{cdm} ratio, indicating that tighter priors on ω_b are consequently narrowing the ω_{cdm} posterior. Notably, the interplay between these priors and the full shape fits implies that in a regime dominated by priors, even small shifts can impact overall parameter accuracy, offering fresh insights into the sensitivity of cosmological inferences.

Observations in configuration space multipoles reveal that varying the baryon density parameter, ω_b, extend their influence well beyond adjustments to the BAO peak's shape or location as shown in Figure 4.8. These shifts suggest that ω_b variations produce complex alterations throughout the multipole structure, highlighting dependencies.

Our exploration continues by relaxing n_s from its Planck-fixed constraint. Comparing the distinct data contours, one where n_s remains fixed and another with a flat prior, reveals that ω_{cdm}'s uncertainty shrinks considerably when n_s is stabilized, whereas the impact on h is relatively modest, as noted in Figure 4.9. Observations derived from table 4.4 indicate that allowing n_s to vary results in uncertainties for ω_{cdm} that are roughly doubled and for h that are about 1.25 times larger, with the scalar amplitude A_s suffering an error increase of approximately 1.4 times. Additionally, our results show that the posterior distributions for both ω_{cdm} and h tend to shift upward when n_s is free, though these shifts remain within the acceptable 1σ interval, suggesting a coupling between the underlying shape of the power spectrum and the amplitude of matter fluctuations.

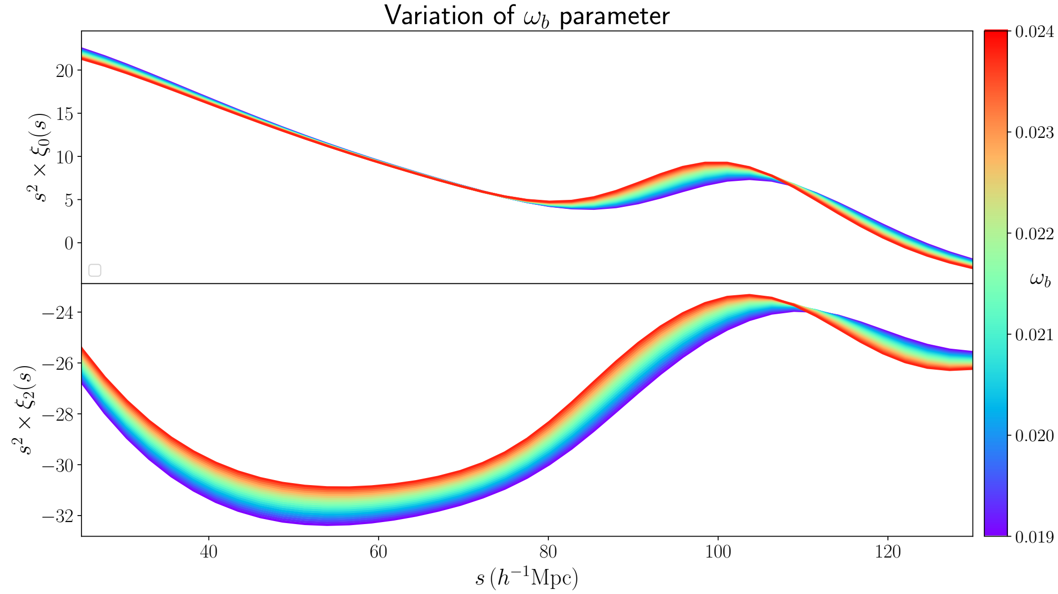

Figure 4.8: plot illustrates multipoles computed using our EFT GSM model to showcase the impact of varying ω_m in a standard setup. Figure extracted from [16].

Parameter	ω_b–10σ, n_s Free	ω_b–BBN, n_s Free	ω_b–10σ, n_s Fixed	ω_b–BBN, n_s Fixed
ω_{cdm}	0.121 ± 0.016	$0.118^{+0.010}_{-0.013}$	0.1119 ± 0.0083	0.1115 ± 0.0050
ω_b	0.0235 ± 0.0034	0.02237 ± 0.00037	0.0227 ± 0.0031	0.02237 ± 0.00037
h	0.701 ± 0.035	$0.692^{+0.014}_{-0.016}$	0.689 ± 0.029	0.688 ± 0.012
$\ln(10^{10}A_s)$	2.95 ± 0.19	2.96 ± 0.18	3.03 ± 0.13	3.03 ± 0.13
n_s	0.925 ± 0.070	$0.930^{+0.070}_{-0.063}$	0.97	0.97

Table 4.4: A 68% confidence level for the monopole and quadrupole fitting the combined $z_1 + z_3$ sample. Initially, the analysis incorporated a Big Bang nucleosynthesis (BBN) prior on ω_b, yet the method was extended by alternatively relaxing this prior to a scenario where its uncertainty followed the standard BBN value, $\sigma = 10 \times \sigma_{\mathrm{BBN}}$, denoted as 10σ. Additionally, the spectral index n_s was allowed to vary. Table extracted from [16].

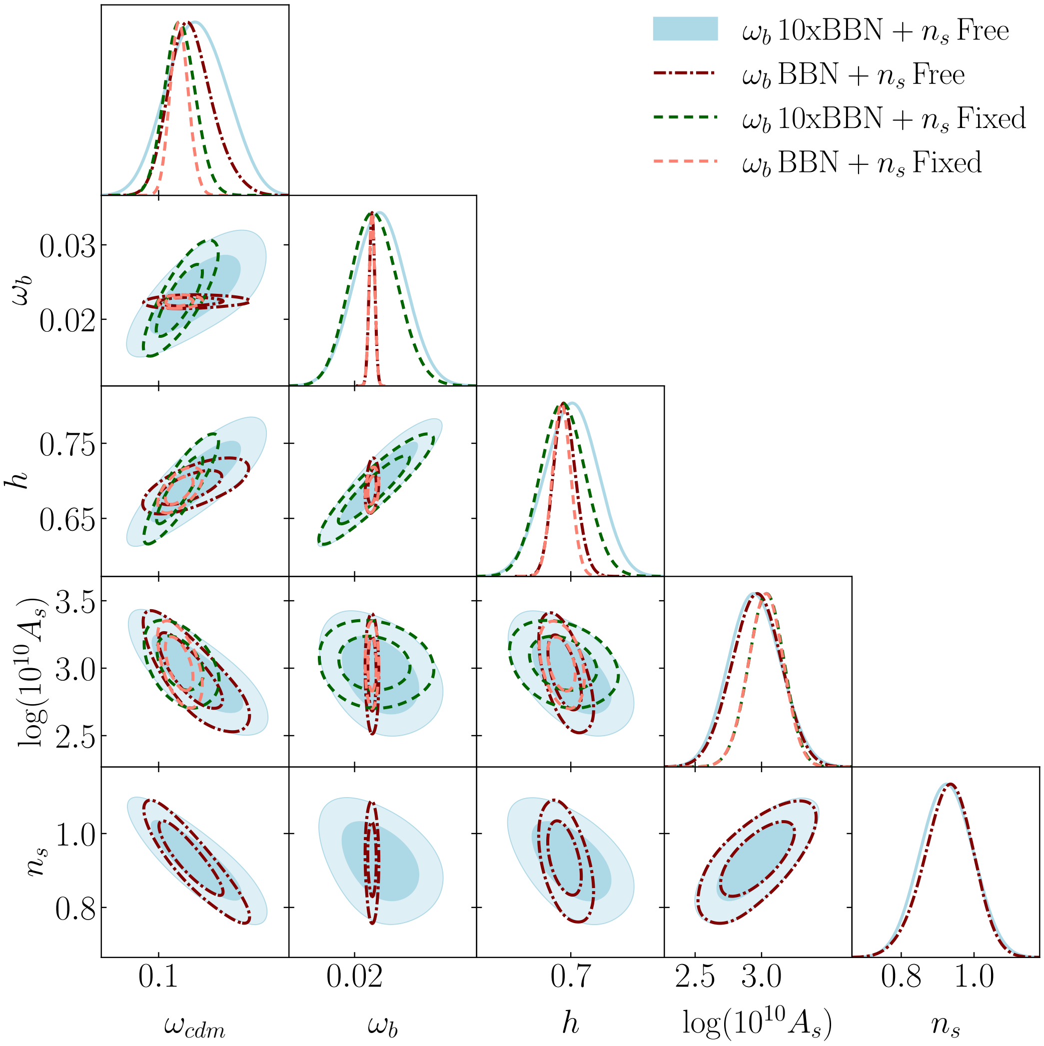

Figure 4.9: The triangular plot illustrates the influence of different prior settings on our GSM EFT full shape analysis applied to the BOSS $z_1 + z_3$ dataset. Our standard configuration, depicted in yellow, serves as the baseline from which all other modifications are compared. Furthermore, distinct adjustments to the priors reveal that wider bounds for ω_b or n_s individually lead to contour shifts shown in green and red, respectively. In contrast, simultaneous expansion of both priors, outlined in blue, illustrates an effect on the likelihood distribution. Figure extracted from [16].

4.3.5 Exploring the Information Content of Multipoles

Initially, we examine how multipole measurements can constrain the information about model parameters. Next, a new Markov Chain Monte Carlo analysis was carried out on the $z_1 + z_3$ dataset, focusing only on the monopole component in line with our standard methodological setup. Moreover, instead of incorporating several multipole orders as in the traditional strategy, this approach isolates the monopole to determine its constraining power. Figure 4.10 shows a visual comparison contrasting the outcomes, indicated by red dashed trends, with those of the comprehensive method that employs both monopole and quadrupole features.

Displayed in a summary table 4.5, the results for both scenarios reveal that isolating the monopole component enables a precise measurement of the key cosmological parameters, specifically, the ω_{cdm} and h with an accuracy almost indistinguishable from methods that incorporate quadrupole data($\Delta\omega_{cdm} = 0.0006$, and $\Delta h = 0.001$) and precision ($\Delta\sigma_{\omega_{cdm}} = 0.0005$, and $\Delta\sigma_h < 0.001$). Note that, the main cosmological signals are encapsulated within the monopole. Additionally, the amplitude parameter, A_s, is determined with considerably less precision.

Redshift-space distortions modify the amplitude of the quadrupole relative to the monopole on large scales, thereby breaking the usual degeneracy inherent in the parameter $\beta \equiv f/b_1$. Alternatively, the strong correlation between the amplitude parameter A_s and large-scale bias means that incorporating quadrupole data sharpens the estimation of A_s. Theoretically, these outcomes are well anticipated and align with existing models. In addition, the quadrupole keeps information about baryon acoustic oscillation scales, which one might expect to slightly refine estimates of parameters like ω_{cdm} and h. Surprisingly, our findings reveal that this anticipated improvement in ω_{cdm} and h does not manifest, challenging our initial expectations.

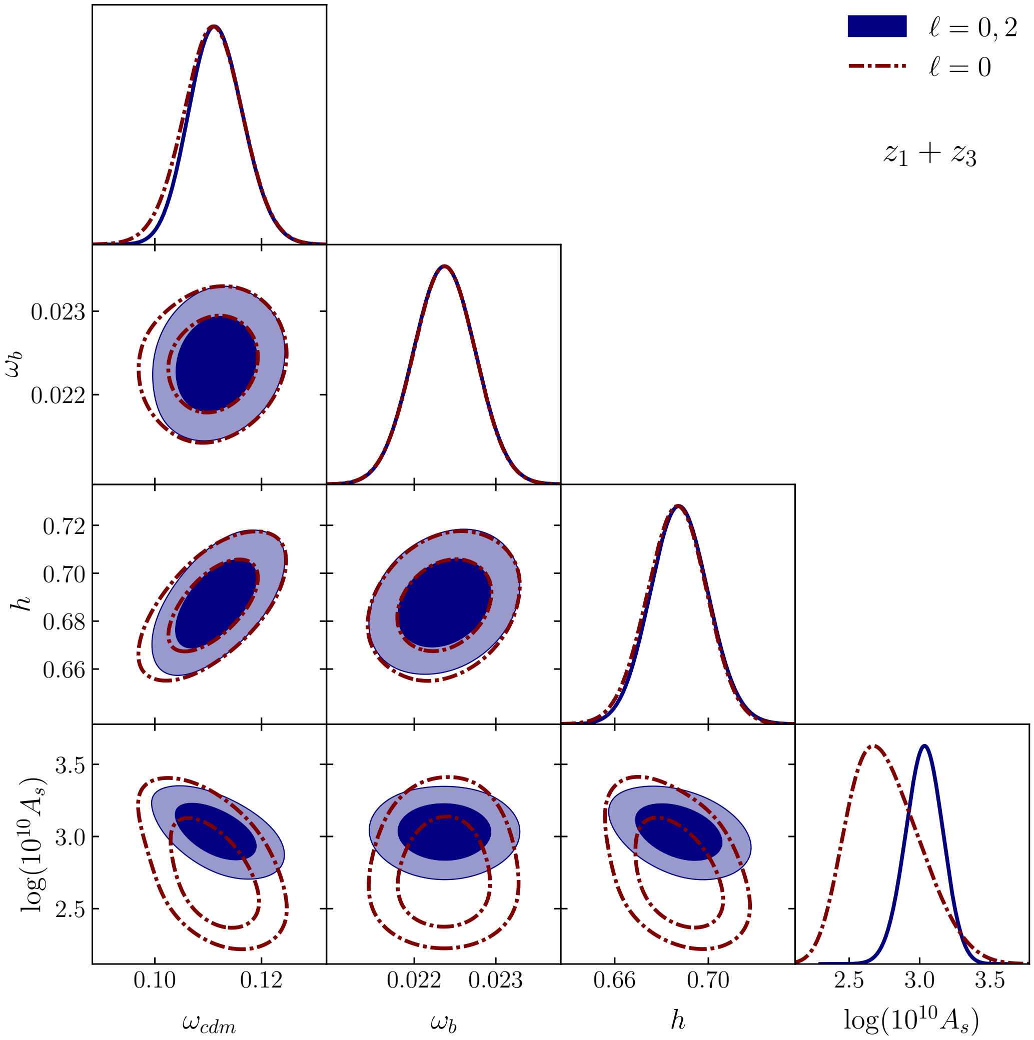

Figure 4.10: The triangle plot graph shows how two different configuration settings for the BOSS dataset $z_1 + z_3$. The diagram contains two distinct color patterns: blue represents the combination of both monopole and quadrupole correlation multipoles, while the red pattern shows only the monopole. Figure extracted from [16].

BOSS Combined z1+z3

	ω_{cdm}	ω_b	h	$\ln(10^{10}A_s)$
ξ_0	0.1109 ± 0.0055	0.02237 ± 0.00037	0.687 ± 0.012	$2.75^{+0.21}_{-0.29}$
$\xi_0 + \xi_2$	0.1115 ± 0.0050	0.02237 ± 0.00037	0.688 ± 0.012	3.03 ± 0.13

Table 4.5: Summary of one sigma contours for our baseline configuration and for the same inconfiguration, excluding the quadrupole. Table extracted from [16].

.

Chapter 5

Full Modeling and Parameter Compression Methods in configuration space for DESI 2024 and beyond

Over the past decade, spectroscopic galaxy surveys have dramatically enhanced our understanding of cosmic structure by mapping the spatial distribution of galaxies with increasing precision. Projects such as 2dF Galaxy Redshift Survey (2dFGRS) [3, 4], the Baryon Oscillation Spectroscopic Survey (BOSS) [5, 6], the extended Baryon Oscillation Spectroscopic Survey (eBOSS) [7, 8], or the Dark Energy Survey (DES) [9, 10] have progressively expanded the volume of observed data, allowing for more refined cosmological analyses. As technology advances, next-generation surveys like *Euclid* mission [11, 12, 88] and The Rubin Observatory Large Synoptic Survey Telescope (LSST) [13] are designed to further increase the data volume. Meanwhile, the Dark Energy Spectroscopic Instrument (DESI) [1] [89–92] is actively conducting one of the most extensive galaxy-mapping efforts ever done, covering an unprecedented sky area $14,000\,\mathrm{deg}^2$ in a volume of $40\,h^{-3}\,\mathrm{Gpc}^3$ and increasing the number of cataloged galaxies. Looking ahead, DESI-II [93] is set to

---

[1] www.desi.lbl.gov

push these observations even deeper into the high-redshift universe $z > 2$, promising new insights into cosmic evolution with observation starting in 2026.

By observing distinct tracers of matter across different redshift ranges, DESI provides aa big amoung of data that help us to study the structure formation. Emission line galaxies (ELG) [94] trace star-forming regions at intermediate redshifts ($0.6 < z < 1.6$), while luminous red galaxies (LRG) [95] at lower redshifts($0.3 < z < 1.0$). Quasars (QSO), the bright centers of active galaxies, extend the reach of the survey to even greater distances ($0.9 < z < 2.1$) [96]. The unprecedented precision of these measurements demands highly accurate theoretical models to interpret clustering patterns and extract cosmological information.

Cosmological modeling requires innovative techniques that compare theoretical predictions with observational data. On the top of our modeling lies Perturbation theory (PT) in the two approaches, in the Lagrangian approach [97–103] and in the Eulerian approach [104–107]. Both frameworks emerge as powerful tools for studying the complex dynamics of large-scale cosmic structures. More recently, effective field theory (EFT) has been included to accouint for the quasi-linear regime of cosmic structure formation. In particular, our modeling of the Gaussian streaming model works in the PT (LPT)/EFT approach [19, 22, 25–27].

As the DESI Year 1 data release approached, the collaboration focused on modeling systematic effects using a new simulation suite, `ABACUS-SUMMIT` [108], which meets the current requirements for resolution and volume. To analyze redshift-space clustering statistics, four different modeling methods were tested.

One method uses a configuration space model known as EFT-GSM [35], which forms the core of this work. Another approach relies on three Fourier space models: PyBird [30, 109–111], FOLPSν [33, 41] and velocileptors [32, 112, 113], models described in the previous section. Finally, earlier comparisons of the Fourier models [31] help guide this paper as it looks at both the configuration space and Fourier space techniques.

Motivated by the then-upcoming DESI first-year release, this study examines potential systematic uncertainties in our modelling and fitting methodology. Developed within a Lagrangian Perturbation Theory framework, the EFT-GSM model captures redshift space correlation functions up to one-loop corrections by employing the Gaussian Streaming Model alongside additional EFT parameters, where we let two EFT parameters vary: $CLF1_{EFT}$ and bs^2. Initiated to handle nonlinear effects on scales as small as $20\,h^{-1}\,\mathrm{Mpc}$, this approach is enhanced with neural networks, as detailed in the previous section. Notably, our methodology diverges from the other three by fitting the correlation function directly, whereas alternative methods operate in Fourier space (see also [111], which fits the correlation function, through Fourier transforming the power spectrum calculated with PyBird.)

As seen in the previous chapter, different techniques have been developed to extract cosmological information from the shape of two-point statistics, namely the compressed methods (as the *Standard* and *ShapeFit* methodologies) and the direct fits (also called *Full Modeling*). In this context, the *Standard* approach, which has been applied for the clustering of redshift space tracers in SDSS-III BOSS [5, 80, 81] and SDSS-IV eBOSS [114–116], starts with the one-time computation of the linear power spectrum within a chosen fiducial cosmology. Then, this power spectrum is integrated into a PT framework that compresses the key cosmological details into three free parameters: the linear growth rate of structures $f\sigma_8$ and two scaling parameters, $\alpha_\perp$ and $\alpha_\parallel$, which account for the Alcock-Paczynski effect [117].

Alternatively, the *ShapeFit* methodology of [40] builds on the *Standard* compressed approach by incorporating an extra free parameter, m, which adjusts the slope of the linear power spectrum to capture the transition from large to small scales. Independently, this additional parameter enriches the compressed framework originally used for BOSS and eBOSS without sacrificing model independence. Furthermore, the inherent flexibility of methods like the *Standard* and *ShapeFit* approaches makes them notably less dependent on specific model assumptions compared to those derived from *Full Modeling* techniques. Additionally, their lower computational demands offer a practical advantage over more intensive analyses. Ultimately, though the *ShapeFit* procedure has been examined primarily in Fourier space [e.g. 40], its effectiveness when applied in configuration space remains a topic of ongoing research.

In contrast to the compressed nature of the *Standard* and *ShapeFit* methodologies, one can chosse to explore the full posterior distribution of all cosmological parameters. Particularly, in this work, we designed a neural network following the procedure of the previous section to tackle this problem, consequently, it is now feasible to model the full shape of the correlation function or power spectrum while simultaneously varying cosmological, bias, and EFT (Effective Field Theory) parameters.

This section aims to assess how three techniques, *Standard*, *ShapeFit*, and *Full Modeling*, extract cosmological information. Subsequently, we compare the outcomes of these approaches under a diverse set of conditions. Moreover, our analysis spans different DESI tracers to ensure a broad performance evaluation for DESI standards. All findings here were published in [44]

5.1 Abacuss Simulations

As we need to asses how well our method performs and identify any systematic uncertainties, we use a mock version of the DESI survey produced with the `ABACUS-SUMMIT` simulation, which features tracers like ELG, LRG, and QSO. Following this, we employ various techniques to see how the correlation functions from our EFT-GSM model align with those from the `ABACUS-SUMMIT` simulation under different parameters and model configurations. In this section we give details of the simulations used for this work.

5.1.1 Mock Data

The correlation function estimates are calculated as the average across 25 cubic boxes, each one having a volume of $(2\,h^{-1}\mathrm{Gpc})^3$. These boxes were created using the `ABACUS-SUMMIT` high-accuracy N-body simulations [108]. As a result, the total volume covered by these simulations reaches $200\,h^{-3}\mathrm{Gpc}^3$.

In each `ABACUS-SUMMIT` simulation box, there are 6912^3 particles, with each particle having a mass of $2 \times 10^9\,h^{-1}\mathrm{M}_\odot$. The simulations use cosmological parameters based on Planck 2018 [118], which include $h = 0.6736$, $\omega\mathrm{cdm} = 0.1200$, $\omega_\mathrm{b} = 0.02237$, $\ln(10^{10}A_s) = 3.0364$, $n_s = 0.9649$, $M_\nu = 0.06\,\mathrm{eV}$, and $w_0 = -1$.

Our mock catalogs are constructed using the halo occupation distribution (HOD) model, which is fitted to the *Survey Validation* 3 (SV3) spectroscopic data from DESI. Three separate sets of catalogs are created, each corresponding to a different tracer: LRG at $z = 0.8$, ELG at $z = 1.1$, and QSO at $z = 1.4$. To populate the simulations with galaxies, we apply the HOD approach outlined in [119] for ELG and in [120] for LRG and QSO. Calibration of the HOD models is achieved by using small-scale wedges (less than $5\,h^{-1}$Mpc) along with information on large-scale bias evolution, when available. Each simulation snapshot undergoes independent calibration to ensure accuracy.

The mock catalogs are constructed using the halo occupation distribution (HOD) model, which is fitted to the *Survey Validation* 3 (SV3) spectroscopic data from DESI. Three separate sets of catalogs are created, each corresponding to a different tracer: LRG at $z = 0.8$, ELG at $z = 1.1$, and QSO at $z = 1.4$. To populate the simulations with galaxies, we apply the HOD approach outlined in [119] for ELG and in [120] for LRG and QSO. Calibration of the HOD models is achieved by using small-scale wedges (less than $5\,h^{-1}$Mpc) along with information on large-scale bias evolution, when available. Each simulation snapshot was obtained through independent calibration to ensure accuracy.

For our analysis, the s-bins are set at intervals of $\Delta s = 4\,h^{-1}$Mpc. Figure 5.1 illustrates the average `ABACUS` multipoles for $\ell = 0, 2, 4$ across the three tracers: LRG (left panel), ELG (middle panel), and QSO (right panel).

5.1.2 Covariance Mocks

To calculate covariance matrices, we use the correlation function derived from 1000 simulations obtained with the Zel'dovich method [121], known for being efficient rather than highly accurate. These simulations share the same cosmological parameters as `ABACUS-SUMMIT`. The covariance matrix C for each tracer, between bins i and j, is determined using:

$$C_s^{(ij)} = \frac{1}{N_{\text{mocks}} - 1} \sum_{m=1}^{N_{\text{mocks}}} \left(\xi_i^m - \bar{\xi}_i\right)\left(\xi_j^m - \bar{\xi}_j\right), \tag{5.1}$$

where $N_{\text{mocks}} = 1000$ is the total number of mocks, and $\bar{\xi}_i$ is the average value for the i^{th} bin.

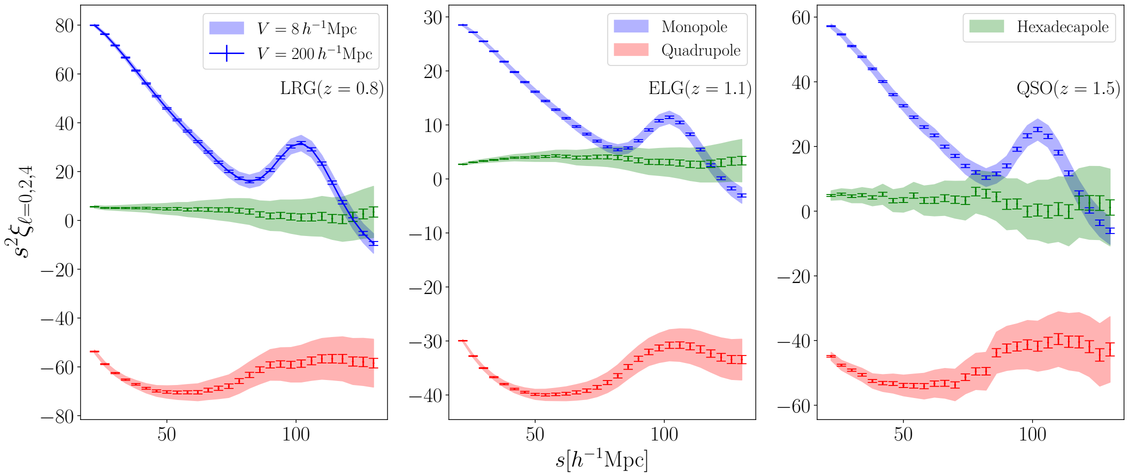

Figure 5.1: The plot illustrates the `ABACUS-SUMMIT` simulations where the multipoles of the correlation function for three tracers are: LRGs (left panel), ELGs (center panel), and QSOs (right panel). Each color corresponds to a different multipole, with the monopole shown in blue, the quadrupole in green, and the hexadecapole in red. The shaded area indicates the errors from the covariance matrix, while the error bars show these errors re-scaled by $1/\sqrt{25}$. Figure extracted from [44].

We decided do not to include the Hartlap factor as it slightly modifies the inverse covariance matrix from 0.94 to 0.95, depending on the scale, resulting in variations under 0.5%.

In our study, we apply three different re-scaling methods for the covariance. The first method involves no re-scaling, which corresponds to a simulation volume of $8\,h^{-3}\,\mathrm{Gpc}^3$. The second method re-scales by a factor of $1/5$, resulting in a volume of $V_5 = 40\,h^{-3}\,\mathrm{Gpc}^3$. The third method uses a re-scaling factor of $1/25$, corresponding to a volume of $V_{25} = 200\,h^{-3}\,\mathrm{Gpc}^3$. Figure 5.1 displays the square root of the diagonal terms of the `EZ mocks` covariance as a shaded region for the three tracers without rescaling: LRG (left panel), ELG (middle panel), and QSO (right panel). Additionally, we include error bars that represent the $1/25$ covariance.

5.2 Testing EFT-GSM Mode with Neural Network Accelerators in the Extended Space

This section focuses on enhancing our neural network by integrating two additional parameters, where the goal is to emulate the GSM-EFT to predict the galaxy correlation function based on specific cosmological parameters. Although the process of evaluating the code directly is relatively quick, conducting a Markov Chain Monte Carlo (MCMC) analysis demands around 10^5 evaluations. This problem drives us to reduce the evaluation time of our models by employing neural network emulators.

The introduction and rigorous testing of these emulators were detailed earlier and documented in the previous chapter. To train the Neural Network, we use 60,000 parameter space points. This approach is generally applicable to most results in this study, except for some advanced analyses that require more complex models and thus a larger training dataset, as elaborated in sections §5.3.2 and §5.3.2. Out of the total points, 50,000 are designated for training the algorithm. Another 5,000 points serve as the validation set, allowing the neural network to evaluate model accuracy during each training cycle. Depending on the model's performance, the algorithm might adjust the learning rate or finalize the model. The final 5,000 points form the test set, used to verify the model's accuracy on new data. Each set is independently created using a Korobov rule [75], ensuring uniform sampling in a high-dimensional space where the Neural Network can improve its learning rate.

The accuracy of our surrogate models is obtained using the test set points. In Figure 5.2, percentile lines illustrate the boundaries where 50%, 70%, and 90% of the errors fall within each bin. Most of the errors remain under the 1% mark for Abacus V_1, which aligns with the anticipated Year-One error for DESI. This indicates that our emulators are suitable for processing DESI Year-1 data.

As expected, more complex configurations of our model tend to result in emulators with slightly reduced accuracy. This is especially noticeable in the extended model detailed in Section 5.3.2, where the emulator's errors can reach levels similar to the expected errors for DESI Year 1 in about 5% of our test data points, even when we increase the data points to 100,000. It is important to note that these points are mostly found in less populated areas of the parameter space, particularly where $b_1 < 0.5$, which is far from the expected parameter values for our tracers. Although this has a negligible effect on the constraints discussed in this paper, we are actively exploring ways to enhance both the model's performance and the emulator's precision.

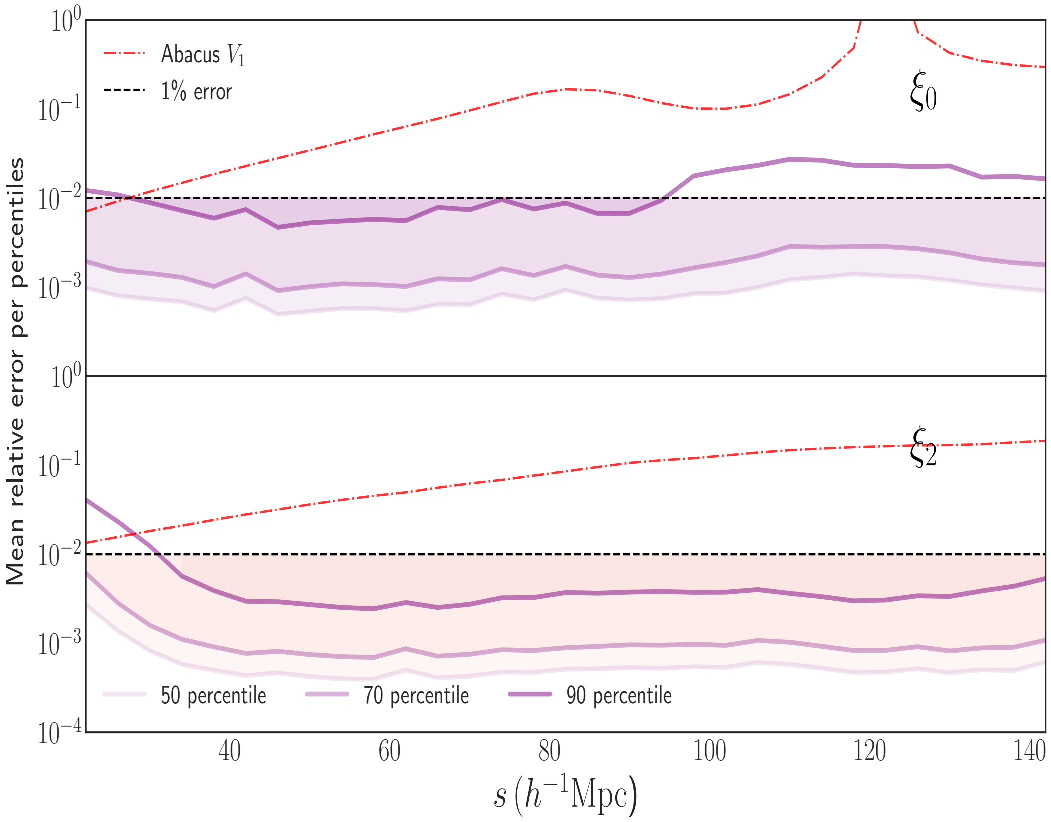

Figure 5.2: The neural network predictions on our test set are compared to the original model values, with percentile errors shown as colored solid lines. A dotted red line indicates the ABACUS error for a similar volume expected at the end of DESI Year One, estimated as $\sqrt{(C_s^{(ii)\ell}/\xi^\ell(r_i))}$. Meanwhile, the black line marks the 1% error threshold. The plot features errors for the Monopole at the top and the Quadrupole at the bottom, revealing that for most scales, the minimizer errors for 90% of our points are about ten times smaller than the anticipated DESI Year One error. Figure extracted from [44].

Free parameters and priors			
`Cosmological`		`Compressed`	
h	$\mathcal{U}[0.55, 0.91]$	$f\sigma_8$	$\mathcal{U}[0, 1]$
ω_b	$\mathcal{N}[0.02237, 0.00037]$	$\alpha_\perp$	$\mathcal{U}[0.5, 1.5]$
$\omega_{\rm cdm}$	$\mathcal{U}[0.08, 0.16]$	$\alpha_\parallel$	$\mathcal{U}[0.5, 1.5]$
$\log(10^{10}A_s)$	$\mathcal{U}[2.0, 4.0]$	m	$\mathcal{U}[-1, 1]$
	`Nuisances`		
b_1	$\mathcal{U}[0, 2.0]$	b_2	$\mathcal{U}[-5, 10]$
$\sigma^2_{\rm EFT}$	$\mathcal{U}[-20, 100]$	$c_{1,\rm EFT}$	$\mathcal{U}[-100, 100]$
b_{s^2}	$\mathcal{U}[-10, 10]$		

Table 5.1: In our baseline *Full Modeling* analyses of the `ABACUS-SUMMIT` simulations, we use the parameters and their priors listed here. Table extracted from [44].

5.3 Cosmological constraints from configuration space clustering statistics

We now use our methodology to explore the parameter space. Here, we utilize the MCMC code `emcee` [122], which is an ensemble sampler that uses the affine invariance as described in [47]. This tool offers several benefits, such as the possibility to run in parallel, making it ideal for leveraging the power of large computing clusters. In our study, we apply the priors listed in Table 5.1, where we include the limits for the two extra parameters, for the three of our independent methods. For all scenarios, we set the spectral index to $n_s = 0.97$, unless specified otherwise, and we keep the effective number of relativistic degrees of freedom at $N_{\rm eff} = 3.046$, along with a massive neutrino abundance of $\omega_{\rm ncdm} = 0.00064$.

To ensure that our results converge, we use the integrated auto-correlation time from [47], where it is checked every 100 steps. We consider a chain converged when the chain's length must exceed 100 times the estimated auto-correlation time and the change in this estimate should be less than 1%.

5.3.1 Baseline Analysis

We examine the behavior of the GSM model across different volumes using *Full Modeling* and *ShapeFit*. Figure 5.3 shows the posterior distributions for the LRG tracer with a minimum scale of $s_{min} = 30\,h^{-1}\,\mathrm{Gpc}$. We consider three volumes: $8\,h^{-3}\,\mathrm{Gpc}^3$ (approximating the first year of DESI), $40\,h^{-3}\,\mathrm{Gpc}^3$ (representing the fifth year of DESI), and $200\,h^{-3}\,\mathrm{Gpc}^3$ (the total combined volume of 25 simulations, used for testing models with small error bars). Table 5.2 provides the central values and 1σ errors for the cosmological parameters, showing that larger volumes result in tighter constraints while maintaining similar central values.

In *Full Modeling*, the constraints for h are approximately 0.4σ for V_1, 0.6σ for V_5, and 1.1σ for V_{25}. For $\ln(10^{10}A_s)$, the constraints are 0.5σ for V_1, 0.1σ for V_5, and 1.4σ for V_{25}. The parameter Ω_m remains below 1σ: 0.1σ for V_1, 0.1σ for V_5, and 0.3σ for V_{25}. Overall, *Full Modeling* shows that increasing the survey volume significantly tightens the constraints on cosmological parameters.

In *ShapeFit*, the constraints for $f\sigma_8$ are 0.04σ for V_1, 0.5σ for V_5, and 1.7σ for V_{25}. The parameter $\alpha_\parallel$ is 0.3σ for V_1, 0.5σ for V_5, and 0.9σ for V_{25}. The parameter $\alpha_\perp$ is 0.4σ for V_1, 1.0σ for V_5, and 2.4σ for V_{25}. Lastly, for m, the constraints are 0.03σ for V_1, 0.4σ for V_5, and 1.1σ for V_{25}. In conclusion, *ShapeFit* also benefits from larger volumes, providing tighter constraints, particularly for parameters like $f\sigma_8$ and $\alpha_\perp$.

Our first observations indicate that the large volume's shift seems related to the limits of our PT 1-loop corrections. Furthermore, our analysis shows that even when using the largest volume, no parameter exhibits a noticeable shift, and, subsequently, the results show that the accuracy of N-body simulations remains below 2 σ for these extensive volumes [e.g. 123]. Moreover, we chose to focus on a volume of $8, h^{-3}, \mathrm{Gpc}^3$ to compare our errors reliably with the current DESI year one analysis, unless otherwise specified. Additional aspects of our model include several nuisance parameters, which make it important to address the volume prior projection effect by comparing the MAP points against the mean values of our MCMC estimations.

Volume (Max. Freedom).

Full Modeling				
Volume	Ω_m	h	$\ln(10^{10}A_s)$	
$V_1 = 8h^{-3}\mathrm{Mpc}^3$	0.3147 ± 0.0099	0.6771 ± 0.0083	2.98 ± 0.12	
$V_5 = 40h^{-3}\mathrm{Mpc}^3$	0.3146 ± 0.0046	0.6768 ± 0.0052	3.003 ± 0.058	
$V_{25} = 200h^{-3}\mathrm{Mpc}^3$	0.3146 ± 0.0021	0.6776 ± 0.0035	2.999 ± 0.026	
ShapeFit				
Volume	$f\sigma_8$	$\alpha_\parallel$	$\alpha_\perp$	m
$V_1 = 8h^{-3}\mathrm{Mpc}^3$	0.449 ± 0.025	1.005 ± 0.018	0.9965 ± 0.0079	-0.001 ± 0.031
$V_5 = 40h^{-3}\mathrm{Mpc}^3$	0.444 ± 0.012	1.0037 ± 0.0077	$0.9960^{+0.0039}_{-0.0043}$	0.007 ± 0.017
$V_{25} = 200h^{-3}\mathrm{Mpc}^3$	0.4408 ± 0.0054	1.0033 ± 0.0035	0.9954 ± 0.0019	$0.0107^{+0.0088}_{-0.0099}$

Table 5.2: We summarize the constraints derived from fitting the ABACUS LRG sample using our *Full Modeling* and *ShapeFit* approaches, with a minimum scale of $s_{\min} = 30\,h^{-1}\mathrm{Mpc}$ across various volumes. Each method provides insights into how the model parameters behave under different conditions. As the volume increases, the precision of the constraints improves, although we notice some small shift. Table extracted from [44].

.

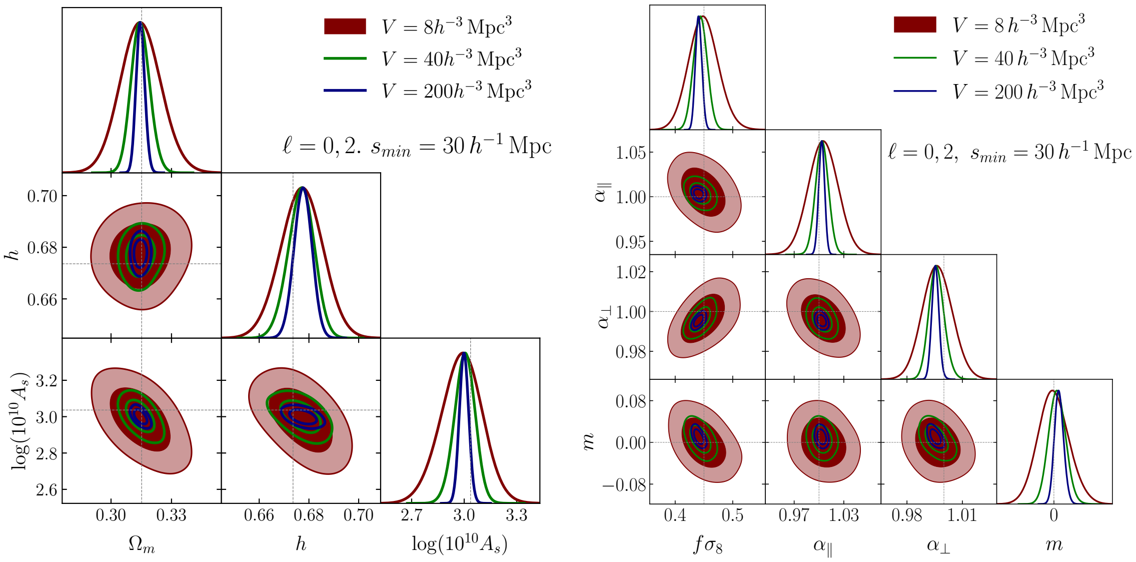

Figure 5.3: The triangular plot illustrates the posterior distribution of our EFT-GSM model, derived from fitting the average of 25 `ABACUS-SUMMIT` simulations of the LRG tracer, with different varying rescaling factors for the covariance matrix. On the left side, the results from *Full Modeling* are displayed, while the right side shows the outcomes from *ShapeFit*. The maroon color represents the non-rescaled covariance matrix, the green indicates a covariance matrix rescaled by a factor of 5, and the blue corresponds to a rescaling factor of 25. These plots have a minimum scale of $s_{min} = 30h^{-1}$Gpc. Figure extracted from [44].

Optimizing the fitting range

We now use our baseline approach for the three methodologies described in the previous chapter. For this purpose, we fit the average of 25 `ABACUS-SUMMIT` simulations. Our baseline approach involves using the monopole and quadrupole of the correlation function, with the covariance calculated for a volume of $8\,h^{-3}\,\mathrm{Gpc}^3$. We set a maximum range scale of $s_{max} = 130\,h^{-1}\,\mathrm{Mpc}$. Note the similarity with other research [103] where it is indicated that larger scales do not provide significant additional information.

To find the optimal minimum range, we perform some fittings on the LRG tracer for different minimum scales in addition to our standard baseline settings. In this subsection, we discus how the GSM-EFT model performs across various minimum scales using all methodologies: *Standard* analysis, *ShapeFit*, and *Full Modeling*. Each approach is evaluated to determine its effectiveness in capturing the relevant features of the data at different scales. By comparing these methodologies, we aim to understand how scale choices impact the model's accuracy and determine which scales should be used as part of our standard settings.

In the *Standard* approach, we constrain all the information through the $f\sigma_8$, $\alpha_\perp$, and $\alpha_\parallel$ parameters, therefore, to recover the cosmological parameters, we need an extra MCMC methoodas explained in section 3.1. In the first step of the compression part,we have a reference template which is based on the `ABACUS-SUMMIT` cosmology, therefore, we anticipate $\alpha_\perp = 1$ and $\alpha_\parallel = 1$. The same aply to *ShapeFit* methodology, but with one extra parameter m, which we expect to be 0. Figure 5.4 displays the constraints for both methodologies following the settings of the previous subsection. We compare the *Standard* methodology (left) and *ShapeFit* (right) across three different minimum scales: $22\,h^{-1}\,\mathrm{Mpc}$, $30\,h^{-1}\,\mathrm{Mpc}$, and $38\,h^{-1}\,\mathrm{Mpc}$. All results align with the `ABACUS-SUMMIT` cosmology within 1σ, which shows the robustness of our methodology. At this stage of compressed parameters, both methodologies show similar performance for the common parameters, which is expected since the m parameter is designed to offer additional information. This extra information about the slope of the linear power spectrum makes *ShapeFit* as effective as *Full Modeling* when translating the compressed parameters from *ShapeFit* into the cosmological

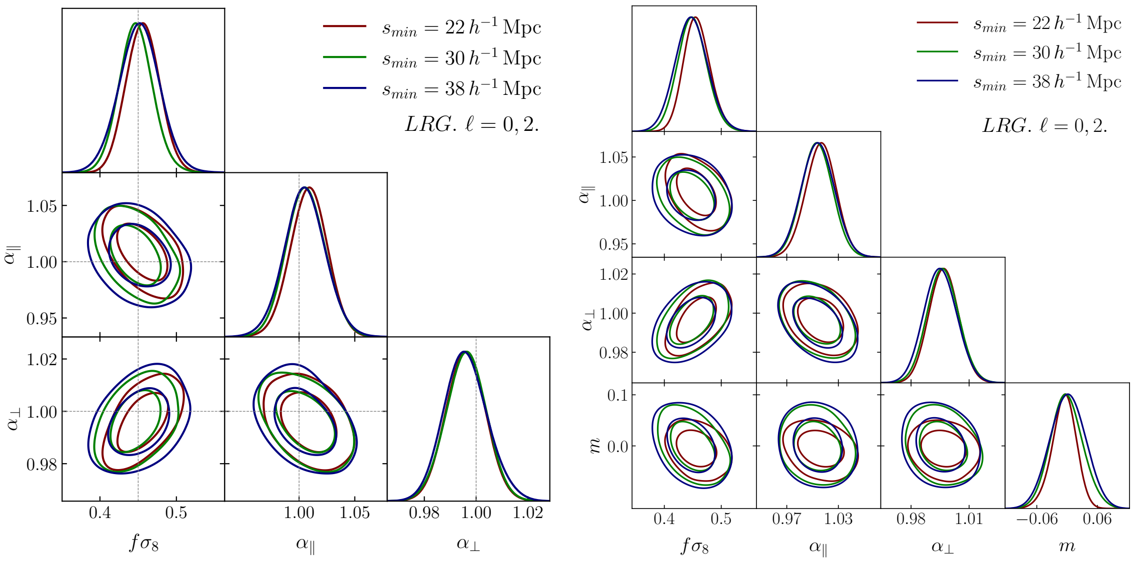

Figure 5.4: The triangular plot illustrates the posterior distribution derived from the average of 25 `ABACUS-SUMMIT` simulations, for the smallest volume, for the LRG tracer for different minimum scales. On the left, we have the constraints for the *Standard* methodology, while the right side displays the constraints for the *ShapeFit* methodology. Figure extracted from [44].

parameters for ΛCDM, as demonstrated in [36, 37].

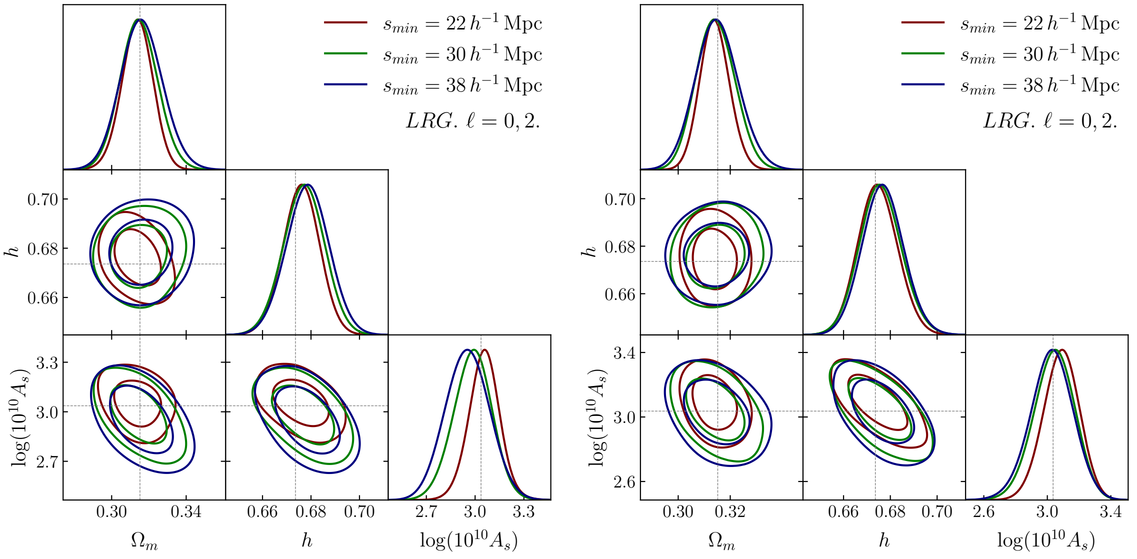

Figure 5.5: The triangular plot illustrates the posterior distribution derived from the average of 25 `ABACUS-SUMMIT` simulations, for the smallest volume, for the LRG tracerfor different minimum scales. On the left, we have the constraints for the *Full Modeling* methodology, while the right side displays the constraints for the *ShapeFit* methodology. Figure extracted from [44].

On the other hand we have that for the non compressed space of *Shape Fit*, and our *Full Modeling* approach, we let all the parameters to be varyed. Therefore, our *Full Modeling* baseline analysis includes four free cosmological parameters: $\{\Omega_m,\ h,\ \ln(10^{10}A_s),\ \omega_b\ \}$. However, as outlined in our methodology, we apply a Gaussian BBN prior on ω_b and flat priors on the nuisance and EFT parameters, as detailed in Table 5.1. Figure 5.5 shows the posterior distribution contours for our baseline analysis in this approach, varying the minimum scale s_{min} across three values with a bin separation $8\,h^{-1}$Mpc.The right side shows a triangular plot for the *Full Modeling* methodology. In the *Shape Fit* approach, we only fit the cosmological parameters from the chains of the compressed parameters. Here, the left side displays the *ShapeFit* methodology for the same minimum scales. Both methodologies align well with the true cosmological parameter values across all scales, with deviations from the true values of $\leq 0.1\sigma$ for Ω_m, $\leq 0.6\sigma$ for h, and $\leq 0.7\sigma$ for A_s at $s_{min} = 22\,h^{-1}$Mpc for both *Full Modeling*. In *ShapeFit*, deviations are similarly small, with $\leq 0.2\sigma$ for Ω_m, $\leq 0.4\sigma$ for h, and $\leq 0.5\sigma$ for $\ln(10^{10}A_s)$. Overall, no significant shifts are detected with either methodology. Figure 5.6 illustrates the correlation function's shape, representing the mean values from the baseline analysis fit, which from previous results, we set a minimum scale of $s_{min} = 20\,h^{-1}$Mpc and maximum scale of $s_{min} = 130\,h^{-1}$Mpc giving a totla 58 bins for the 9 free parameters. For this parameters we obtain a $\chi^2 = 60.57$

As observed in this section, while all choices of minimum scales yield results within 1σ, using $s_{min} = 22\,h^{-1}$Mpc results in the smallest error bars without any noticeable bias. This finding aligns with conclusions from other studies [35]. Consequently, we identify $s_{\rm range} = 22 - 130\,h^{-1}$Mpc as the optimal fitting range. Unless stated otherwise, the following sections' results will be presented using this range.

Maximal and minimal freedom

In this subsection, we investigate two different configurations for the parameter space of galaxy bias terms, labeled as *maximal* and *minimal* freedom. In the *maximal* configuration, all parameters are allowed to vary simultaneously, serving as our baseline setup. On the other hand, the *minimal* configuration assumes a co-evolution scenario where non-local biases emerge solely through gravitational evolution, meaning they do not appear in the

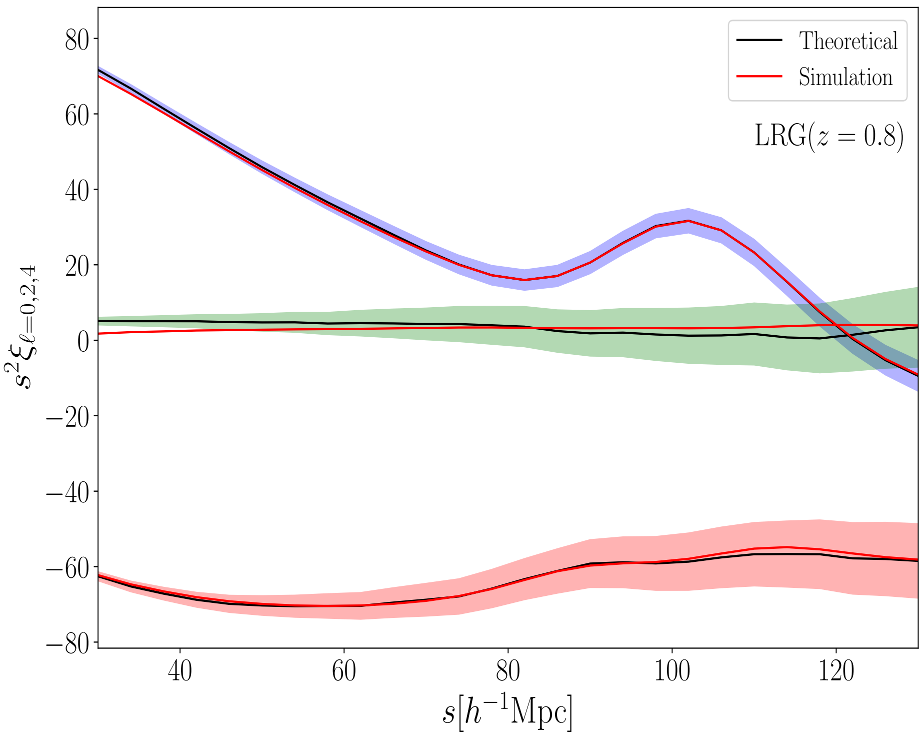

Figure 5.6: This plot illustrates the shape of the multipoles of the correlation function derived from the average of 25 `ABACUS-SUMMIT` simulations for the LRG tracer, alongside the mean of the posterior distribution for our standard settings, within the range of $s_{min} = 30\,h^{-1}$Mpc to $s_{max} = 130\,h^{-1}$Mpc. The shaded areas represent the error bars for the small volume. Figure extracted from [44].

Lagrangian coordinates framework, which corresponds to setting $b_{s^2} = 0$.

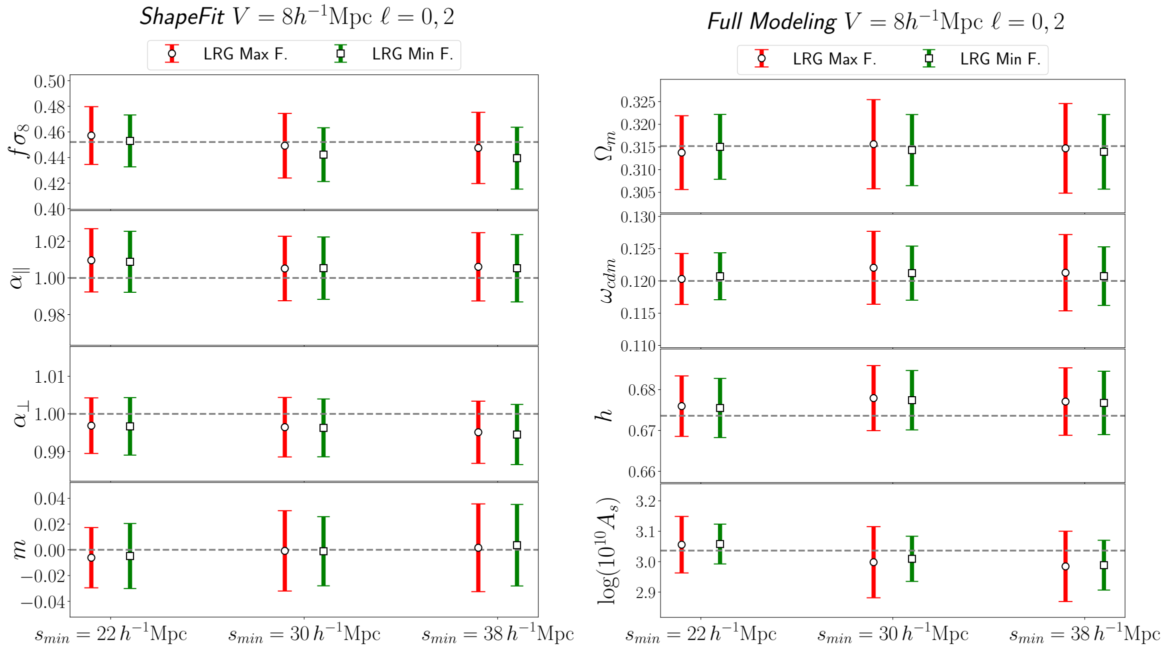

Figure 5.7: We compare error bars for LRG using a volume of $8\,h^{-3}\mathrm{Gpc}^3$. The red bars represent results from the maximal freedom configuration, while the blue bars correspond to the minimal freedom. The model's performance is shown at different minimum scales, each separated by a bin of 8 h^{-1}Mpc. On the left, the *ShapeFit* fits, and on the right, the *Full Modeling* fits. Here we show different distances, although our standard configuration uses $s_{min} = 30\,h^{-1}$Mpc. Figure extracted from [44].

In figure 5.7, we can observe the posteriors that our modeling gives for the LRG tracer. The average values of the posterior distribution for parameters obtained using both *maximal* and *minimal* freedom configurations are within 1 σ applying the *ShapeFit* and *Full Modeling* methods. In both the *Full Modeling* and *ShapeFit* approaches, the central values are quite consistent. However, the *Minimal* freedom configuration results in approximately 10% smaller error bars, which is expected as it has one parameter less. However, when examining the compressed parameter space specifically at the $22\,h^{-1}$Mpc scale, we find that the *Maximal* freedom configuration provides mean value predictions for $f\sigma_8$ that are closer to the true simulation values in addition of smaller error bars. Meanwhile, the *Minimal* freedom setup offers more accurate predictions for $\alpha_\parallel$, although with larger error bars.

Our study began by comparing the predicted central values with the expected ones, finding that both the Minimal and Maximal freedom configurations perform similarly.

Analysis in the compressed parameter space, as well as in the cosmological parameter space, produced consistent outcomes. Results indicated that the fitted parameters have not shifted significantly from their true values for the volume considered. Examination with a baseline of $s_{min} = 22, h^{-1}$Mpc revealed that employing the more constrained Minimal freedom configuration can reduce most error bars by approximately 10%, whether one adopts the compressed or Full Modeling approach. For this work, we decided to choose the conservative scenario by setting the Maximal freedom scenario as our standard configuration.

5.3.2 Extensions to Baseline Analysis

We extend our main analysis with three additional extensions to our baseline settings. Initially, we examine the extra information provided by the hexadecapole of the correlation function. Next, we consider the possibility of placing constraints on the spectral index n_s. Finally, we allow the Dark Energy equation of state to evolve over time by employing the (w_o, w_a) prescription.

Effect of including hexadecapole

We now assess the effect of incorporating the hexadecapole into our baseline analysis. This multipole is notably challenging to model accurately due to limitations in first-loop Perturbation Theory (PT) when addressing small-scale phenomena. Our investigation aims to clarify these challenges and improve our understanding of the modeling complexities associated with the hexadecapole. Discrepancies become more pronounced at scales below $30\,h^{-1}$Mpc in configuration space, where non-linear effects significantly influence the results. This breakdown in PT accuracy underscores the difficulties in capturing the full complexity of the hexadecapole at these smaller scales. Consequently, this approach helps us better understand how the additional multipole influences our overall model predictions.

In Figure 5.8, the impact of including the hexadecapole on both the compressed and cosmological parameter spaces is shown for three different minimum scales: 22, 30, and $38\,h^{-1}$Mpc. Results from the baseline analysis are indicated in red, while those with

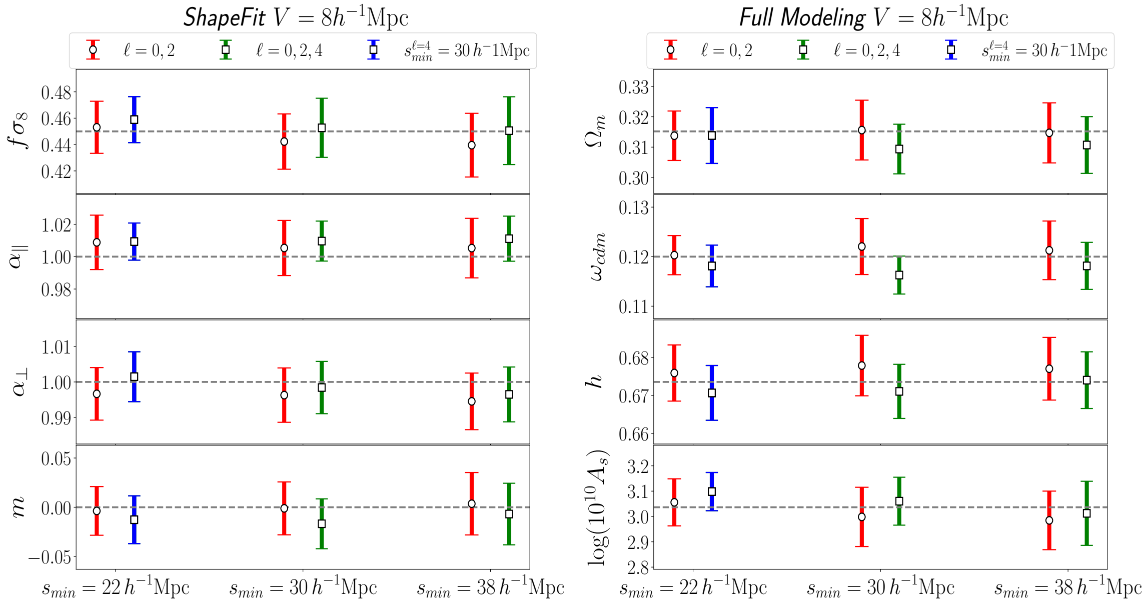

Figure 5.8: These plots show Constraints for the two methodologies, they are presented as a function of the minimum scale for LRGs, using the smallest volume of 8 h^{-1}Mpc. Results from the baseline analysis that do not include the hexadecapole are shown in red, while those that incorporate the hexadecapole appear in green. The left panel displays the fits obtained with the *ShapeFit* method, and the right panel illustrates the ones from the *Full Modeling* approach. Additionally, the lower bin, colored in blue, follows the same settings, except that the bin at $s_{min} = 22\,h^{-1}$Mpc is excluded from the analysis as it is not well modeled. Figure extracted from [44].

the added hexadecapole are shown in green. The bin at $22\,h^{-1}$Mpc is removed for the hexadecapole because small-scale modeling is particularly difficult, although it is still used for the monopole and quadrupole, which help us to have better quotes as when we indluded the bin, multimodes appears in the chains. Inclusion of the small hexadecapole bin creates a bimodal posterior, with a secondary high-likelihood region that Planck has already ruled out at more than 3σ. Since our model is less reliable at these small scales, this bin is excluded. Slight variations within 1σ are observed in the compressed parameters $f\sigma_8$, $\alpha_\parallel$, $\alpha_\perp$, and m, and similarly, the cosmological parameters Ω_m, h, and A_s remain consistent within 1σ.

In the ShapeFit approach, including the hexadecapole leads to modest reductions in error bars across the three minimum scales. For $f\sigma_8$, the error bar decreases by roughly 0.3% to 0.1% compared with the baseline, and for $\alpha_\parallel$, the reduction is about 0.3%. Meanwhile, the change in $\alpha_\perp$ remains nearly constant at around 0.1%, and the parameter m

varies from an increase of approximately 0.0% to a decrease of about 0.1%. In contrast, for Full Modeling, the error bar for Ω_m shifts from an increase of roughly 0.1% to a decrease of 0.2%, while for h, the reduction ranges from nearly 0.0% to 0.1%, and for $\ln(10^{10}A_s)$, the error bar decreases by between 0% and 0.2%.

In summary, incorporating the hexadecapole does not lead to major improvements, even after excluding the lowest distance bin to avoid issues on quasi-linear scales. Overall, the most notable effect is a modest reduction in the error margins for the Full Modeling analysis. Moreover, for the volume used here, all the parameter constraints remain within 1σ of the true values when the hexadecapole is included.

Exploring the impact of n_s free

The scalar spectral index n_s, characterizes how the amplitude of primordial density fluctuations depends on the wave number k. In our standard analysis, we have fixed n_s to the Planck 2018 value of 0.9649 [118]. Now, we extend our baseline analysis by treating n_s as a free parameter, which allows us to explore a larger region of the parameter space and examine its effects in greater detail and look for possible systematics for the smallest volume of our simulations, which corresponds to a larger volume compared to the year 1 of DESI for one tracer. For this exploration, we adopt the following prior on n_s:

$$n_s : \mathcal{U}\left[0.5, 1.5\right]. \tag{5.2}$$

Introducing the new degree of freedom increases the overall complexity of our models, so to retain the precision of our emulator we have increased the number of multipoles used in the neural network to 90,000. Allocating these multipoles, we assign 80,000 for training and reserve 5,000 as a validation set. Setting aside the remaining 5,000, we use them as our test set to gather statistics for evaluating the model's performance. All the remaining architecture remains the same.

The findings discussed in this subsection are derived exclusively using our Full Modeling methodology. Figure 5.9 presents the triangular plots of our analysis for the LRG sample using the Full Modeling approach, with n_s now treated as a free parameter. Results are shown for three different minimum scales, and each scale yields a distinct level of precision in parameter recovery, although all parameters, including n_s, consistently fall within the 1σ uncertainty contour. At $s_{\text{min}} = 22\,h^{-1}$Mpc, the mean value of h and $\log(10^{10}A_s)$ are particularly precise, with deviations of roughly 0.1σ from the true value. When s_{min} is increased to $30\,h^{-1}$Mpc, improvements are evident for Ω_m, h, and n_s, with deviations of about 0.1σ, 0.2σ, and 0.1σ, respectively. At $s_{\text{min}} = 38\,h^{-1}$Mpc, the results for Ω_m and n_s are pretty close, with shifts near to 0.1σ.

This analysis reveals that the precision of parameter recovery depends strongly on the chosen minimum scale, s_{min}, when n_s is allowed to vary. Different s_{min} values result in distinct levels of performance, showing that careful selection of this scale is crucial in full modeling analysis. Notably, the scale of $30\,h^{-1}$Mpc proves to be the most effective for most parameters, with the exception of A_s, which is still estimated within $1\ \sigma$ of the true value. Overall, our findings demonstrate that the method reliably captures all parameters within $1\ \sigma$ uncertainty..

Exploring the impact of w_0 and w_a as free parameters

In this section, we explore a parametrization of the dark energy, where these parameters allow us to study an evolving equation of state instead of assuming a constant dark energy density. The parameters w_0 and w_a are essential in modeling the evolution of dark energy according to the Chevallier–Polarski–Linder framework [124, 125]. Instead of assuming the constant behavior of dark energy as in the ΛCDM model, this approach allows for a time-varying equation of state. By defining

$$w(a) = w_0 + w_a(1 - a), \tag{5.3}$$

In this subsection, we build only on our Full Modeling approach by investigating two separate extensions. One approach introduces variations in the parameter w_0 alone, which reflects the parametrization $w(a) = w_0$ into our baseline model. The second approach varies both w_0 and w_a, having the full parametrization.

As we extend our standard setup by adding two parameters: w_0 and w_a, we have to

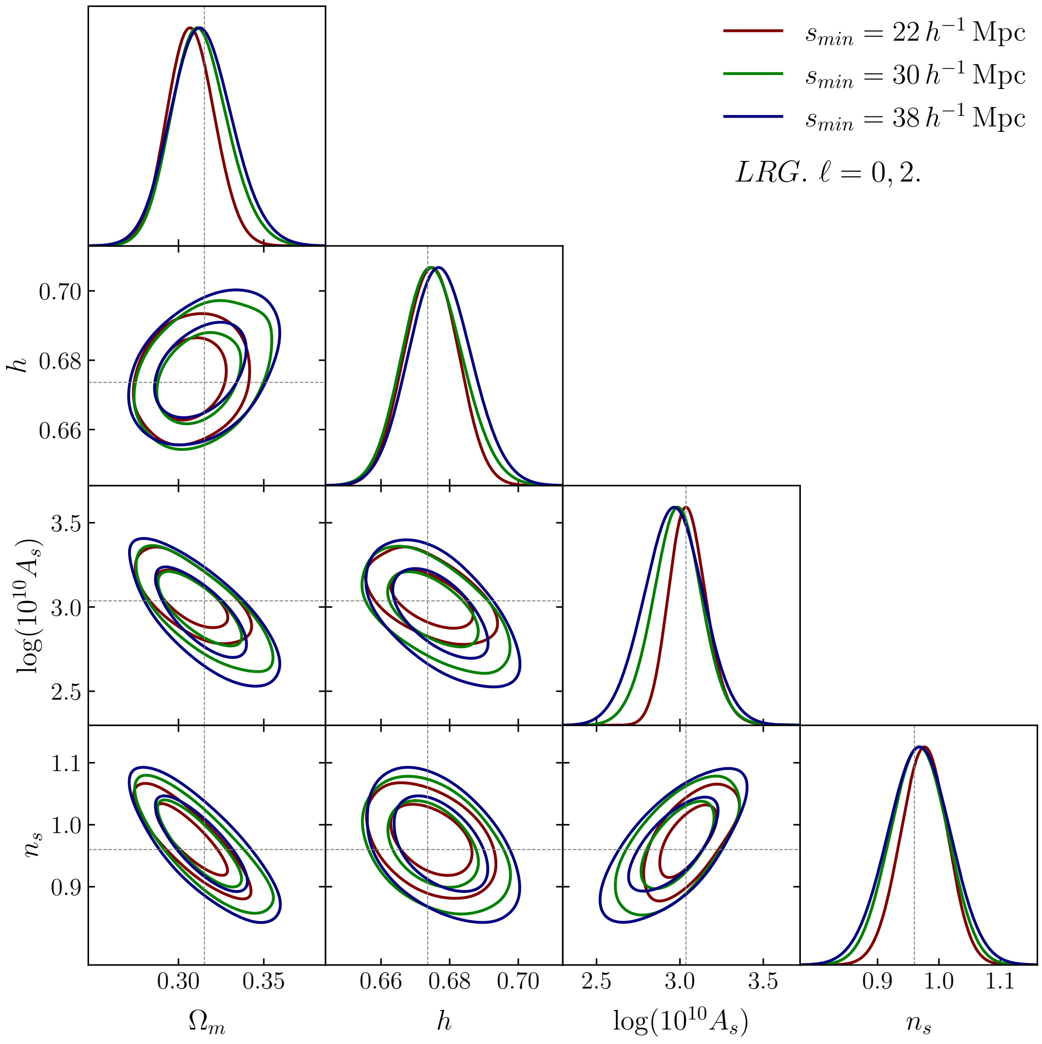

Figure 5.9: The triangular plot presents the 1 σ and 2 σ posterior contours for our baseline analysis on the LRG sample using the Full Modeling approach with n_s as a free parameter. Each contour reflects the uncertainties in our parameter estimates, showing how predictions shift as the minimum scale is adjusted. Figure extracted from [44].

modify the number of points used to generate the emulator. To maintain our emulator's accuracy, we have increased the number of multipoles used in training and testing to 170,000 points. Of these, 150,000 are allocated for training, 10,000 serve as a validation set, and the remaining 10,000 form a test set to evaluate the model's performance. All the architecture remains the same. Additionally, we impose the following priors on our parameters:

$$w_0 : \mathcal{U}[-2, 0] \tag{5.4}$$

$$w_a : \mathcal{U}[-3, 3]\,. \tag{5.5}$$

To begin, we report our findings for the parametrization that incorporates only the parameter w_0. A triangular plot, Figure 5.10, shows the constraints on the cosmological parameters, We note that for different distances, all the parameters are recovered within 1 sigma from their true values. The case with $s_{\min} = 22\,h^{-1}$Mpc produces notably tighter contours and presents more precise predictions for A_s, something we expect as it includes more information through one extra bin. Aditionally, the model exhibits deviations of 0.3σ for Ω_m, 0.0 σ for h, 0.1σ for $\log(10^{10}A_s)$, and 0.1σ for w_0. Overall, these results show that our extended model recovers all parameters, including the first parameter of the parametrization w_0.

Now, we focus on the extended model's performance, incorporating both w_0 and w_a as free parameters, which reveals complementary information to the previous analysis, showing that all parameter constraints are successfully recovered within a one standard deviation of their true values. The three different ranges for the model demonstrate comparable effectiveness in their accuracy in most of the parameters. Note that a better performance was observed for a minimum scales of $s_{\min} = 30, h^{-1}$Mpc and $s_{\min} = 38, h^{-1}$Mpc specifically for recovering A_s and Ω_m parameters, although the model with $s_{\min} = 22, h^{-1}$Mpc maintains the advantage of producing more precise measurements with reduced error margins.

The results of our analysis demonstrate high precision in parameter recovery when examining the model with $s_{\min} = 22\,h^{-1}$Mpc. Statistical deviations remain consistently within one standard deviation, we find small variations: Ω_m shows a minimal deviation of $0\,\sigma$, and h exhibits $0.0\,\sigma$ variation. Furthermore, $\log(10^{10}A_s)$ maintains a $0.6\,\sigma$ difference, while w_0 and w_a shows variations of $0.5\,\sigma$ and 0.7σ respectively. These results demonstrate the robustness of our extended model and its capability to accurately recover cosmological parameters, particularly in an extension using the parametrization of the equation of state, which includes both w_0 and w_a components.

5.3.3 All DESI tracers: LRG, ELG, QSO mocks

So far, we have explored the systematics for the LRG tracer, now, we extend our study by incorporating fits for two additional tracers: ELG and QSO mock samples, presented in Section 4.3.1. Subsequently, a new dataset is obtained from combining the three individual data vectors into a single one. Furthermore, each vector contributes to the total logarithmic likelihood, which is then modeled accordingly. For this section, we present two approaches: one based on the ShapeFit template and another based on the Full Modeling methodology.

Figure 5.14 shows the posterior limits for the two methodologies, *ShapeFit* and *Full Modeling*, and it incorporates all the tracers, LRG, ELG and QSO. Every measurement falls within one standard deviation of its expected value, showing clear consistency and reliable accuracy. Consequently, since the results for the smaller minimum scale are a bit more precise and all data agree closely with expectations, we have chosen this set $s_{\min} = 22, h^{-1}$Mpc as our main outcome .

Typically, the level of precision for both LRG and ELG is quite similar, with ELG showing a slightly tighter constraint, while QSO results looks less precise. These results match our expectations based on the effective volumes of the samples, as ELG is the largest DESI sample and QSO is the smallest; it is clear that we will have less precision in the QSO. Furthermore, the combined fit reveals that all parameters agree with their expected values within one sigma, showing deviations of $0.2\,\sigma$ for Ω_m, $0.7\,\sigma$ for h, and $0.3\,\sigma$ for $\log(10^{10}A_s)$.

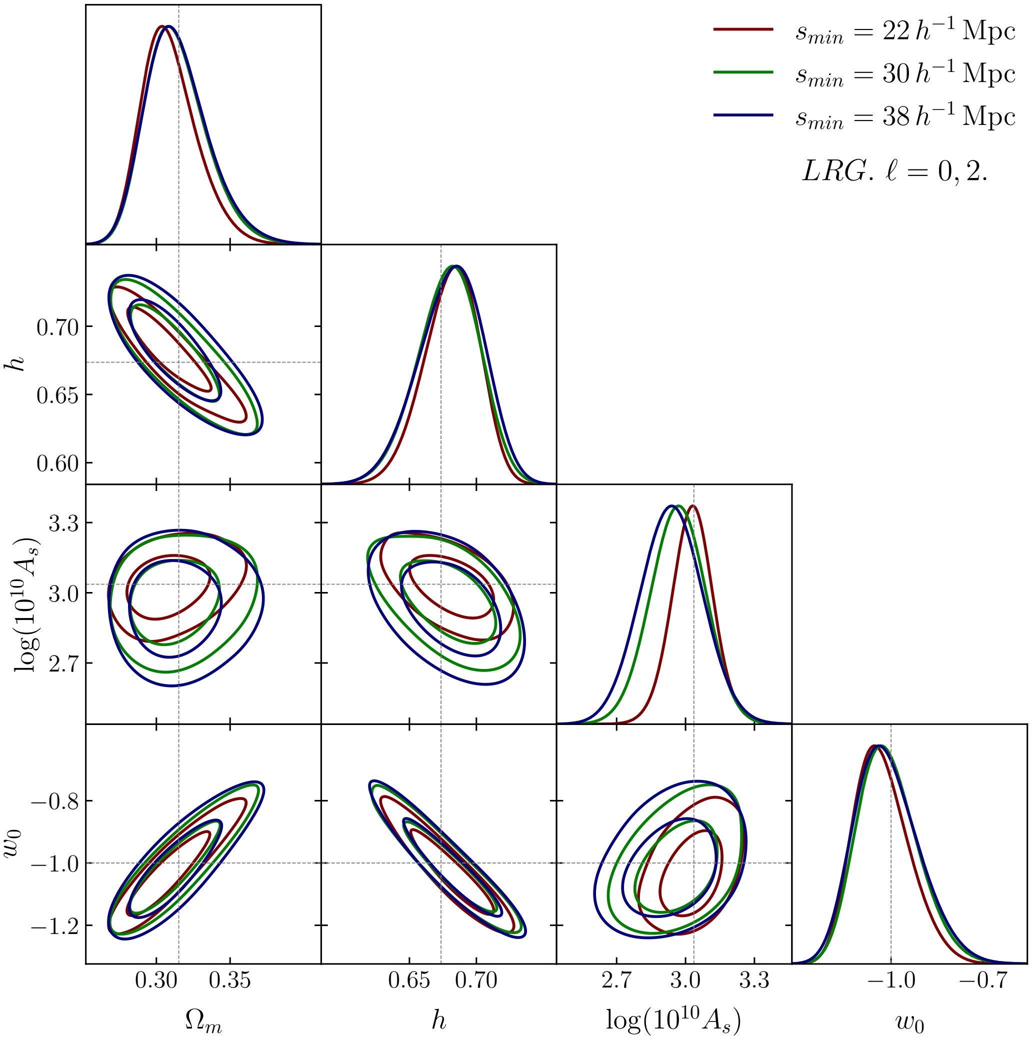

Figure 5.10: The triangular plot presents contours representing one-sigma and two-sigma posterior probabilities derived from the baseline analysis with w_0 as a free parameter. Subsequently, the full modeling technique is employed on the LRG sample to extract information from the data for different minimum scales. Figure extracted from [44].

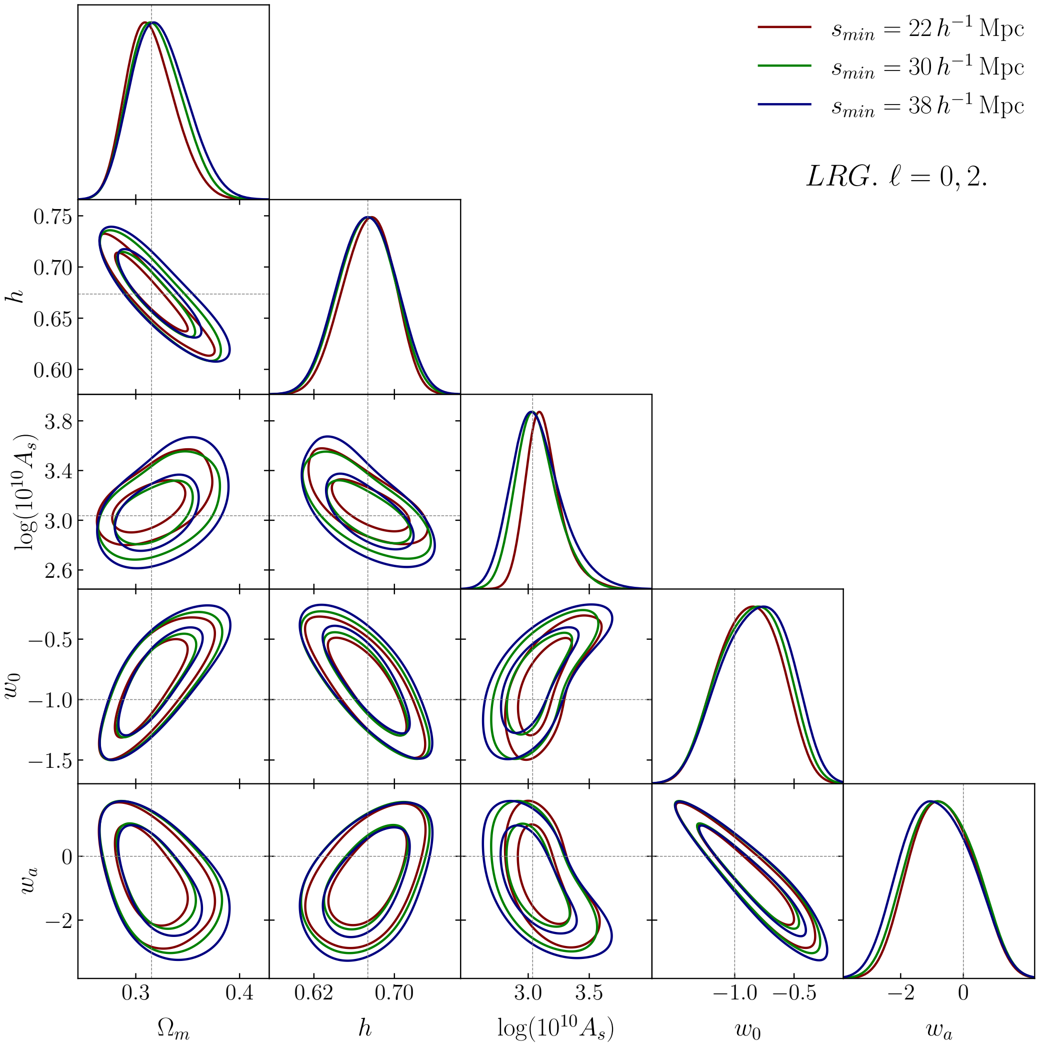

Figure 5.11: The triangular plot presents contours representing one-sigma and two-sigma posterior probabilities derived from the baseline analysis with w_0 and w_a as free parameters. Subsequently, the full modeling technique is employed on the LRG sample to extract information from the data for different minimum scales. Figure extracted from [44].

5.3.4 *ShapeFit* and *Full Modeling* methodology comparison

For this section, we convert *ShapeFit* from its compressed form into cosmological parameters to enable a fair comparison against *Full Modeling* within the same parameter space. In Figure 5.13, we show some minor differences when fitting the correlation function with LRG data over a volume of $V = 8\,h^{-3}\text{Gpc}^3$ and on a minimum scale of $s_{min} = 30\,h^{-1}$Mpc, which based on the previous finding, we set as our standard setting. Here, we find a consistency across both methods for the Maximal Freedom and Minimal Freedom configuration,

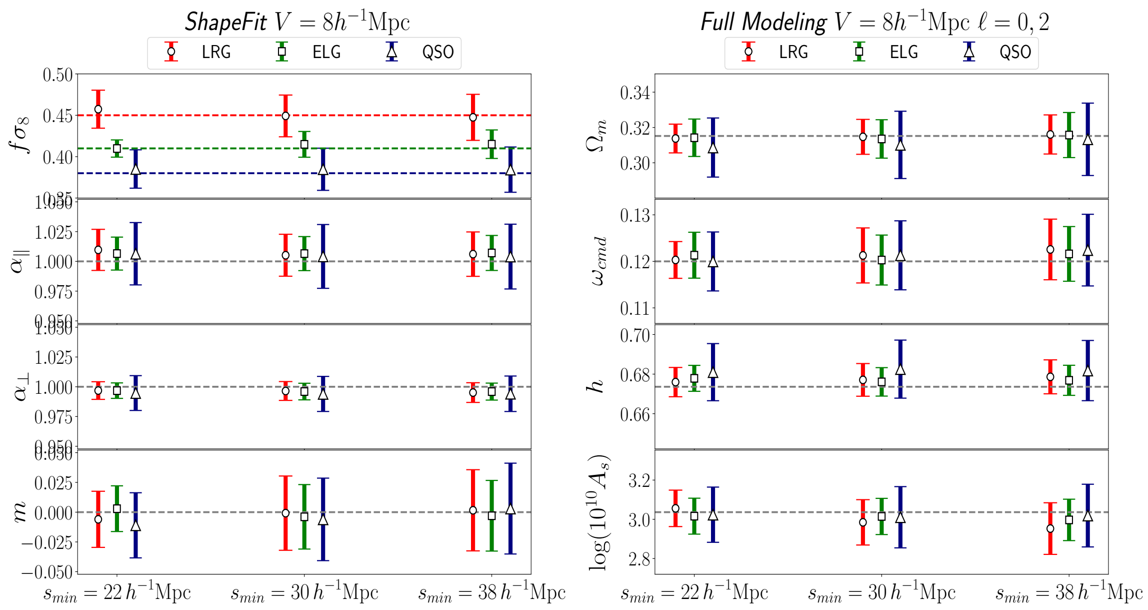

Figure 5.12: Error bar plots display the constraints on LRG (red), ELG (green), and QSO (blue) across three different minimum scales while comparing *ShapeFit* (left) and *Full Modeling* (right) methodologies. Meanwhile, black dashed lines indicate the true values of the parameters. Additionally, it should be noted that the true value of $f\sigma_8$r depends on redshift, and the values should not coincide. Figure extracted from [44].

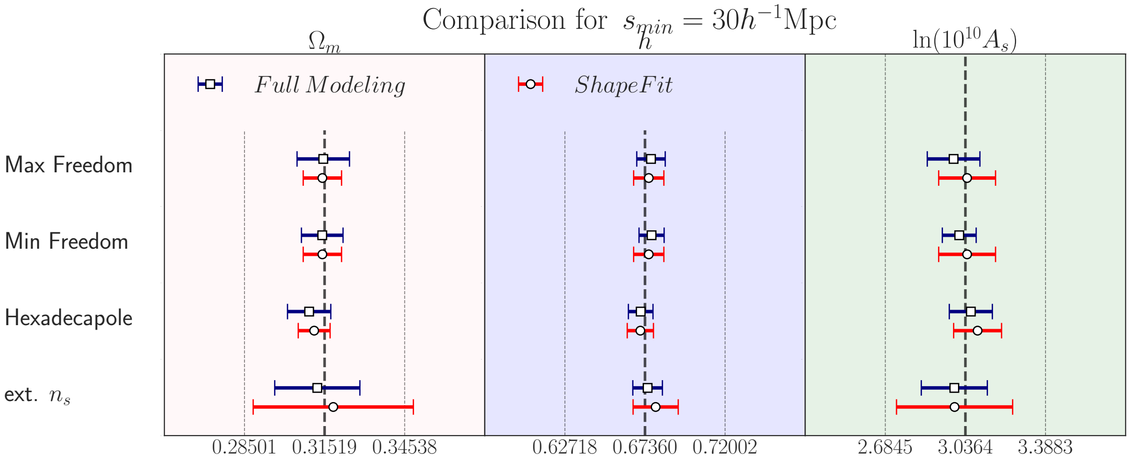

Figure 5.13: In this plot, we show a comparison between the *Full Modeling* and *ShapeFit* methodologies in the cosmological parameter space. Here, the red bars indicate the *ShapeFit* outcomes and blue bars show the *Full Modeling* results. Each result was obtained by fitting the mean of 25 realizations from the `ABACUS-SUMMIT` simulations for the LRG tracer for the monopole and quadrupole. Figure extracted from [44].

the inclusion of the hexadecapole, and the n_s extension, all of which display a comparable performance to constrain the parameters.

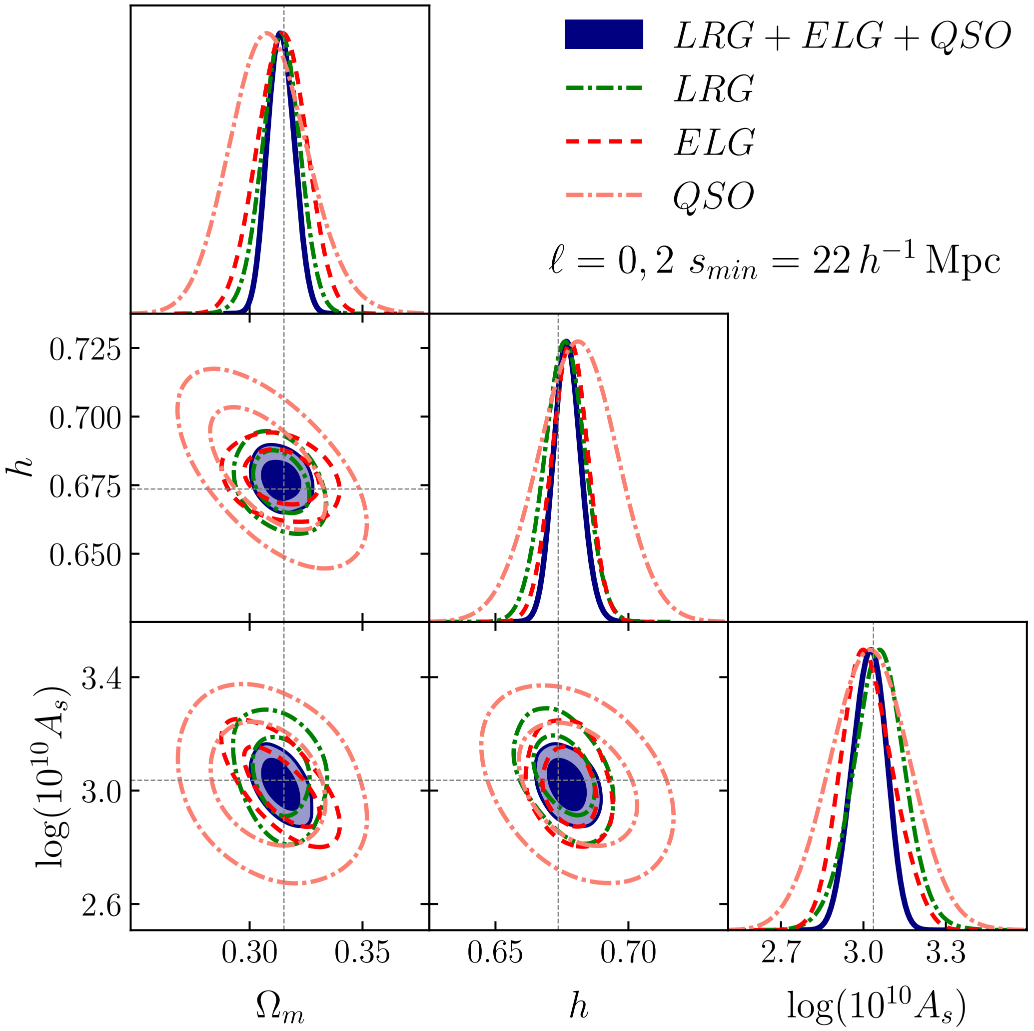

Figure 5.14: In this plot, we show the combined and individual posterior contours for all the Abacus simulations tracers. Figure extracted from [44].

In the previous section, we presented the baseline analysis, which demonstrated solid performance. Later, we found that for Ω_m, the results from both methodologies are almost identical, with a central value difference around 10^{-3} and error ranges $0.68\sigma_{FM} < \sigma_{SF} < 0.74\sigma_{FM}$, making *ShapeFit* slightly tighter. Subsequently, similar findings for h indicate a central difference in the order of 10^{-3} while the error bars range from $1.11\sigma_{FM}$ to $1.0\sigma_{FM}$, meaning that in this case, *ShapeFit* is somewhat less restrictive. Finally, the parameter $\ln(10^{10}A_s)$ also shows a central value shift around 10^{-3}, with the error bars varying between $1.0\sigma_{FM}$ and $1.18\sigma_{FM}$, confirming that *ShapeFit* again provides less constrictive constraints.

Then, we review the Minimal Freedom scenario, which demonstrates comparable performance, with statistics provided in Table 5.3. The findings in for this configurations shows that for Ω_m, we have nearly identical central values, differing by around 10^{-3}, and error bars set between $0.76\sigma_{FM}$ and $0.80\sigma_{FM}$, indicating that *ShapeFit* is somewhat more constrictive in this case. Then, measurements for h reveal a similar central value discrepancy about 10^{-3}, but with error bars ranging from $1.05\sigma_{FM}$ to $1.1\sigma_{FM}$, which suggests that *ShapeFit* provides slightly looser constraints. Finally, for $\ln(10^{10}A_s)$, a central value difference of approximately 10^{-2} is observed, accompanied by error bars between $1.09\sigma_{FM}$ and $1.33\sigma_{FM}$, again showing that *ShapeFit* is less constrictive.

Including the hexadecapole in the baseline analysis reveals similar constraining power for both methods. Observing the results for Ω_m, the central value difference is around 10^{-3} with error bars of $0.71\sigma_{FM} < \sigma_{SF} < 0.77\sigma_{FM}$, which indicates that *ShapeFit* is more constrictive across all minimum scales. Considering the h parameter, a central value difference of about 10^{-3} is found along with error bars of $1.0\sigma_{FM} < \sigma_{SF} < 1.11\sigma_{FM}$, showing that in this case, *ShapeFit* is less constrictive. Lastly, for $\ln(10^{10}A_s)$, the central value differs by approximately 10^{-2} and the error bars lie between $1.0\sigma_{FM} < \sigma_{SF} < 1.09\sigma_{FM}$, again suggesting that *ShapeFit* provides poorer constraints.

Lastly, when extending the model to allow vary n_s, an extra parameter n is introduced in the *ShapeFit* approach, which forced us to vary 5 parameters: $\alpha_\perp$, $\alpha_\parallel$, $f\sigma_8$, m, and n. Subsequently, these variables are converted into cosmological parameters via the relation $n = n_s - n_{\texttt{ref}}$ [36]. All outcomes are presented in Table 5.3, where we can observe that for Ω_m the central value varies by roughly 10^{-3} and the error bars fall in the range $0.72\sigma_{FM} < \sigma_{SF} < 0.93\sigma_{FM}$, showing that *ShapeFit* offers more constrictive limits. Moreover, the h parameter also displays a central shift around 10^{-3} with uncertainties between $0.99\sigma_{FM}$ and $1.11\sigma_{FM}$, indicating that in this case, *ShapeFit* is less tight. Furthermore, the $\ln(10^{10}A_s)$ parameter reveals a central variation below 10^{-2} and error bars spanning from $0.88\sigma_{FM}$ to $1.18\sigma_{FM}$. Finally, the n_s parameter itself shows a central difference of approximately 10^{-3}, with error bars between $1.32\sigma_{FM}$ and $1.54\sigma_{FM}$, which suggests that *ShapeFit* provides looser constraints for n_s.

So far, we have investigated the performance of two methods for predicting cosmological parameters: *ShapeFit* and *Full Modeling*. Based on our findings, we confirms that both techniques accurately recover parameters in the baseline framework as well as in extended models, where, for the volume considered in this study, our results show that both methodologies yield equivalent performance.

5.4 Comparisons with other Effective Field Theory models

The analysis of galaxy Full Shape in DESI year one data involves four distinct EFT theories, with our `EFT-GSM` model being a crucial part of this group as it is the only one built to model the correlation function in configuration space. Recent research presented in [31] examines three Fourier space models, including the `Velocileptors` approach from [113], alongside the `Folps`ν framework by [41], and the `PyBird` system detailed in [111]. Building upon these comparisons, we now explore how these three Fourier space models compare with our configuration space method by Fourier transforming the power spectrum into a correlation function.

Our analytical approach is structured in two main parts to ensure comprehensive model comparison. First, we evaluate how configuration space parameters from our model compare against the established Fourier space results produced by `Velocileptors`, `Folps`ν, and `PyBird`. Subsequently, we employ the `desilike`[2] system, which serves as a unified analytical pipeline for DESI data processing. This powerful tool enables standardized analysis procedures, including MCMC sampling, likelihood estimations, and power spectrum generation. Here, when implementing `desilike`, the difference among the four models lies in their intrinsic methods for deriving redshift space multipoles from the base linear power spectrum.

Beyond its role in method standardization, `desilike` offers powerful advantages like transforming multipoles from Fourier to configuration space. This essential feature enables us to conduct new evaluations of the `Velocileptors`, `Folps`ν, and `PyBird` results directly in configuration space. By utilizing these transformations, we can examine how different

---

[2] https://github.com/cosmodesi/desilike/

Cosmological Parameter Constraints

		Min Freedom			
T	s	Ω_m	h	$\ln(10^{10}A_s)$	
FM	22	0.3150 ± 0.0072	0.6755 ± 0.0072	3.058 ± 0.065	-
FM	30	0.3139 ± 0.0082	0.6767 ± 0.0078	2.989 ± 0.082	-
FM	38	0.314 ± 0.010	0.6786 ± 0.0080	2.96 ± 0.11	-
SF	22	0.3144 ± 0.0058	0.6743 ± 0.0080	3.082 ± 0.087	-
SF	30	0.3146 ± 0.0062	0.6758 ± 0.0082	3.01 ± 0.10	-
SF	38	0.3155 ± 0.0077	0.6783 ± 0.0085	22.97 ± 0.12	-
		Hexadecapole			
FM	22	$0.3138^{+0.0099}_{-0.0085}$	0.6707 ± 0.0072	$3.098^{+0.066}_{-0.082}$	-
FM	30	0.3094 ± 0.0082	0.6711 ± 0.0071	3.060 ± 0.095	-
FM	38	0.3107 ± 0.0093	0.6741 ± 0.0075	3.01 ± 0.13	-
SF	22	0.3142 ± 0.0065	0.6684 ± 0.0072	3.117 ± 0.081	-
SF	30	0.3113 ± 0.0060	0.6709 ± 0.0076	3.09 ± 0.11	-
SF	38	0.3128 ± 0.0072	0.6730 ± 0.0083	3.07 ± 0.13	-
		n_s			
T	s	Ω_m	h	$\ln(10^{10}A_s)$	n_s
FM	22	0.307 ± 0.014	0.6746 ± 0.0076	$3.05^{+0.10}_{-0.12}$	0.975 ± 0.037
FM	30	$0.313^{+0.015}_{-0.017}$	0.6752 ± 0.0086	2.99 ± 0.15	0.967 ± 0.045
FM	38	0.314 ± 0.018	0.6773 ± 0.0089	2.96 ± 0.17	0.967 ± 0.050
SF	22	$0.326^{+0.033}_{-0.038}$	0.681 ± 0.013	$2.97^{+0.25}_{-0.29}$	0.93 ± 0.14
SF	30	$0.319^{+0.031}_{-0.037}$	$0.680^{+0.013}_{-0.015}$	2.99 ± 0.27	0.96 ± 0.14
SF	38	$0.317^{+0.031}_{-0.038}$	$0.680^{+0.013}_{-0.015}$	3.01 ± 0.30	0.97 ± 0.15

Table 5.3: We show the parameter estimation using both the *Full Modeling* and *ShapeFit* methodologies across three different minimum scales. Both methods deliver comparable results, revealing small differences in performance depending on the scale applied. Table extracted from [44].

Model	Description
`Folps-`ν	EPT/EFT with beyond-EdS kernels [41]
`PyBird`	EPT /EFT [111]
`Velocileptors`	LPT/EFT with the possibility to run EPT/EFT [113].

Table 5.4: A brief overview of the models analyzed in Fourier space. For a complete understanding of each model's specific details and performance metrics, readers should refer to the companion papers. Table extracted from [44].

approaches perform when analyzed in the same space.

In our comparative analysis, we consistently present results using both *Full Modeling* and *ShapeFit* approaches. For quick reference, Table 5.4 summarizes the main features of each code included in this comparison. Readers interested in more detailed information about these codes and their individual effectiveness should consult the companion papers, which provide comprehensive explanations and performance evaluations.

Our comparative analysis focuses on specific cases where parameter mappings are well defined across all four models. While the companion paper [113] examines both *minimal* and *maximal freedom* scenarios, we limit our current comparison to the *minimal freedom* setup only. This focused approach is necessary because each model uses different conventions for galaxy-halo bias parameters, making direct comparisons challenging. Specifically, our GSM-EFT model in its maximal freedom version lacks the tidal bias parameter discussed in [31], which means its definition of *maximal* freedom differs from the other models. Therefore, to ensure meaningful comparisons and avoid confusion, we concentrate on cases where parameter relationships are well-established and comparable.

The analysis follows specific guidelines for comparing models, with the *minimal freedom* definitions for Fourier space models clearly outlined in equations 4.15 and 4.16 of [31]. In our comparative work, we fit the average values from the Abacus mocks using our original covariance matrix without any adjustments. This approach corresponds to a volume measurement of $V = 8\,h^{-1}$Mpc, ensuring consistency across all comparisons. By maintaining these standardized conditions, we create a reliable framework for evaluating the different models' performance.

5.4.1 Comparison with EFT models in Fourier Space

The Fourier space analyses presented in this section maintain consistent settings with the companion paper, examining both monopole and quadrupole measurements up to $k_{\text{max}} = 0.25\,h\text{Mpc}^{-1}$. Our investigation compares configuration space results against the original Fourier space results from the three models described in the previous chapter: `Velocileptors`, `Folps`ν, and `PyBird`. The study uses all three galaxy tracers: LRG, ELG, and QSO, while implementing the prior constraints given in Tables 1, 2, and 3 of [31]. Through these common settings, we ensure a fair comparison across different analytical methods.

The results of our analysis are shown in Figure 5.15, where the triangular plots that compare *Full Modeling* (left panels) and *ShapeFit* (right panels) approaches compare the fittings for the LRG data and the other tracers. The Fourier space posteriors shown here match those discussed in the companion paper, as they were generated using the original codes. Our `GSM-EFT` model demonstrates strong agreement with Fourier space models, where all agree within 1-σ contours to the true values, though our constraints appear somewhat tighter. These minor variations in results are reasonable given that the analyses operate in different spaces and use distinct frameworks, and there is no direct range equivalence from distance to modes. We found comparable patterns with other tracers.

A consistent pattern can be observed in our findings, where configuration space constraints produce lower values than those from the three Fourier space codes across all tracers. Several factors might explain these differences, including variations in MCMC sampling methods, likelihood calculations, and non-identical fitting ranges between configuration and Fourier space analyses. To address these methodological differences, we conducted additional testing using `desilike` as a common framework for analyzing all Fourier space models. The outcomes of this standardized comparison between our GSM-EFT approach and the other models are presented in Figure 5.16.

The following step of our study consists of the analysis by using `desilike` to reprocess all likelihood calculations in configuration space. This additional testing helps us to verify whether previously observed differences persist when analyses are standardized through the `desilike` framework. Before proceeding with `desilike` comparisons, we confirmed that both *Full Modeling* and *ShapeFit* results maintained consistency between the original pipeline and `desilike` implementations. Here, the differences between Fourier and configuration space results remain evident in the configuration space version of Fourier space models, with these differences exceeding those found between different Fourier space pipelines.

5.4.2 Comparison with EFT models in Configuration space with `DesiLike`

In this last section, we examine how all four models perform in configuration space, having in mind that while `Velocileptors`, `Folps`ν, and `PyBird` were originally designed for Fourier space analysis, their configuration space versions are now available through fourier transforming them with `desilike`. The conversion between these spaces relies on applying the Hankel transform to Fourier space multipoles, enabling direct comparisons in configuration space, following,

$$\xi_\ell(s) = \frac{(-i)^\ell}{2\pi^2} \int dk k^2 P_\ell(k) j_\ell(ks) \tag{5.6}$$

To transform between spaces, we utilize the `fftlog` algorithm implemented within the `cosmoprimo`[3] package to perform the Hankel transform. The transformation parameters are carefully selected, with calculations ranging from $k_{min} = 10^{-4}\,h\text{Mpc}^{-1}$ to $k_{max} = 0.6\,h\text{Mpc}^{-1}$. Beyond the maximum k-value of 0.6, the transform applies a Gaussian damping effect to ensure smooth results. This mathematical approach allows for accurate conversion of model predictions between Fourier and configuration spaces.

A comprehensive comparison of model performances is illustrated in Figure 5.17, which presents our `GSM-EFT` findings alongside configuration space versions of `Velocileptors`, `Folps`ν, and `PyBird` implemented through `desilike`. The analysis displays both *Full Modeling* (left panel) and *ShapeFit* (right panel) results for the LRG sample across various

---

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

s_{min} values, showing strong agreement between models at all scales, though with slightly more variation than in Fourier space analysis, which we asosiate to the fourier transforming.

Figure 5.18 shows some triangular plots at $s_{min} = 30\,h^{-1}\mathrm{Mpc}$, revealing well-aligned contours mostly within 1 σ to the true values. Notably, configuration space analyses consistently produce smaller error bars than their Fourier space counterparts, with `GSM-EFT` showing 5-10 percent smaller errors than average. This level of agreement seems reasonable, considering `GSM-EFT` is the only model originally developed for configuration space, while others rely on multipole transformations that may introduce numerical uncertainties and increase the loss of information. Future research should explore how configuration space constraints depend on Hankel transform precision and investigate space-dependent precision differences among different models.

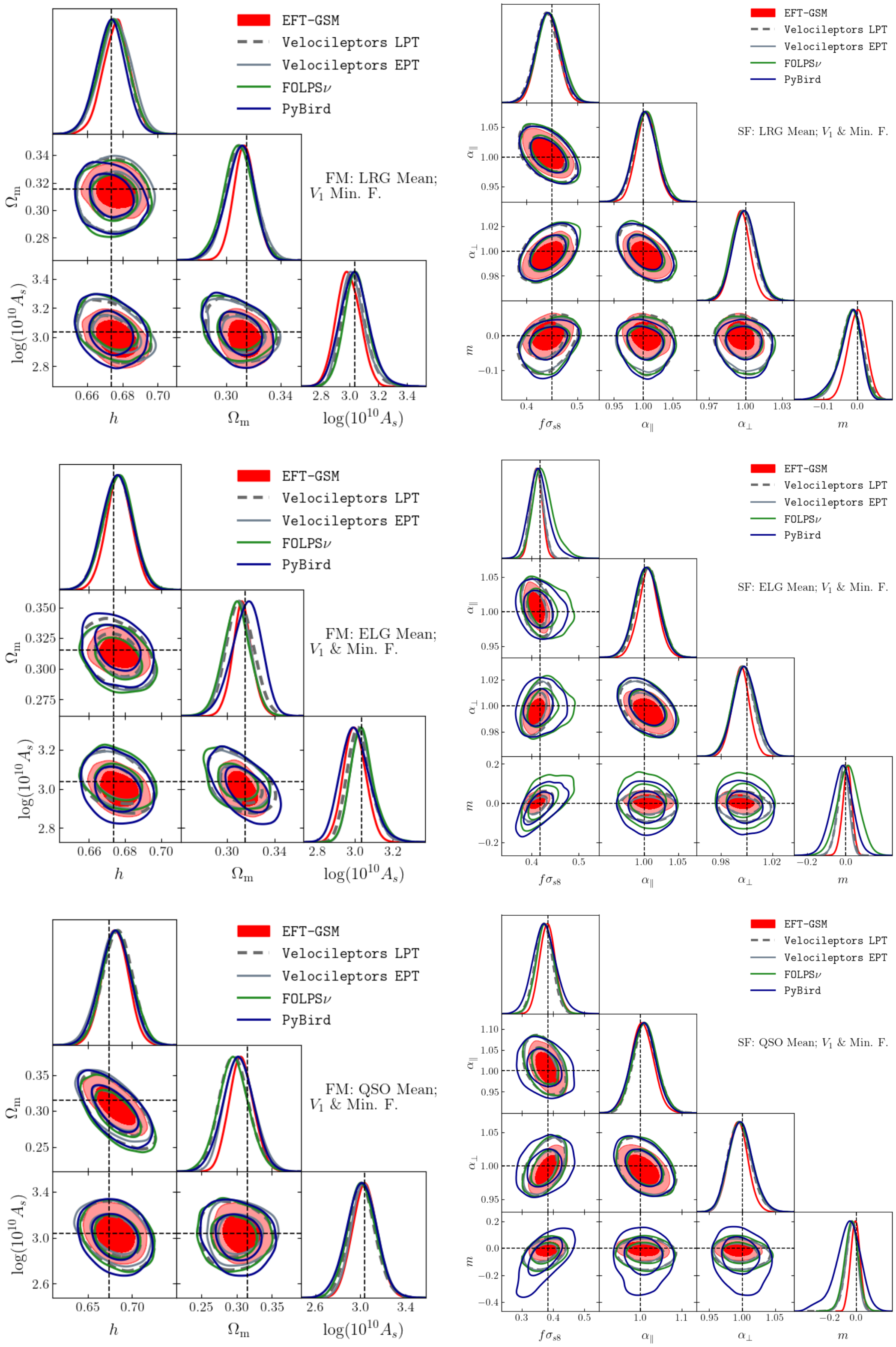

Figure 5.15: All models posteriors using their baseline analysis for all the tracers LRG , ELG and QSO. Figure extracted from [44].

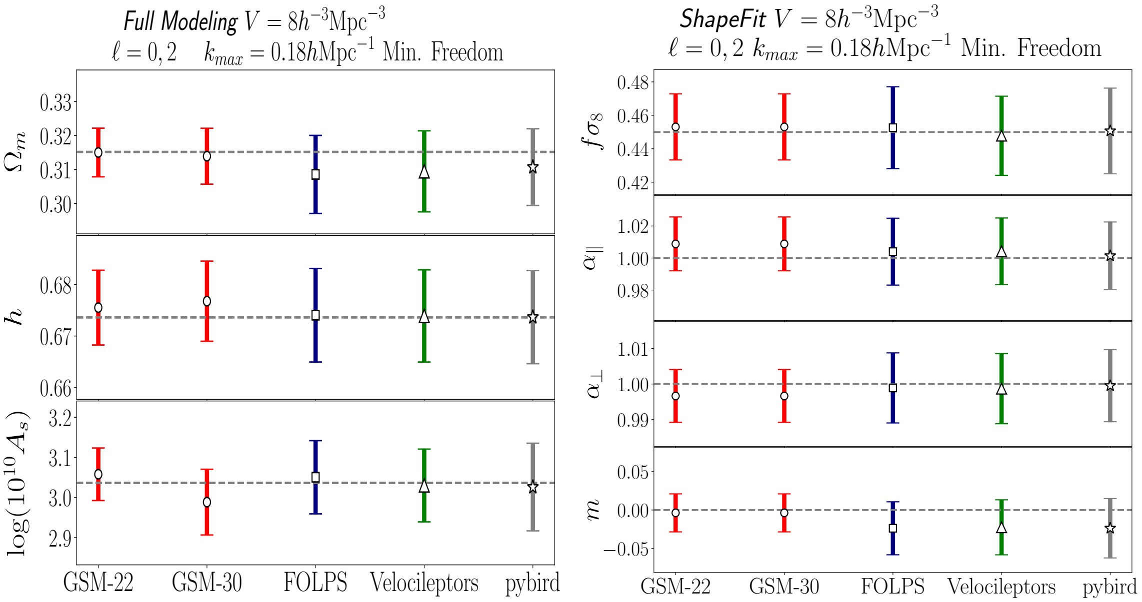

Figure 5.16: Comparison of the three models using their baseline analysis for the minimal freedom in Fourier space and the GSM model working with the same settings in configuration space. Figure extracted from [44].

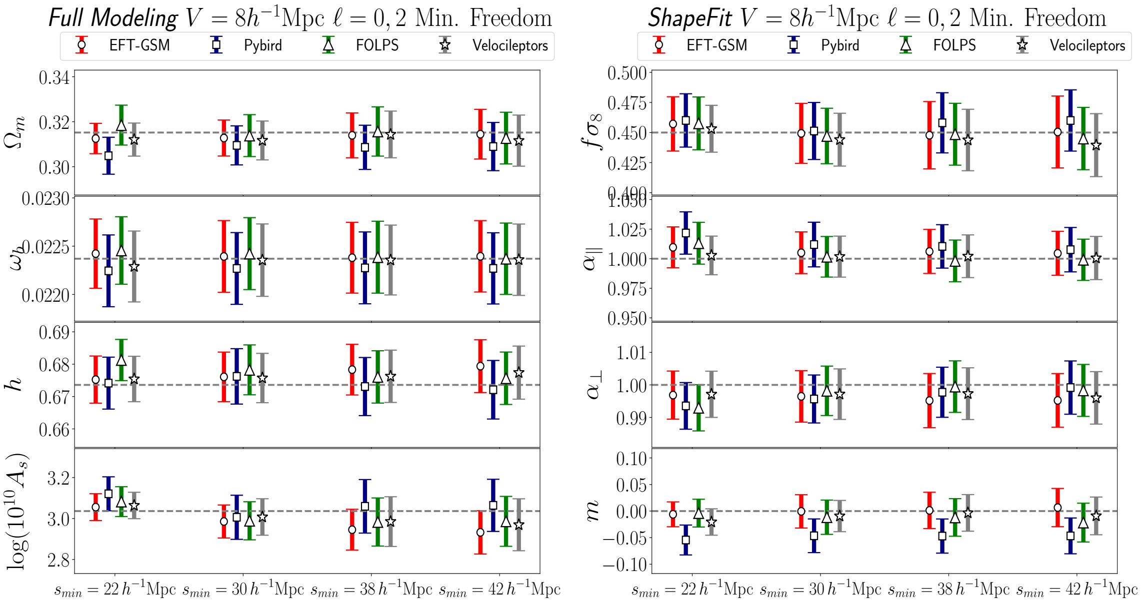

Figure 5.17: Comparison of the erros of the three models using their baseline analysis for the minimal freedom transformed to configuration space and the GSM model working with the same settings in configuration space. Figure extracted from [44].

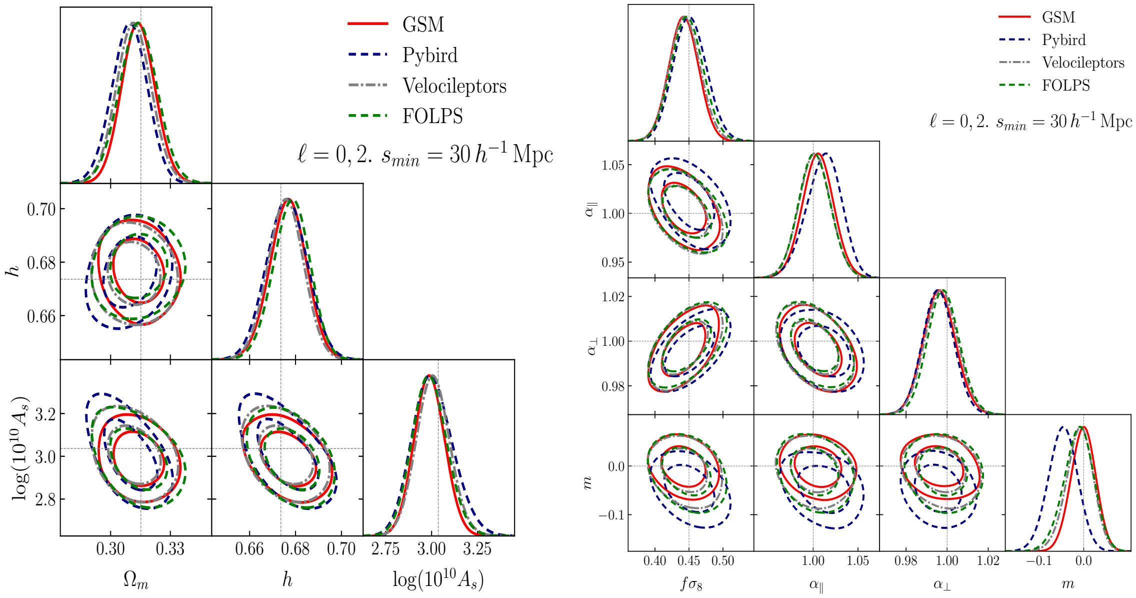

Figure 5.18: Comparison of the posteriors of the three models using their baseline analysis for the minimal freedom transformed to configuration space and the GSM model working with the same settings in configuration space. Figure extracted from [44].

Chapter 6

Conclusions

The analysis of large-scale structure (LSS) has followed two philosophically different approaches to extract cosmological information from the Universe, the compressed methodology and the full modelling methodology. Scientists have been using the compressed methodology, which employs a fixed template to examine specific features like BAO and RSD measurements. This basic approach intentionally limits itself to a few key variables, making it less dependent on specific models but also more restricted in what it can analyze. The compressed approach works with fewer number of free parameters, which makes it easier to use but less comprehensive in its scope. Recent improvements to the compressed analysis method have shown promising results, matching the precision of the second approach.

On the other hand, the full-shape modeling approach has been shown to provide tighter constraints than traditional compressed methods, although it requires significantly more computational resources. Modern developments have helped make full shape analysis more efficient, reducing some of its computational challenges. At the same time, scientists have improved the compressed method to achieve accuracy levels similar to full modeling. Today's stage IV experiments are actively comparing both approaches to find the most effective way to study the cosmos. Each method offers distinct benefits and limitations, suggesting that the best solution might involve using both techniques together. The scientific community continues to refine these methods, seeking the optimal balance between accuracy, efficiency, and practical application in cosmological research.

This work explores the two approaches to analyze the configuration space correlation function through a comprehensive methodology. While previous studies, particularly those involving recent surveys, have primarily utilized full-shape techniques in Fourier space, our work takes a new direction. The potential benefits of examining full-shape analysis within configuration space remain largely unexplored, making this investigation particularly valuable. By shifting away from traditional Fourier space methods, we aim to understand how configuration space analysis can provide fresh insights into survey data. The motivation behind this study arises from the need to diversify analytical approaches, as most modern survey examinations have been limited to Fourier space applications.

Our study implements the EFT-GSM model to examine clustering patterns in configuration space, specifically focusing on creating correlation function templates using second-order perturbation theory. Running a single instance of our model is relatively quick, taking about two seconds, but the challenge lies in the complete MCMC chain process. The full analysis would demand approximately 48 hours using 128 CPUs, which presents a considerable computational challenge. To address this time-consuming issue, we developed a solution using neural network emulators that create substitute models of our EFT-GSM templates. These neural networks prove to be remarkably efficient, reducing the processing time to just 5 minutes with the same number of CPUs (128), making the analysis much more practical and accessible.

Our research involved a two-part validation process to ensure methodology reliability. First, we examined model accuracy by testing our approach against the NSERIES simulation with an effective volume of $V_{\rm eff} = 40\ h^{-1}$Mpc, evaluating three different minimum scales $s_{\rm min} = 20,\ 30,\ 40\ h^{-1}$Mpc. The results showed that using a minimum scale of $s_{\rm min} = 20 h^{-1}$Mpc provided the best balance of accuracy and precision, though A_s predictions only aligned within 1σ at this scale. Second, we assessed the emulator's accuracy, since our work relies on neural network surrogate models rather than direct EFT-GSM calculations. Through comparison testing across various parameter space points, we found that the surrogate models performed exceptionally well, with differences typically below 1% and for some scales as low as 0.1% compared to full EFT-GSM calculations. The MCMC analysis further confirmed this accuracy, producing nearly identical parameter estimations between the surrogate and direct modeling.

Following our methodology validation, we applied our baseline analysis to examine BOSS data, specifically using the combined sample from BOSS DR12's. The study involved separate analyses of two redshift bins (z1 and z3), followed by a combined analysis of both bins. Results from the joint analysis showed improved precision in cosmological parameter measurements. Our study compares findings with seven previous analyses of BOSS data, including six full-shape studies (five in Fourier space by [55],[54], [56],[65],[66], plus one in configuration space by [70], and the Shape Fit methodology by [37]. The comparison reveals strong agreement in H_0 and A_s measurements, with all results falling within 1 σ confidence levels. Regarding Ω_m predictions, four studies showed agreement within 1σ, while three others aligned within 2σ, with discrepancies potentially explained by different prior choices for n_s. Our measurement precision matches five of the comparative studies, though two works achieved better precision by incorporating post-reconstruction power spectrum information, which our analysis did not include.

Some additional testings were conducted to understand how different prior choices affect our parameter constraints. Our investigation expanded upon the baseline analysis in two ways: first by widening the ω_b prior range to ten times its original spread of $\mathcal{N}[0.02237, 0.00037]$, and second by treating n_s as an unconstrained parameter with a flat prior between 0.5 and 1.5. The results showed distinct patterns: loosening ω_b priors led to notably reduced precision in h measurements and slightly affected ω_{cdm} constraints. Meanwhile, allowing n_s to vary had a larger impact on ω_{cdm} precision but a minimal effect on h measurements, which can be explained by the strong influence that both n_s and ω_{cdm} have on multipole slopes. These findings align with observations from previous research in this field. We also find that when comparing monopole-only analysis against our baseline study that uses both monopole and quadrupole data, the monopole-only approach demonstrated remarkable effectiveness, producing constraints on ω_{cdm} and h that nearly matched the precision and accuracy of the combined multipole analysis. This finding was unexpected, particularly since the quadrupole contains additional BAO information that theoretically should enhance the results. However, the analysis showed one notable limitation: A_s constraints became considerably less precise when using only monopole data, which aligns with theoretical predictions. These results suggest that the correlation function's monopole carries most of the crucial cosmological information.

Later, the upcoming release of the first year DESI data required a thorough evaluation of analytical tools designed for extracting cosmological insights. Our research focuses on examining various model configurations and parameters that influence predictions of galaxy two-point statistics' full shapes. A key area of interest lies in comparing different methodological approaches: the *Standard* and *ShapeFit* methods, which offer model-independent compressed analysis techniques, versus the Full Modeling approach that requires pre-selecting specific cosmological models. Understanding the strengths and limitations of these contrasting methods, particularly their relative accuracy and precision, becomes crucial for optimizing future DESI data analysis.

Within the DESI collaboration, four distinct approaches for analyzing redshift space clustering statistics were employed to study systematics. The three methods PyBird, FOLPSν, and velocileptors, which focus on analyzing the full-shape power spectrum, and our EFT-GSM methodology working configuration space served for this purpose. Our research evaluates the EFT-GSM model's effectiveness using a new set of 25 `ABACUS-SUMMIT` mocks, which simulate different matter tracers (ELGs, LRGs, and QSOs). The study includes individual analysis of each tracer type and a combined sample assessment, with results compared against the other three methods' predictions converted to configuration space. Implementation of our approach relies on a newly optimized C-code tool for calculating configuration space multipoles within the GSM-EFT model in an extended space, including 2 extra parameters. To enhance computational efficiency, we incorporated our neural network surrogate models, significantly reducing the time needed to derive cosmological constraints.

Our study tested the Full Modeling approach's accuracy in recovering different tracers from ABACUS-SUMMIT simulations across different volume sizes. The analysis examined three specific volumes: $8, h^{-3}$Gpc^3 (representing DESI's first-year observations), $40, h^{-3}$Gpc^3 (approximating DESI's fifth-year data), and $200, h^{-3}$Gpc^3 (exceeding DESI's volume). Results demonstrated the methodology's consistency across these varying volumes, with most parameters falling within 1σ of their true values. In this work, we focus on results equivalent to the first year volume of DESI.

An evaluation of scale-dependent effects in our LRG sample analysis compared both *Full Modeling* and *Shapefit* approaches across three minimum scale values: $s_{\min} = 22, h^{-1}$Mpc. Both methods showed strong agreement with each other and with simulation values, maintaining accuracy within 1σ across all tested scales. Since the smallest scale $s_{\min} = 22, h^{-1}$Mpc provided the most precise constraints, we adopted this as our baseline for subsequent analyses. Additionally, we investigated the impact of parameter freedom by comparing scenarios where b_{s^2} was either unconstrained or fixed at zero, with the unconstrained case serving as our maximal freedom definition, serving at the same time as our baseline configuration. While both approaches produced consistent results within 1σ for V_1, the constrained case showed approximately 10% tighter parameter constraints.

In this work, we have also analyzed several extensions to our standard configuration. The first of these extensions is the inclusion of the hexadecapole of the correlation function into our analysis, which is considered as the multipole where first-loop PT starts to find difficulties modeling our scales of interest. In this extension, we include a model of the hexadecapole but only down to a minimum scale of $s_{\min} = 30\, h^{-1}$Mpc as we do not trust our model at smaller scales. We note that this extension does not significantly improve our results, with its only advantage being a small improvement in the precision of the *Full Modeling* estimates. We also highlight that all predictions recover the true value of all parameters within 1σ.

We also tested variations of our baseline analysis, like incorporating the spectral index n_s as an additional parameter in our analysis of the LRG sample, specifically using the Full Modeling configuration. The extended model successfully predicted all parameters, including n_s, maintaining accuracy within 1σ confidence level. Interestingly, this modification enhanced the recovery of original parameters compared to baseline results across all scales. The analysis revealed that a minimum scale of $s_{\min} = 30, h^{-1}$Mpc produced optimal results when using this extended parameter.

Our research explored an extended model incorporating the Chevallier-Polarski-Linder parameterization of dark energy, focusing on the parameters w_0 and w_a. The analysis involved two parts using the LRG sample: an initial phase examining only w_0 variations (with w_a fixed at its known value), followed by the whole analysis where both parameters were allowed to vary. The Full Modeling approach demonstrated excellent parameter recovery from the ABACUS-SUMMIT simulations, with both w_0 and w_a predictions remaining within 1σ confidence.The exceptional accuracy achieved can be partially attributed to our configuration space methodology, which minimizes the influence of volume effects on dark energy parameter estimations.

Moving beyond our LRG sample analyses, we extended our methodology testing to include QSO and ELG samples from `ABACUS-SUMMIT` simulations. The evaluation applied to our standard configuration across three different minimum scales for new tracers reveals that the ELG analyses are better in accuracy and precision compared to LRG results, while QSO analyses demonstrated lower precision overall. These variations in precision align with expectations based on the different tracer populations in each sample. Despite these differences, our methodology proved robust, successfully recovering all parameters within 1σ confidence levels across all three sample types.

An additional comparative analysis was conducted between Full Modeling and ShapeFit approaches across various cosmological parameters, examining different configurations including bias settings and minimum scale selections. The study also evaluated three model extensions, incorporating hexadecapole, minimal and maximal freedom, and the extension for the n_s parameter. Standard model configurations showed consistent central value estimates between both methods, though error estimates varied by 30% depending on specific parameters and test cases. More significant differences emerged when testing extended models with additional cosmological parameters, affecting both central values and uncertainty ranges. A notable finding revealed that the Full Modeling approach provided tighter constraints than ShapeFit when n_s was treated as a free parameter, particularly in constraining n_s itself.

Our final investigation compared our configuration space methodology `GSM-EFT` with three Fourier space codes: `Velocileptors`, `Folps`ν, and `PyBird`. This comparison complements a broader analysis of RSD methodologies detailed in a companion paper. The results demonstrated strong consistency between our `GSM-EFT` approach and all Fourier space models, with our method achieving smaller uncertainty ranges in both Full Modeling and ShapeFit configurations. Further validation using the desilike framework confirmed the compatibility of results across both spaces while highlighting the enhanced precision achieved in configuration space analysis. These findings support the effectiveness of our configuration space approach in cosmological parameter estimation.

All of this work contributed to the development of the methodology that was used in the cosmological analysis of the DESI Year 1 full-shape data. Through the testing of simulations, comparison of modeling approaches, and validation of redshift-space clustering statistics, a robust framework has been established to extract precise cosmological information from the observed LSS.

Bibliography

[1] J. T. A. de Jong, G. A. Verdoes Kleijn, K. H. Kuijken and E. A. Valentijn, *The Kilo-Degree Survey*, *Experimental Astronomy* **35** (Jan., 2013) 25–44, [1206.1254].

[2] T. W. Shimwell, C. L. Hale, P. N. Best, A. Botteon, A. Drabent, M. J. Hardcastle et al., *The LOFAR Two-metre Sky Survey: Deep Fields Data Release 2. I. The ELAIS-N1 field*, *arXiv e-prints* (Jan., 2025) arXiv:2501.04093, [2501.04093].

[3] M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole et al., *The 2dF Galaxy Redshift Survey: spectra and redshifts*, **328** (Dec., 2001) 1039–1063, [astro-ph/0106498].

[4] W. J. Percival, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, S. Cole et al., *The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe*, **327** (Nov., 2001) 1297–1306, [astro-ph/0105252].

[5] K. S. Dawson, D. J. Schlegel, C. P. Ahn, S. F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, **145** (Jan., 2013) 10, [1208.0022].

[6] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, **470** (Sept., 2017) 2617–2652, [1607.03155].

[7] K. S. Dawson, J.-P. Kneib, W. J. Percival, S. Alam, F. D. Albareti, S. F. Anderson

et al., *The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, **151** (Feb., 2016) 44, [1508.04473].

[8] S. Alam, M. Aubert, S. Avila, C. Balland, J. E. Bautista, M. A. Bershady et al., *Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory*, **103** (Apr., 2021) 083533, [2007.08991].

[9] J. Zuntz, E. Sheldon, S. Samuroff, M. A. Troxel, M. Jarvis, N. MacCrann et al., *Dark Energy Survey Year 1 results: weak lensing shape catalogues*, **481** (Nov., 2018) 1149–1182, [1708.01533].

[10] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen et al., *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, **98** (Aug., 2018) 043526, [1708.01530].

[11] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *arXiv e-prints* (Oct., 2011) arXiv:1110.3193, [1110.3193].

[12] G. D. Racca, R. Laureijs, L. Stagnaro, J.-C. Salvignol, J. Lorenzo Alvarez, G. Saavedra Criado et al., *The Euclid mission design*, in *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave* (H. A. MacEwen, G. G. Fazio, M. Lystrup, N. Batalha, N. Siegler and E. C. Tong, eds.), vol. 9904 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 99040O, July, 2016. 1610.05508. DOI.

[13] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman et al., *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, **873** (Mar., 2019) 111, [0805.2366].

[14] DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, **164** (Nov., 2022) 207, [2205.10939].

[15] A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu et al., *Report of the Dark Energy Task Force*, *arXiv e-prints* (Sept., 2006) astro–ph/0609591, [`astro-ph/0609591`].

[16] S. Ramirez, M. Icaza-Lizaola, S. Fromenteau, M. Vargas-Magaña and A. Aviles, *Full Shape Cosmology Analysis from BOSS in configuration space using Neural Network Acceleration*, *arXiv e-prints* (Oct., 2023) arXiv:2310.17834, [`2310.17834`].

[17] B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro et al., *SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues*, **455** (Jan., 2016) 1553–1573, [`1509.06529`].

[18] C. Uhlemann, M. Kopp and T. Haugg, *Edgeworth streaming model for redshift space distortions*, *Phys. Rev. D* **92** (2015) 063004, [`1503.08837`].

[19] Z. Vlah, E. Castorina and M. White, *The Gaussian streaming model and convolution Lagrangian effective field theory*, *JCAP* **12** (2016) 007, [`1609.02908`].

[20] G. Valogiannis, R. Bean and A. Aviles, *An accurate perturbative approach to redshift space clustering of biased tracers in modified gravity*, **2020** (Jan., 2020) 055, [`1909.05261`].

[21] T. Matsubara, *Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture*, *Phys. Rev. D* **78** (2008) 083519, [`0807.1733`].

[22] A. Aviles, *Renormalization of Lagrangian bias via spectral parameters*, *Phys. Rev. D* **98** (2018) 083541, [`1805.05304`].

[23] V. Desjacques, D. Jeong and F. Schmidt, *Large-Scale Galaxy Bias*, *Phys. Rept.* **733** (2018) 1–193, [`1611.09787`].

[24] G. Valogiannis, R. Bean and A. Aviles, *An accurate perturbative approach to redshift space clustering of biased tracers in modified gravity*, *JCAP* **2001** (2020) 055, [`1909.05261`].

[25] T. Matsubara, *Resumming Cosmological Perturbations via the Lagrangian Picture: One-loop Results in Real Space and in Redshift Space*, *Phys. Rev. D* **77** (2008) 063530, [0711.2521].

[26] J. Carlson, B. Reid and M. White, *Convolution Lagrangian perturbation theory for biased tracers*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 1674, [1209.0780].

[27] Z. Vlah, M. White and A. Aviles, *A Lagrangian effective field theory*, *JCAP* **09** (2015) 014, [1506.05264].

[28] Z. Vlah and M. White, *Exploring redshift-space distortions in large-scale structure*, *JCAP* **1903** (2019) 007, [1812.02775].

[29] S. Ramirez-Solano, M. Icaza-Lizaola, H. E. Noriega, M. Vargas-Magaña, S. Fromenteau, A. Aviles et al., *Full modeling and parameter compression methods in configuration space for desi 2024 and beyond*, *arXiv e-prints* (2024) arXiv:2404.07268, [2404.07268].

[30] G. d'Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang et al., *The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, **2020** (May, 2020) 005, [1909.05271].

[31] M. Maus, Y. Lai, H. E. Noriega, S. Ramirez-Solano, A. Aviles, S. Chen et al., *A comparison of effective field theory models of redshift space galaxy power spectra for desi 2024 and future surveys*, *arXiv e-prints* (2024) arXiv:2404.07272, [2404.07272].

[32] S.-F. Chen, Z. Vlah and M. White, *Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory*, **2020** (July, 2020) 062, [2005.00523].

[33] H. E. Noriega, A. Aviles, S. Fromenteau and M. Vargas-Magaña, *Fast computation of non-linear power spectrum in cosmologies with massive neutrinos*, **2022** (Nov., 2022) 038, [2208.02791].

[34] A. Taruya, T. Nishimichi and S. Saito, *Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory*, **82** (Sept., 2010) 063522, [1006.0699].

[35] S. Ramirez, M. Icaza-Lizaola, S. Fromenteau, M. Vargas-Magaña and A. Aviles, *Full Shape Cosmology Analysis from BOSS in configuration space using Neural Network Acceleration*, *arXiv e-prints* (Oct., 2023) arXiv:2310.17834, [2310.17834].

[36] S. Brieden, H. Gil-Marín and L. Verde, *ShapeFit: extracting the power spectrum shape information in galaxy surveys beyond BAO and RSD*, **2021** (Dec., 2021) 054, [2106.07641].

[37] S. Brieden, H. Gil-Marín and L. Verde, *Model-independent versus model-dependent interpretation of the SDSS-III BOSS power spectrum: Bridging the divide*, **104** (Dec., 2021) L121301, [2106.11931].

[38] S. Brieden, H. Gil-Marín and L. Verde, *PT challenge: validation of ShapeFit on large-volume, high-resolution mocks*, **2022** (June, 2022) 005, [2201.08400].

[39] S. Brieden, H. Gil-Marín and L. Verde, *Model-agnostic interpretation of 10 billion years of cosmic evolution traced by BOSS and eBOSS data*, **2022** (Aug., 2022) 024, [2204.11868].

[40] S. Brieden, H. Gil-Marín and L. Verde, *ShapeFit: extracting the power spectrum shape information in galaxy surveys beyond BAO and RSD*, **2021** (Dec., 2021) 054, [2106.07641].

[41] H. E. Noriega, A. Aviles, H. Gil-Marín, S. Ramirez-Solano, S. Fromenteau, M. Vargas-Magaña et al., *Comparing compressed and full-modeling analyses with folps: Implications for desi 2024 and beyond*, *arXiv e-prints* (2024) arXiv:2404.07269, [2404.07269].

[42] S.-F. Chen, Z. Vlah and M. White, *Consistent Modeling of Velocity Statistics and Redshift-Space Distortions in One-Loop Perturbation Theory*, *JCAP* **07** (2020) 062, [2005.00523].

[43] S.-F. Chen, Z. Vlah, E. Castorina and M. White, *Redshift-Space Distortions in Lagrangian Perturbation Theory*, *JCAP* **03** (2021) 100, [`2012.04636`].

[44] S. Ramirez-Solano, M. Icaza-Lizaola, H. E. Noriega, M. Vargas-Magaña, S. Fromenteau, A. Aviles et al., *Full Modeling and parameter compression methods in configuration space for DESI 2024 and beyond*, **2025** (Jan., 2025) 129, [`2404.07268`].

[45] M. Karamanis and F. Beutler, *Ensemble slice sampling: Parallel, black-box and gradient-free inference for correlated multimodal distributions*, *arXiv preprint arXiv: 2002.06212* (2020) .

[46] M. Karamanis, F. Beutler and J. A. Peacock, *zeus: A python implementation of ensemble slice sampling for efficient bayesian parameter inference*, *arXiv preprint arXiv:2105.03468* (2021) .

[47] J. Goodman and J. Weare, *Ensemble samplers with affine invariance*, *Communications in Applied Mathematics and Computational Science* **5** (Jan., 2010) 65–80.

[48] D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, **125** (Mar., 2013) 306, [`1202.3665`].

[49] P. McDonald, *Clustering of dark matter tracers: Renormalizing the bias parameters*, *Phys. Rev. D* **74** (2006) 103512, [`astro-ph/0609413`].

[50] P. McDonald and A. Roy, *Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS*, *JCAP* **08** (2009) 020, [`0902.0991`].

[51] D. Baumann, A. Nicolis, L. Senatore and M. Zaldarriaga, *Cosmological Non-Linearities as an Effective Fluid*, *JCAP* **07** (2012) 051, [`1004.2488`].

[52] J. J. M. Carrasco, M. P. Hertzberg and L. Senatore, *The Effective Field Theory of Cosmological Large Scale Structures*, *JHEP* **09** (2012) 082, [`1206.2926`].

[53] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, *Large scale structure of the universe and cosmological perturbation theory*, *Phys. Rept.* **367** (2002) 1–248, [`astro-ph/0112551`].

[54] M. M. Ivanov, M. Simonović and M. Zaldarriaga, *Cosmological Parameters from the BOSS Galaxy Power Spectrum*, *JCAP* **05** (2020) 042, [`1909.05277`].

[55] G. D'Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang et al., *The Cosmological Analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, *JCAP* **05** (2020) 005, [`1909.05271`].

[56] O. H. E. Philcox, M. M. Ivanov, M. Simonović and M. Zaldarriaga, *Combining full-shape and BAO analyses of galaxy power spectra: a 1.6% CMB-independent constraint on H_0*, **2020** (May, 2020) 032, [`2002.04035`].

[57] S.-F. Chen, Z. Vlah and M. White, *A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO*, *JCAP* **02** (2022) 008, [`2110.05530`].

[58] O. H. E. Philcox and M. M. Ivanov, *BOSS DR12 full-shape cosmology: ΛCDM constraints from the large-scale galaxy power spectrum and bispectrum monopole*, *Phys. Rev. D* **105** (2022) 043517, [`2112.04515`].

[59] O. H. E. Philcox, M. M. Ivanov, G. Cabass, M. Simonović, M. Zaldarriaga and T. Nishimichi, *Cosmology with the redshift-space galaxy bispectrum monopole at one-loop order*, *Phys. Rev. D* **106** (2022) 043530, [`2206.02800`].

[60] A. Chudaykin, K. Dolgikh and M. M. Ivanov, *Constraints on the curvature of the Universe and dynamical dark energy from the Full-shape and BAO data*, *Phys. Rev. D* **103** (2021) 023507, [`2009.10106`].

[61] G. D'Amico, Y. Donath, L. Senatore and P. Zhang, *Limits on Clustering and Smooth Quintessence from the EFTofLSS*, `2012.07554`.

[62] P. Carrilho, C. Moretti and A. Pourtsidou, *Cosmology with the EFTofLSS and BOSS: dark energy constraints and a note on priors*, *JCAP* **01** (2023) 028, [2207.14784].

[63] L. Piga, M. Marinucci, G. D'Amico, M. Pietroni, F. Vernizzi and B. S. Wright, *Constraints on modified gravity from the BOSS galaxy survey*, *JCAP* **04** (2023) 038, [2211.12523].

[64] A. He, R. An, M. M. Ivanov and V. Gluscevic, *Self-Interacting Neutrinos in Light of Large-Scale Structure Data*, 2309.03956.

[65] T. Tröster, A. G. Sánchez, M. Asgari, C. Blake, M. Crocce, C. Heymans et al., *Cosmology from large-scale structure. Constraining ΛCDM with BOSS*, **633** (Jan., 2020) L10, [1909.11006].

[66] S.-F. Chen, Z. Vlah and M. White, *A new analysis of galaxy 2-point functions in the BOSS survey, including full-shape information and post-reconstruction BAO*, **2022** (Feb., 2022) 008, [2110.05530].

[67] A. Semenaite, A. G. Sánchez, A. Pezzotta, J. Hou, A. Eggemeier, M. Crocce et al., *Beyond - ΛCDM constraints from the full shape clustering measurements from BOSS and eBOSS*, **521** (June, 2023) 5013–5025, [2210.07304].

[68] R. Neveux, E. Burtin, V. Ruhlmann-Kleider, A. de Mattia, A. Semenaite, K. S. Dawson et al., *Combined full shape analysis of BOSS galaxies and eBOSS quasars using an iterative emulator*, **516** (Oct., 2022) 1910–1922, [2201.04679].

[69] M. Maus, S.-F. Chen and M. White, *A comparison of template vs. direct model fitting for redshift-space distortions in BOSS*, **2023** (June, 2023) 005, [2302.07430].

[70] P. Zhang, G. D'Amico, L. Senatore, C. Zhao and Y. Cai, *BOSS Correlation Function analysis from the Effective Field Theory of Large-Scale Structure*, **2022** (Feb., 2022) 036, [2110.07539].

[71] J. DeRose, S.-F. Chen, M. White and N. Kokron, *Neural network acceleration of large-scale structure theory calculations*, **2022** (Apr., 2022) 056, [2112.05889].

[72] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, **538** (2000) 473–476, [astro-ph/9911177].

[73] A. Lewis and S. Bridle, *Cosmological parameters from CMB and other data: A Monte Carlo approach*, **66** (2002) 103511, [astro-ph/0205436].

[74] K. Heitmann, D. Higdon, M. White, S. Habib, B. J. Williams, E. Lawrence et al., *The Coyote Universe. II. Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum*, **705** (Nov., 2009) 156–174, [0902.0429].

[75] A. Korobov, *The approximate computation of multiple integrals*, in *Dokl. Akad. Nauk SSSR*, vol. 124, pp. 1207–1210, 1959.

[76] J. Alsing, H. Peiris, J. Leja, C. Hahn, R. Tojeiro, D. Mortlock et al., *SPECULATOR: Emulating Stellar Population Synthesis for Fast and Accurate Galaxy Spectra and Photometry*, **249** (July, 2020) 5, [1911.11778].

[77] V. Springel, *The Cosmological simulation code GADGET-2*, *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105–1134, [astro-ph/0505010].

[78] F. S. Kitaura, G. Yepes and F. Prada, *Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing.*, **439** (Mar., 2014) L21–L25, [1307.3285].

[79] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release*, **456** (Mar., 2016) 4156–4173, [1509.06400].

[80] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson et al., *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe,*

the Milky Way, and Extra-Solar Planetary Systems, **142** (Sept., 2011) 72, [1101.1529].

[81] S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton et al., *The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III*, **219** (July, 2015) 12, [1501.00963].

[82] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger et al., *The 2.5 m Telescope of the Sloan Digital Sky Survey*, **131** (Apr., 2006) 2332–2359, [astro-ph/0602326].

[83] S. A. Smee, J. E. Gunn, A. Uomoto, N. Roe, D. Schlegel, C. M. Rockosi et al., *The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey*, **146** (Aug., 2013) 32, [1208.2233].

[84] A. S. Bolton, D. J. Schlegel, É. Aubourg, S. Bailey, V. Bhardwaj, J. R. Brownstein et al., *Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey*, **144** (Nov., 2012) 144, [1207.7326].

[85] L. Anderson, É. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, **441** (June, 2014) 24–62, [1312.4877].

[86] A. J. Ross, F. Beutler, C.-H. Chuang, M. Pellejero-Ibanez, H.-J. Seo, M. Vargas-Magaña et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function*, **464** (Jan., 2017) 1168–1191, [1607.03145].

[87] L. Anderson, É. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, **441** (June, 2014) 24–62, [1312.4877].

[88] Euclid Collaboration, Y. Mellier, Abdurro'uf, J. A. Acevedo Barroso, A. Achúcarro, J. Adamek et al., *Euclid. I. Overview of the Euclid mission*, *arXiv e-prints* (May, 2024) arXiv:2405.13491, [`2405.13491`].

[89] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen et al., *The DESI Experiment Part I: Science,Targeting, and Survey Design*, *arXiv e-prints* (Oct., 2016) arXiv:1611.00036, [`1611.00036`].

[90] DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, **164** (Nov., 2022) 207, [`2205.10939`].

[91] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering et al., *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, **167** (Feb., 2024) 62, [`2306.06307`].

[92] DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering et al., *The Early Data Release of the Dark Energy Spectroscopic Instrument*, *arXiv e-prints* (June, 2023) arXiv:2306.06308, [`2306.06308`].

[93] D. J. Schlegel, S. Ferraro, G. Aldering, C. Baltay, S. BenZvi, R. Besuner et al., *A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5*, *arXiv e-prints* (Sept., 2022) arXiv:2209.03585, [`2209.03585`].

[94] A. Raichoor, D. J. Eisenstein, T. Karim, J. A. Newman, J. Moustakas, D. D. Brooks et al., *Preliminary Target Selection for the DESI Emission Line Galaxy (ELG) Sample*, *Research Notes of the American Astronomical Society* **4** (Oct., 2020) 180, [`2010.11281`].

[95] R. Zhou, J. A. Newman, K. S. Dawson, D. J. Eisenstein, D. D. Brooks, A. Dey et al., *Preliminary Target Selection for the DESI Luminous Red Galaxy (LRG) Sample*, *Research Notes of the American Astronomical Society* **4** (Oct., 2020) 181, [`2010.11282`].

[96] C. Yèche, N. Palanque-Delabrouille, C.-A. Claveau, D. D. Brooks, E. Chaussidon, T. M. Davis et al., *Preliminary Target Selection for the DESI Quasar (QSO) Sample*, *Research Notes of the American Astronomical Society* **4** (Oct., 2020) 179, [2010.11280].

[97] B. A. Reid, L. Samushia, M. White, W. J. Percival, M. Manera, N. Padmanabhan et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at $z = 0.57$ from anisotropic clustering*, **426** (Nov., 2012) 2719–2737, [1203.6641].

[98] L. Samushia, B. A. Reid, M. White, W. J. Percival, A. J. Cuesta, G.-B. Zhao et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring growth rate and geometry with anisotropic clustering*, **439** (Apr., 2014) 3504–3519, [1312.4899].

[99] S. Satpathy, S. Alam, S. Ho, M. White, N. A. Bahcall, F. Beutler et al., *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: on the measurement of growth rate using galaxy correlation functions*, **469** (Aug., 2017) 1369–1382, [1607.03148].

[100] A. Tamone, A. Raichoor, C. Zhao, A. de Mattia, C. Gorgoni, E. Burtin et al., *The completed SDSS-IV extended baryon oscillation spectroscopic survey: growth rate of structure measurement from anisotropic clustering analysis in configuration space between redshift 0.6 and 1.1 for the emission-line galaxy sample*, **499** (Dec., 2020) 5527–5546, [2007.09009].

[101] S.-F. Chen, Z. Vlah and M. White, *Consistent modeling of velocity statistics and redshift-space distortions in one-loop perturbation theory*, **2020** (July, 2020) 062, [2005.00523].

[102] Y. Zhang, A. R. Pullen, S. Alam, S. Singh, E. Burtin, C.-H. Chuang et al., *Testing general relativity on cosmological scales at redshift $z \sim 1.5$ with quasar and CMB lensing*, **501** (Feb., 2021) 1013–1027, [2007.12607].

[103] J. E. Bautista, R. Paviot, M. Vargas Magaña, S. de la Torre, S. Fromenteau, H. Gil-Marín et al., *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1*, **500** (Jan., 2021) 736–762, [2007.08993].

[104] L. Fonseca de la Bella, D. Regan, D. Seery and S. Hotchkiss, *The matter power spectrum in redshift space using effective field theory*, **2017** (Nov., 2017) 039, [1704.05309].

[105] V. Desjacques, D. Jeong and F. Schmidt, *The galaxy power spectrum and bispectrum in redshift space*, **2018** (Dec., 2018) 035, [1806.04015].

[106] A. Chudaykin, M. M. Ivanov, O. H. E. Philcox and M. Simonović, *Nonlinear perturbation theory extension of the Boltzmann code CLASS*, **102** (Sept., 2020) 063533, [2004.10607].

[107] G. d'Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang et al., *The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, **2020** (May, 2020) 005, [1909.05271].

[108] N. A. Maksimova, L. H. Garrison, D. J. Eisenstein, B. Hadzhiyska, S. Bose and T. P. Satterthwaite, *ABACUSSUMMIT: a massive set of high-accuracy, high-resolution N-body simulations*, **508** (Dec., 2021) 4017–4037, [2110.11398].

[109] T. Colas, G. d'Amico, L. Senatore, P. Zhang and F. Beutler, *Efficient cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, **2020** (June, 2020) 001, [1909.07951].

[110] G. D'Amico, L. Senatore and P. Zhang, *Limits on wCDM from the EFTofLSS with the PyBird code*, **2021** (Jan., 2021) 006, [2003.07956].

[111] Y. Lai, C. Howlett, M. Maus, H. Gil-Marín, H. E. Noriega, S. Ramírez-Solano et al., *A comparison between Shapefit compression and Full-Modelling method with*

PyBird for DESI 2024 and beyond, *arXiv e-prints* (2024) arXiv:2404.07283, [2404.07283].

[112] S.-F. Chen, Z. Vlah, E. Castorina and M. White, *Redshift-space distortions in Lagrangian perturbation theory*, **2021** (Mar., 2021) 100, [2012.04636].

[113] M. Maus, S. Chen, M. White, J. Aguilar, S. Ahlen, A. Aviles et al., *An analysis of parameter compression and full-modeling techniques with Velocileptors for DESI 2024 and beyond*, *arXiv e-prints* (2024) arXiv:2404.07312, [2404.07312].

[114] A. Raichoor, A. de Mattia, A. J. Ross, C. Zhao, S. Alam, S. Avila et al., *The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: large-scale structure catalogues and measurement of the isotropic BAO between redshift 0.6 and 1.1 for the Emission Line Galaxy Sample*, **500** (Jan., 2021) 3254–3274, [2007.09007].

[115] A. J. Ross, J. Bautista, R. Tojeiro, S. Alam, S. Bailey, E. Burtin et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Large-scale structure catalogues for cosmological analysis*, **498** (Oct., 2020) 2354–2371, [2007.09000].

[116] B. W. Lyke, A. N. Higley, J. N. McLane, D. P. Schurhammer, A. D. Myers, A. J. Ross et al., *The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release*, **250** (Sept., 2020) 8, [2007.09001].

[117] C. Alcock and B. Paczynski, *An evolution free test for non-zero cosmological constant*, **281** (Oct., 1979) 358.

[118] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, **641** (Sept., 2020) A6, [1807.06209].

[119] A. Rocher, V. Ruhlmann-Kleider, E. Burtin and A. de Mattia, *Halo occupation distribution of Emission Line Galaxies: fitting method with Gaussian processes*, **2023** (May, 2023) 033, [2302.07056].

[120] S. Yuan, H. Zhang, A. J. Ross, J. Donald-McCann, B. Hadzhiyska, R. H. Wechsler et al., *The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit*, *arXiv e-prints* (June, 2023) arXiv:2306.06314, [`2306.06314`].

[121] C.-H. Chuang, F.-S. Kitaura, F. Prada, C. Zhao and G. Yepes, *EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics*, *Monthly Notices of the Royal Astronomical Society* **446** (11, 2014) 2621–2628, [`https://academic.oup.com/mnras/article-pdf/446/3/2621/10448149/stu2301.pdf`].

[122] D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306–312, [`1202.3665`].

[123] C. Grove, C.-H. Chuang, N. C. Devi, L. Garrison, B. L'Huillier, Y. Feng et al., *The DESI N-body simulation project - I. Testing the robustness of simulations for the DESI dark time survey*, **515** (Sept., 2022) 1854–1870, [`2112.09138`].

[124] M. Chevallier and D. Polarski, *Accelerating Universes with Scaling Dark Matter*, *International Journal of Modern Physics D* **10** (Jan., 2001) 213–223, [`gr-qc/0009008`].

[125] E. V. Linder, *Exploring the Expansion History of the Universe*, **90** (Mar., 2003) 091301, [`astro-ph/0208512`].