Alma Mater Studiorum · Università di Bologna

SCHOOL OF SCIENCE
Department of Physics and Astronomy
Master Degree Programme in Astrophysics and Cosmology

Accelerating linear-theory clustering models: a new evolution-mapping based emulator

Graduation Thesis

Supervisor:
Prof. Federico Marulli
Co-Supervisor:
Dr. Sofia Contarini

Submitted by:
Emmanuel Agati

Academic Year 2025–2026

Contents

Abstract

Large-scale structure observations provide one of the most powerful probes of the cosmological model. The extraction of cosmological information from clustering measurements relies on Bayesian parameter inference, which typically requires tens of thousands to millions of evaluations of theoretical predictions, such as the linear matter power spectrum. Although Einstein–Boltzmann solvers, including `CAMB` and `CLASS`, provide highly accurate predictions, their computational cost represents a major bottleneck for modern cosmological analyses.

In this Thesis, a neural-network emulator for the linear matter power spectrum is developed and integrated within the `CosmoBolognaLib` framework. The proposed approach exploits the evolution-mapping technique, which separates the cosmological dependence of the power spectrum into components associated with its shape and redshift evolution. This allows a large fraction of the parameter dependence to be treated analytically or through physically motivated rescalings, while only the residual dependence is learned by the neural network.

As a result, the emulator was trained in a significantly reduced parameter space, depending only on $\Omega_{\rm b}$ and $\Omega_{\rm cdm}$, while retaining the ability to reconstruct predictions for a broader set of cosmological models. The training was performed on a set of 80 000 linear matter power spectra generated with `CAMB`, covering the parameter ranges $\Omega_{\rm b} \in [0.03, 0.06]$ and $\Omega_{\rm cdm} \in [0.07, 0.70]$. The spectra were sampled over the interval $k \in [1.5 \times 10^{-4}, 75]\,\mathrm{Mpc}^{-1}$, adopting a finer sampling around the baryon acoustic oscillation scales to improve the description of the oscillatory features. Its accuracy was first evaluated on an independent test set of 20 000 previously unseen power spectra, verifying that the deviations from the reference `CAMB` predictions were statistically insignificant in the context of cosmological analyses.

The effective performance was then determined through complete Bayesian analyses of mock clustering measurements in ΛCDM, $k\Lambda$CDM, and w_0w_aCDM cosmologies. Overall, all the validation tests demonstrated residuals below the 0.1% level over the adopted range of scales and redshifts, showing posterior distributions statistically consistent with those obtained using direct `CAMB` evaluations. The resulting speed-up of Bayesian analyses reaches approximately three orders of magnitude in each case, reducing the execution time of a typical cosmological analysis from several tens of hours to about one minute on identical computational resources.

These results demonstrate that physically motivated emulation techniques can substantially accelerate cosmological parameter inference without compromising accuracy. The integration of the emulator within `CosmoBolognaLib` provides users with a powerful tool for all cosmological probes relying on the linear matter power spectrum, including, for example, clustering analyses based on correlation-function multipoles, halo mass function and void size function. Existing inference pipelines can therefore be used without significant modification, while achieving speed-ups of approximately three orders of magnitude. Although developed within

`CosmoBolognaLib`, the emulator can be straightforwardly used in external applications through the power-spectrum predictions provided by the library. Finally, the proposed framework provides a flexible and modular basis for future extensions towards more complex cosmological scenarios and cosmological observables.

Introduction

Cosmology seeks to understand the composition, geometry, and dynamical evolution of the Universe through the comparison between theoretical models and observational data. A key source of information is provided by the large-scale distribution of matter, whose observed properties reflect both the composition of the Universe and its dynamical evolution. Since the large-scale distribution of galaxy was first mapped, it has become clear that the matter distribution is not uniform on all scales (de Lapparent et al. 1986). It exhibits a complex network of galaxy clusters, filaments, and voids. These structures are collectively referred to as the large-scale structure (LSS), and their statistics provide some of the most powerful tools for probing the fundamental properties of the Universe. Indeed, the observed distribution of matter is the result of the interplay between the initial conditions of the Universe, its matter and energy content, and the laws governing its time evolution. Consequently, measurements of the LSS contain valuable information about the physical processes that have shaped the Universe throughout cosmic history.

The interpretation of these observations relies on cosmological models, namely theoretical frameworks that specify the physical constituents of the Universe and predict the behaviour of cosmological observables through a set of physical parameters. Among the various cosmological models proposed over the years, the Λ-cold dark matter (ΛCDM) model has emerged as the current standard model thanks to its remarkable agreement with a wide range of observations, including measurements of the cosmic microwave background (CMB, see e.g. Planck collaboration 2020; Madhavacheril et al. 2024), type Ia supernovae (see e.g. Brout et al. 2022; DES Collaboration et al. 2024), and the clustering of galaxies (see e.g. Alam et al. 2021; Adame et al. 2025). It includes six parameters: the baryon energy density $\Omega_{\rm b}$, the dark matter energy density $\Omega_{\rm cdm}$, the primordial scalar amplitude $A_{\rm s}$, the primordial spectral index $n_{\rm s}$, the Hubble constant H_0 and the optical depth τ.

Despite its remarkable observational success, the physical nature of both dark matter and dark energy remains unknown. Moreover, persistent discrepancies between independent cosmological probes, including for example the disagreement between local and CMB-based determinations of the Hubble constant, continue to motivate extensions of the standard cosmological model (see Di Valentino et al. 2021a,b,c, for a review). For these reasons, a number of extensions of the standard model have been proposed, including scenarios with non-zero spatial curvature ($k\Lambda$CDM), dynamical dark energy ($w_0 w_a$CDM), and modifications of General Relativity.

To test the cosmological model, we compare observations with theoretical predictions, focusing on different regions and scales of the Universe. For example, in the context of LSS, one of the main observable is the matter power spectrum: it measures the variance of cosmic density fluctuations as a function of spatial scale, quantifying the degree of matter clustering at various distances. This key statistic encodes information about the primordial density fluctuations, the composition of the Universe through its dependence on cosmological parameters, and the growth of cosmic structures. Specifically, the exact shape and amplitude of the power spectrum are dictated by the parameters governing the cosmological model. This fundamental dependence forms the cornerstone of modern LSS analyses, where predicting the matter power spectrum serves as the essential building block for modelling a wide range of clustering statistics.

Theoretical predictions of the power spectrum are typically obtained using codes called Boltzmann solvers. Several solvers are currently available to compute the power spectrum numerically given a set cosmological parameters, and well known examples include `CAMB` (Lewis et al. 2000) and `CLASS` (Lesgourgues 2011). Although these tools are extremely precise, their use in cosmological analyses can be challenging due to the computational time required to predict the matter power spectrum.

This computational burden becomes especially apparent during the analysis of cosmological data sets, which are typically processed within a Bayesian framework. To determine the best-fit parameters and their uncertainties, one must sample the posterior distribution, most commonly using Markov chain Monte Carlo (MCMC) techniques. Because MCMC algorithms explore the parameter space by generating samples proportionally to their probability, the procedure requires a new Boltzmann solver prediction for every proposed point. A full Bayesian inference can therefore easily require hundreds of thousands, or even millions, of matter power spectrum evaluations. While a single evaluation takes mere seconds, the cumulative cost is massive; extracting parameter constraints from power spectrum measurements often consumes tens of hours on clusters of tens to hundreds of CPUs.

For this reason, emulators of Boltzmann solvers have been widely adopted in recent years. An emulator is a tool that reproduces the predictions of a theoretical model without explicitly solving the underlying physical equations, thereby significantly reducing the computational cost. When using emulators, a Bayesian analysis such as the one mentioned above can be completed in times ranging from seconds to a few minutes. Examples of cosmological emulators available in the literature include `CosmoPower` (Spurio Mancini et al. 2022), `BACCO` (Aricò et al. 2022) and `Aletheia` (Sánchez et al. 2025). The construction of accurate emulators, however, is itself a challenging task. The volume of the parameter space grows rapidly with the number of free parameters, requiring increasingly large training sets and more complex neural-network (NN) architectures to maintain a given level of accuracy. This difficulty is commonly referred to as the curse of dimensionality and represents one of the main limitations of conventional emulation approaches.

The primary objective of this Thesis is to develop a highly accurate and computationally efficient NN emulator for the linear matter power spectrum, directly applicable within modern cosmological inference pipelines. Implemented within the `CosmoBolognaLib` (Marulli et al. 2015) framework, the core innovation of this work lies in its application of the evolution-mapping technique (Sánchez et al. 2022; Esposito et al. 2024; Fiorilli et al. 2025). By treating part of the cosmological dependence analytically, this approach significantly reduces the dimensionality of the training space that the NN must learn. This dimensional reduction is key to its performance, allowing the emulator to achieve an exceptional accuracy of better than 0.1% across all physical scales and redshifts relevant to current cosmological analyses. Consequently, this tool is highly competitive with—and in several respects superior to—existing codes, thanks to both its stringent precision and the broad variety of cosmological models it encompasses. Furthermore, its natively modular structure ensures it can be seamlessly extended to explore even more complex alternative scenarios in the future. Ultimately, this framework reduces the computational cost of Bayesian analyses by several orders of magnitude without sacrificing the rigorous accuracy demanded by modern LSS probes.

The path towards these results shapes the overall structure of this Thesis, which is organized as follows:

- Chapter 1 introduces the cosmological framework adopted throughout this work, reviewing the foundations of General Relativity, the dynamics of the expanding Universe, the standard ΛCDM model, and its extensions.
- Chapter 2 presents the theoretical description of structure formation and matter clustering, introducing linear perturbation theory, the growth of cosmic structures, and the main clustering statistics employed in cosmological analyses.
- Chapter 3 describes the machine-learning techniques adopted in this work and introduces the evolution-mapping framework underlying the emulator.
- Chapter 4 presents the implementation of the emulator within `CosmoBolognaLib`, including the generation of the training set, the NN architecture, and the validation methodology.
- Chapter 5 assesses the emulator performance through accuracy tests, speed-up measurements, and complete Bayesian analyses performed on mock data sets.
- Finally, Chapter 6 summarises the main results of this work and discusses future developments and possible extensions of the proposed framework.

Chapter 1

Cosmological background

In this chapter, we provide the theoretical background required for the analysis presented in this work. A comprehensive treatment of these topics can be found in Vittorio (2018) and Dodelson & Schmidt (2021), which we follow closely throughout this chapter.

Modern cosmology is grounded in General Relativity (GR), introduced by Einstein, which provides a consistent and testable framework for describing the dynamics of the Universe. In this context, both the expansion history and the evolution of matter perturbations are governed by GR. It is therefore necessary to briefly introduce the basic concepts underlying this theory and its main cosmological implications.

We begin by introducing the description of an expanding Universe, including the Hubble–Lemaître law and the definition of redshift. We then present the Friedmann equations, which relate the expansion rate to the energy content of the Universe. Finally, we discuss how different assumptions on the matter–energy components lead to the definition of specific cosmological models.

1.1 The Hubble–Lemaître law

Observational evidence indicates that the Universe is expanding. This was first established by Hubble (1929), who observed a linear relation between the distances of galaxies and their recession velocities, as illustrated in Fig. 1.1. As a consequence, the separation between cosmic objects increases with time, implying that it was smaller in the past.

It is therefore convenient to describe the expansion by introducing the scale factor $a(t)$, which quantifies the expansion of the Universe and is conventionally normalised to unity at the present time, $a(t_0) = 1$. The time t is called cosmic time and corresponds to the time measured by a clock in a reference system at rest in the expanding universe. In this framework, the physical distance $x(t)$ between two comoving points can be written as

$$x(t) = a(t)\, x_0 \,, \tag{1.1}$$

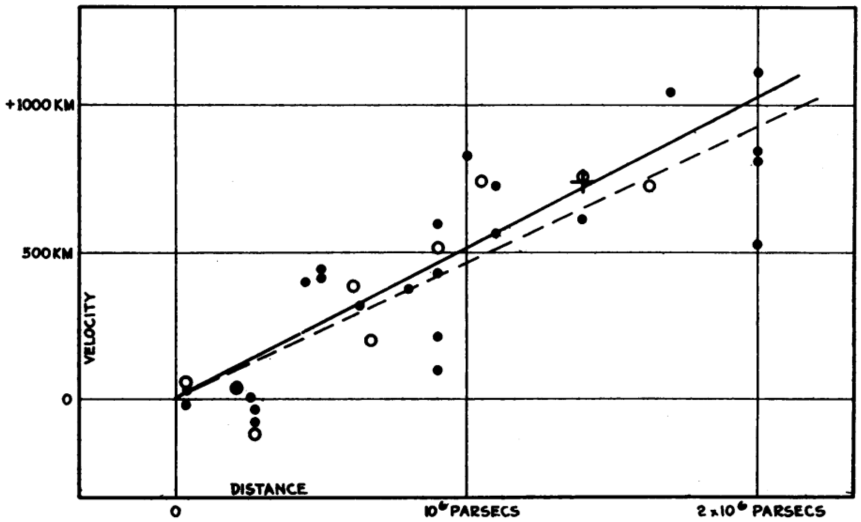

Figure 1.1: Original velocity–distance relation for galaxies from Hubble (1929). The recession velocity is shown as a function of distance, providing the first observational evidence for a linear relation and hence for the expansion of the Universe.

where x_0 is their comoving separation, which remains constant in time. This relation shows that the physical distance evolves proportionally to the scale factor.

The expansion of the Universe naturally leads to a recession velocity between distant objects. By differentiating the previous relation with respect to time, we obtain

$$v \equiv \dot{x} = \dot{a}\, x_0 = \frac{\dot{a}}{a}\, x \,. \tag{1.2}$$

Defining the Hubble parameter as

$$H(t) \equiv \frac{\dot{a}}{a}\,, \tag{1.3}$$

we recover the Hubble–Lemaître law,

$$v = H(t)\, x \,. \tag{1.4}$$

At the present time, this relation reduces to $v = H_0\, x$, where $H_0 \equiv H(t_0)$ is the Hubble constant, representing the current expansion rate of the Universe.

Another key observational consequence of the expansion is the cosmological redshift. As the Universe expands, the wavelength of photons propagating through spacetime is stretched proportionally to the scale factor. As a result, the observed wavelength $\lambda_{\rm obs}$ is larger than the emitted one $\lambda_{\rm emit}$. The redshift z is defined as

$$1 + z \equiv \frac{\lambda_{\rm obs}}{\lambda_{\rm emit}} = \frac{a(t_0)}{a(t_{\rm emit})}\,, \tag{1.5}$$

which, using the normalisation $a(t_0) = 1$, simplifies to $1 + z = 1/a(t_{\rm emit})$.

1.2 The Friedmann equations

In this section, we present the equations governing the expansion of the Universe within the framework of GR. Under the assumptions of homogeneity and isotropy, the spacetime geometry is described by the Friedmann–Lemaître–Robertson–Walker (FLRW) metric. The time evolution of the scale factor is then determined by Einstein's field equations, which relate the geometry of spacetime to its energy content. In the following, we first introduce the FLRW metric and the description of the cosmic fluid, and then present Einstein's equations, from which the Friedmann equations are derived.

1.2.1 The FLRW metric

The theoretical framework of cosmology is based on GR, which provides a description of gravity in terms of the geometry of spacetime. A fundamental requirement of any physical theory, including GR, is general covariance, i.e. the laws of physics must take the same form in any coordinate system. In this context, gravity is described by a curved spacetime, whose geometry is fully specified by the metric tensor. Let us consider a point in spacetime described by the coordinates $x^\mu = (x^0, x^1, x^2, x^3)$ and an infinitesimally displaced point $x^\nu = x^\mu + \mathrm{d}x^\mu$. The spacetime interval between these two points is defined as

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu\,, \tag{1.6}$$

where $g_{\mu\nu}$ is the symmetric metric tensor and we have adopted the Einstein summation convention over repeated indices.

The specific form of the metric tensor depends on the physical properties of the spacetime under consideration. On scales larger than 60–70 h^{-1} Mpc observations support the validity of the cosmological principle, according to which the Universe is homogeneous and isotropic (Yadav et al. 2005). Under these assumptions, the spacetime geometry is described by the FLRW metric. In comoving coordinates (ct, r, θ, ϕ), with c the speed of light, the FLRW line element can be written as

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right]\,, \tag{1.7}$$

where $a(t)$ is the scale factor and $k = 0, \pm 1$ determines the spatial curvature of the Universe, corresponding to flat, open, and closed geometries, respectively.

1.2.2 The energy–momentum tensor

The matter–energy content of the Universe is characterised by the energy–momentum tensor $T_{\mu\nu}$. Assuming the cosmological principle, the cosmic fluid can be modelled as a perfect fluid at rest in comoving coordinates, fully characterised by its energy density ρ and pressure $\mathcal{P}$. In this frame, the energy–momentum tensor takes the diagonal form $T^\mu_{\ \nu} = \mathrm{diag}(-\rho c^2, \mathcal{P}, \mathcal{P}, \mathcal{P})$.

The evolution of the fluid is governed by the conservation of the energy–momentum tensor,

$$\nabla_\mu T^\mu_\nu = 0\,, \tag{1.8}$$

where ∇_μ denotes the covariant derivative. This relation provides the general-relativistic analogue of the continuity and Euler equations. In a homogeneous and isotropic Universe described by the FLRW metric, this relation reduces to the continuity equation

$$\dot{\rho} + 3H\left(\rho + \frac{\mathcal{P}}{c^2}\right) = 0. \tag{1.9}$$

It describes how the energy density evolves as a consequence of the expansion of the Universe.

It is then convenient to characterise each component of the cosmic fluid (see Sect. 1.3) through an equation of state parameter w_s, defined for a given species s as

$$w_s \equiv \frac{\mathcal{P}_s}{\rho_s c^2}\,. \tag{1.10}$$

Under this definition, the evolution of the energy density can be written as

$$\rho_s(a) \propto \exp\left\{-3\int^a \frac{\mathrm{d}a'}{a'}\left[1 + w_s(a')\right]\right\}\,, \tag{1.11}$$

which, in the case of a constant equation of state parameter, reduces to

$$\rho_s(a) \propto a^{-3(1+w_s)}\,. \tag{1.12}$$

For example, relativistic matter and radiation have $w = 1/3$, while non-relativistic matter have $w = 0$.

1.2.3 The Einstein equations

The relation between the geometry of spacetime and its matter–energy content is encoded in Einstein's field equations. In order to introduce these equations, it is useful to briefly recall the main geometrical quantities that describe spacetime curvature.

The connection coefficients, or Christoffel symbols, are defined as

$$\Gamma^\alpha_{\beta\gamma} \equiv \frac{1}{2}\, g^{\alpha\delta}\left(\partial_\beta g_{\gamma\delta} + \partial_\gamma g_{\beta\delta} - \partial_\delta g_{\beta\gamma}\right)\,, \tag{1.13}$$

where ∂_μ denotes the partial derivative with respect to the coordinate x^μ. These quantities do not form a tensor on the spacetime manifold, but rather correspond to the coordinate-dependent components of the affine connection on the tangent bundle. Nevertheless, they play a central role in defining covariant derivatives and describe how vectors are parallel transported in curved spacetime. In particular, they appear in the geodesic equation, which governs the motion of freely falling particles (Rovelli 2021).

The curvature of spacetime is quantified by the Riemann tensor, defined as

$$R^{\alpha}{}_{\mu\beta\nu} = \partial_\beta \Gamma^{\alpha}_{\mu\nu} - \partial_\nu \Gamma^{\alpha}_{\mu\beta} + \Gamma^{\alpha}_{\beta\gamma}\Gamma^{\gamma}_{\mu\nu} - \Gamma^{\alpha}_{\nu\gamma}\Gamma^{\gamma}_{\mu\beta} \,. \tag{1.14}$$

From the Riemann tensor, one can construct the Ricci tensor $R_{\mu\nu} \equiv R^{\alpha}{}_{\mu\alpha\nu}$, and the Ricci scalar $R \equiv g^{\mu\nu}R_{\mu\nu}$. A particularly important combination of these quantities is the Einstein tensor,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \,, \tag{1.15}$$

which fully characterises the geometric properties of spacetime entering the field equations.

On the other hand, we saw that the matter–energy content of the Universe is described by the energy–momentum tensor $T_{\mu\nu}$, so Einstein's field equations can be written as

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}\,T_{\mu\nu} \,, \tag{1.16}$$

where G is Newton's gravitational constant and Λ is the cosmological constant. Eq. (1.16) establishes a direct link between spacetime geometry and the energy content of the Universe. From the field equations it is possible to derive the growth of structure at late times, as we will see in Chapter 2.

When applied to the FLRW metric, Eq. (1.16) leads to the Friedmann equations, which govern the time evolution of the scale factor. In particular, the time–time and space–space components yield, respectively,

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\,\rho \,, \tag{1.17}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3\mathcal{P}}{c^2}\right) \,, \tag{1.18}$$

which are called first and second Friedmann equations.

1.3 The cosmological models

The framework introduced so far is completely general and does not depend on the specific composition of the Universe. In particular, Eq. (1.11) describes the evolution of the energy density for any component once its equation of state is specified, and these quantities enter the Friedmann equations to determine the expansion history.

Cosmological models are obtained by specifying the different contributions to the total energy density, their corresponding equations of state, and the spatial curvature. In this section, we describe the main components of the cosmic fluid and show how different assumptions on these quantities lead to the definition of specific cosmological models.

1.3.1 The concordance model of cosmology

The currently accepted cosmological framework is the ΛCDM model, which is supported by a wide range of independent observations, most notably measurements of the CMB anisotropies by the Planck collaboration (2020). This model describes a spatially flat Universe composed of several components contributing to the total energy density.

It is convenient to express all energy densities in common units by introducing the critical density:

$$\rho_{\rm cr} \equiv \frac{3H_0^2}{8\pi G} = 1.88\, h^2 \times 10^{-29}\ {\rm g\, cm^{-3}}, \tag{1.19}$$

with h the dimensionless Hubble parameter, defined as $H_0 = 100\, h\, {\rm km\, s^{-1}\, Mpc^{-1}}$. The density parameter of a given species s is then defined as

$$\Omega_s \equiv \frac{\rho_s(t_0)}{\rho_{\rm cr}}. \tag{1.20}$$

It is also common to define the physical density parameter

$$\omega_s \equiv \Omega_s\, h^2, \tag{1.21}$$

which is directly constrained by cosmological observations and will be used in the following (see Sect. 3.3).

The first component we consider is radiation, whose dominant contribution today is given by photons in the form of the CMB. This radiation decoupled from matter at $z \approx 1100$ and has since propagated freely. A full-sky map of the CMB temperature anisotropies, as measured by the Planck satellite, is shown in Fig. 1.2. Its present-day temperature is measured to be $T_0 = 2.7255 \pm 0.0006\,{\rm K}$ (Fixsen 2009), which implies a photon density parameter $\omega_\gamma \simeq 2.47 \times 10^{-5}$. Therefore, photons provide a negligible contribution to the present-day energy budget.

The baryonic component consists of ordinary matter described by the Standard Model. Its physical density is tightly constrained by CMB observations, yielding $\omega_{\rm b} = 0.0224 \pm 0.0001$ (Planck collaboration 2020). This corresponds to approximately 5% of the total energy density of the Universe.

A much larger fraction of the matter content is in the form of cold dark matter (CDM), a non-relativistic and non-baryonic component that interacts predominantly through gravity. Its physical density is measured to be $\omega_{\rm cdm} = 0.120 \pm 0.001$ (Planck collaboration 2020), implying that matter accounts for about 30% of the total energy density, with CDM being the dominant contribution.

The remaining 70% of the energy density is attributed to dark energy, which is responsible for the observed accelerated expansion of the Universe. In the ΛCDM model, dark energy is described by a cosmological constant Λ, characterised by an equation-of-state parameter $w = -1$. This interpretation is consistent with multiple cosmological probes, including CMB, baryon acoustic oscillations (BAO), and supernova observations (Planck collaboration 2020).

Finally, a subdominant contribution arises from massive neutrinos. Their present-day energy density can be expressed as (Mangano et al. 2005)

$$\omega_\nu = \frac{\sum_i m_{\nu,i}}{93.14\,{\rm eV}}, \tag{1.22}$$

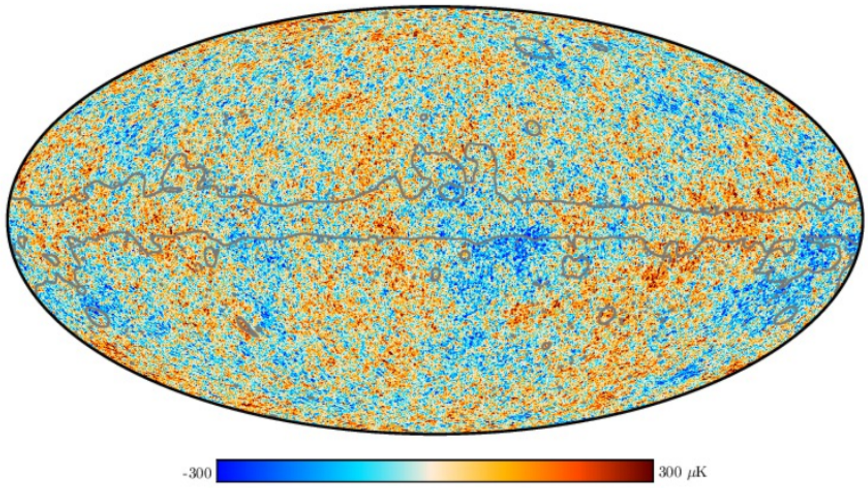

Figure 1.2: Full-sky map of CMB temperature anisotropies (units of ΔT) from the Planck 2018 SMICA reconstruction. The map is shown in Galactic coordinates and has had the Galactic plane masked. The colour scale highlights fluctuations at the $\sim 100\,\mu K$ level (relative to the mean CMB temperature, i.e. $\Delta T/T \sim 10^{-5}$). Credit: ESA/Planck Collaboration (Planck Collaboration 2020).

which, given current constraints $\sum m_\nu \lesssim 0.12\,\mathrm{eV}$ (Planck collaboration 2020), corresponds to $\omega_\nu \lesssim 0.0013$, i.e. a contribution at the level of about 0.1–0.3% of the total energy budget.

In this work, we neglect the contribution of massive neutrinos. This approximation is justified on large scales, where their impact is negligible. As a first approximation, we therefore ignore their effect in order to simplify the analysis and reduce parameter degeneracies. Possible strategies to include massive neutrinos will be discussed in Sect. 6.2.

As anticipated in the Introduction, the ΛCDM model is characterised by six parameters: Ω_{cdm}, Ω_{b}, A_{s}, n_{s}, H_0, and τ. These parameters fully specify both the background expansion and the properties of the primordial density fluctuations. In particular, Ω_{cdm} and Ω_{b} determine the present-day matter content of the Universe, while A_{s} and n_{s} describe the amplitude and spectral tilt of the primordial power spectrum, respectively. The Hubble constant H_0 sets the current expansion rate, and the optical depth τ accounts for the effects of reionisation on the CMB.

Given this parameterisation, the evolution of the different cosmological components can be described in a simple and unified way. In the ΛCDM framework, each component is characterised by a constant equation-of-state parameter w, which implies a power-law scaling of the corresponding energy density with the scale factor (see Eq. 1.12). This property allows one to express the expansion history directly in terms of the cosmological parameters introduced above. In particular, using the definition of the Hubble parameter, Eq. (1.17) can be rewritten in the convenient form

$$\frac{H^2(a)}{H_0^2} = \Omega_{\mathrm{m}}\, a^{-3} + \Omega_{\mathrm{r}}\, a^{-4} + \Omega_\Lambda\,, \tag{1.23}$$

where we have defined $\Omega_\mathrm{m} \equiv \Omega_\mathrm{b} + \Omega_\mathrm{cdm}$.

1.3.2 Non-standard cosmological scenarios

Non-standard cosmological scenarios are extensions of the ΛCDM model in which one or more of its underlying assumptions are relaxed, including modifications of the dark energy sector or of the theory of gravity itself. In the following, we briefly discuss some commonly considered extensions beyond the concordance cosmological model.

If the spatial geometry of the Universe is not flat, the cosmological model can be extended by introducing the curvature density parameter Ω_k. This extended model is referred to as $k\Lambda$CDM. In this case, Eq. (1.23) becomes

$$\frac{H^2(a)}{H_0^2} = \Omega_\mathrm{m}\, a^{-3} + \Omega_\mathrm{r}\, a^{-4} + \Omega_\Lambda + \Omega_k\, a^{-2}\,, \tag{1.24}$$

where the additional term accounts for the contribution of spatial curvature to the expansion rate. Current cosmological observations place tight constraints on the curvature parameter, yielding results consistent with a spatially flat Universe, with only very small deviations from $\Omega_k = 0$ allowed (Planck collaboration 2020). This justifies the common assumption of spatial flatness in the standard ΛCDM model.

If dark energy is not a cosmological constant, its energy density must be modelled through a time-dependent equation-of-state parameter. In this case, Eq. (1.11) is used. A widely adopted parametrisation is the Chevallier–Polarski–Linder (CPL) form (Chevallier & Polarski 2001; Linder 2003),

$$w(a) = w_0 + w_a(1 - a)\,, \tag{1.25}$$

which describes a linear evolution of the equation-of-state parameter with the scale factor. It is then convenient to introduce the function

$$f_\mathrm{DE}(a) = a^{-3(1+w_0+w_a)} \exp\left[-3w_a(1-a)\right]\,, \tag{1.26}$$

which encodes the redshift evolution of the dark energy density.

In this case, Eq. (1.23) takes the form

$$\frac{H^2(a)}{H_0^2} = \Omega_\mathrm{m}\, a^{-3} + \Omega_\mathrm{r}\, a^{-4} + \Omega_\mathrm{DE}\, f_\mathrm{DE}(a)\,. \tag{1.27}$$

This scenario is referred to as w_0w_aCDM, as the model is characterised by the two additional parameters w_0 and w_a, which describe the present value and the time evolution of the dark energy equation of state. Recent results from the Dark Energy Spectroscopic Instrument (DESI) have provided constraints on the dark energy equation of state, showing mild preferences for dynamical dark energy models when combined with the CMB and supernova data, thereby strengthening the case for parametrisations such as w_0w_aCDM as viable extensions of the standard ΛCDM scenario (Adame et al. 2025b).

Beside the remarkable success of GR, a variety of alternative theories of gravity have been proposed, commonly referred to as modified gravity models. Among these, a particularly well-studied class is represented by the $f(R)$ theories, in which the Einstein–Hilbert action is generalised. Neglecting, for simplicity, the matter contribution, the Einstein–Hilbert action can be written as (Rovelli 2021)

$$S[g] = \frac{1}{16\pi G} \int \mathrm{d}^4x \sqrt{-g}\, (R - 2\Lambda) \,, \tag{1.28}$$

where $g \equiv \det(g_{\mu\nu})$, R is the Ricci scalar, and Λ is the cosmological constant. In $f(R)$ theories, the action in Eq. (1.28) is generalised by replacing the linear dependence on R with a generic function $f(R)$, leading to

$$S[g] = \frac{1}{16\pi G} \int \mathrm{d}^4x \sqrt{-g}\, f(R) \,. \tag{1.29}$$

In this framework, the cosmological constant can be either included explicitly or effectively generated by the functional form of $f(R)$. The field equations derived from this action remain generally covariant and Lorentz invariant, as in GR. However, in contrast to Einstein's equations, they are typically fourth-order in derivatives of the metric, reflecting the increased complexity of the theory. The standard GR limit is recovered for $f(R) = R - 2\Lambda$. A key feature of $f(R)$ theories is the presence of an additional scalar degree of freedom, which can be interpreted as an effective scalar field coupled to gravity. From a cosmological perspective, $f(R)$ models have attracted considerable interest as they can naturally give rise to accelerated expansion, both at early times (inflation) and at late times, without necessarily invoking a cosmological constant (Clifton et al. 2012). Despite these appealing features, $f(R)$ theories are tightly constrained by observations and theoretical consistency. Viable models must closely reproduce GR on small scales, often requiring screening mechanisms, to avoid instabilities or unphysical solutions. These requirements significantly restrict the allowed functional forms of $f(R)$.

The inclusion of modified gravity scenarios within an emulator framework is beyond the scope of this work and will be deferred to future studies. Nevertheless, the methodological framework presented in the following can be naturally extended to include such models, allowing for a consistent analysis within these extended cosmological scenarios (see Sect. 6.2).

Chapter 2

Structure formation and clustering

In this chapter we describe the linear evolution of cosmological perturbations leading to the formation of LSS and its observables. This requires solving the coupled Einstein–Boltzmann system for the metric and matter perturbations.

Although the early Universe was highly homogeneous, small perturbations were already present. The formation of LSS arises from the gravitational growth of these initial inhomogeneities, which can be described by perturbing Eq. (1.16). The typical amplitude of these perturbations is of the order of 10^{-5} (Planck collaboration 2020). They are imprinted in cosmological observables such as the CMB anisotropies and the large-scale distribution of matter. Therefore, a linear treatment provides an excellent approximation on sufficiently large scales. Thus, we consider a perturbed FLRW spacetime and work at first order in the perturbations, assuming that their amplitude is small.

We restrict our analysis to scalar perturbations and adopt the conformal Newtonian gauge, in which the metric takes the form (Dodelson & Schmidt 2021)

$$g_{00} = -1 - 2\Psi(\boldsymbol{x}, t) \ , \tag{2.1}$$

$$g_{0i} = 0 \ , \tag{2.2}$$

$$g_{ij} = a^2(t)\delta_{ij} \left[1 + 2\Phi(\boldsymbol{x}, t)\right] \ , \tag{2.3}$$

where Ψ and Φ are scalar potentials. The function Ψ corresponds to the Newtonian gravitational potential and governs the motion of non-relativistic particles, whereas Φ describes perturbations to the spatial curvature. Given this metric, the Einstein tensor can be expanded at first order, providing the left-hand side of the perturbed Einstein equations. The right-hand side of the Einstein equations is given by the energy–momentum tensor, which depends on the energy density and pressure of the species present in the Universe. These components are described statistically in terms of phase-space distribution functions. For a species s, the distribution function $f_s(\boldsymbol{x}, \boldsymbol{p}, t)$ is defined such that the number of particles in a phase-space volume element is

$$\mathrm{d}N = \frac{g_s}{(2\pi\hbar)^3} f_s(\boldsymbol{x}, \boldsymbol{p}, t)\, \mathrm{d}^3x\, \mathrm{d}^3p \ , \tag{2.4}$$

where g_s denotes the number of internal degrees of freedom. The distribution function encodes the microscopic state of the system and allows one to compute the macroscopic quantities that appear in the energy–momentum tensor.

The particle trajectories are described by their four-momentum

$$P^\mu \equiv \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \,, \tag{2.5}$$

where λ is an affine parameter along the particle worldline. For massive particles, λ can be identified with the proper time, up to a constant rescaling, while for massless particles it provides a convenient parametrisation of null geodesics. The spatial components P^i are related to the physical momentum p^i through the expansion of the Universe as

$$p^i = a(t)\, P^i \,, \tag{2.6}$$

so that p^i represents the physical momentum measured by a comoving observer. It is convenient to decompose the momentum into its magnitude and direction as

$$p^i = p\,\hat{p}^i \,, \tag{2.7}$$

where $p \equiv |p^i|$ is the magnitude of the physical momentum, and $\hat{p}^i$ is the unit vector specifying its direction. This decomposition is particularly useful when describing anisotropies in the distribution function. The energy–momentum tensor can then be written as a moment of the distribution function (Dodelson & Schmidt 2021),

$$T^\mu_\nu(\boldsymbol{x}, t) = \frac{g}{\sqrt{-\det[g_{\alpha\beta}]}} \int \frac{\mathrm{d}P_1 \mathrm{d}P_2 \mathrm{d}P_3}{(2\pi\hbar)^3} \frac{P^\mu P_\nu}{P^0} f_s(\boldsymbol{x}, \boldsymbol{p}, t) \,, \tag{2.8}$$

where g denotes the degeneracy factor, accounting for the number of internal degrees of freedom described by the distribution function f. The energy–momentum tensor provides the source term for the Einstein equations and represents the flux of four-momentum carried by the particles.

The evolution of the distribution function is governed by the Boltzmann equation, which includes the derivatives of the distribution function. In an expanding and perturbed Universe, this can be written as

$$\frac{\partial f_s}{\partial t} + \frac{\mathrm{d}x^i}{\mathrm{d}t}\frac{\partial f_s}{\partial x^i} + \frac{\mathrm{d}p}{\mathrm{d}t}\frac{\partial f_s}{\partial p} + \frac{\mathrm{d}\hat{p}^i}{\mathrm{d}t}\frac{\partial f_s}{\partial \hat{p}^i} = C[f_s] \,, \tag{2.9}$$

where $C[f_s]$ denotes the collision term, which accounts for particle interactions. Expanding both the metric and the distribution function around their homogeneous background values, the Boltzmann equation generates a hierarchy of coupled differential equations for the perturbations of each species. Together with the perturbed Einstein equations, this defines the Einstein–Boltzmann system, which must be solved numerically to obtain cosmological observables such as the CMB anisotropies and the matter power spectrum.

2.1 Linear perturbation theory

We now focus on the anisotropies in the radiation field, on the inhomogeneities in the matter distribution, and on how these perturbations source the gravitational field within GR. Since gravity couples to the total energy–momentum tensor, all species are dynamically coupled through the metric perturbations. As a consequence, the full system must be solved simultaneously. In this section, all equations are expanded at linear order around a homogeneous and isotropic FLRW background. This approximation is valid as long as perturbations remain small.

2.1.1 Boltzmann equations

We begin by deriving the Boltzmann equation for photons. The photon distribution function follows a Bose–Einstein distribution. At zeroth order, the photon temperature depends only on time, $T(t)$, reflecting the homogeneity and isotropy of the background Universe. At linear order, however, small spatial and angular variations are present. These are encoded in the fractional temperature perturbation

$$\Theta(\boldsymbol{x}, \hat{\boldsymbol{p}}, t) \equiv \frac{\delta T}{T} , \tag{2.10}$$

which depends both on position and on the photon propagation direction $\hat{\boldsymbol{p}}$, thus describing inhomogeneities and anisotropies respectively. The zeroth order of the perturbed Bose–Einstein distribution yields the well-known scaling $T \propto a^{-1}$, which reflects the redshifting of photon momenta due to the expansion of the Universe. The first-order expansion instead encodes the evolution of anisotropies and will be inserted into the Boltzmann equation.

Before writing the full equation, it is useful to introduce a few key quantities. The conformal time (or comoving horizon) η is defined as

$$\eta \equiv \int_0^t \frac{c\, \mathrm{d}t'}{a(t')} , \tag{2.11}$$

and represents the comoving distance travelled by light since the beginning of the Universe. It sets the causal structure: only regions within the comoving horizon η can be in causal connection. Another important quantity is the optical depth

$$\tau(\eta) \equiv \int_\eta^{\eta_0} \mathrm{d}\eta'\, a\, n_e \sigma_T , \tag{2.12}$$

which measures the probability that a photon has scattered between time η and today. Its derivative $\tau' = -a\, n_e \sigma_T$ quantifies the interaction rate. When $|\tau'|$ is large, photons are tightly coupled to electrons; when it becomes small, photons free–stream. Finally, it is convenient to decompose the angular dependence of the temperature perturbation in Fourier space using Legendre polynomials:

$$\Theta_\ell(k, \eta) \equiv \frac{1}{(-i)^\ell} \int_{-1}^{1} \frac{\mathrm{d}\mu}{2}\, \mathcal{P}_\ell(\mu)\, \Theta(k, \mu, \eta) , \tag{2.13}$$

where $\mu \equiv \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{p}}$ is the cosine of the angle between the Fourier mode $\boldsymbol{k}$ and the photon direction. The multipole moments Θ_ℓ have a clear physical interpretation: the monopole ($\ell = 0$) corresponds to the local temperature fluctuation, the dipole

($\ell = 1$) to the bulk velocity, and higher multipoles encode anisotropies of increasing angular complexity.

Expressing derivatives with respect to conformal time, and working in Fourier space, the Boltzmann equation for photons is

$$\Theta' + ik\mu\Theta = -\Phi' - ik\mu\Psi - \tau' \left[\Theta_0 - \Theta + \mu u_b - \frac{1}{2}\mathcal{P}_2(\mu)\,\Pi \right] . \tag{2.14}$$

The left–hand side of Eq. (2.14) describes the free streaming of photons. The term $ik\mu\Theta$ accounts for the propagation of anisotropies along the photon direction. The first two terms on the right–hand side encode gravitational effects. The term involving Φ' represents gravitational redshift, while the $ik\mu\Psi$ term accounts for Doppler shifts induced by spatial gradients of the gravitational potential. Together, they describe how metric perturbations modify photon energies along their trajectories. The last term, proportional to τ', is the collision term and describes Thomson scattering between photons and free electrons. Physically, this term drives the photon distribution towards local equilibrium with the baryon fluid. The term Π represents the anisotropic scattering contribution, and is defined as

$$\Pi \equiv \Theta_2 + \Theta_{P,2} + \Theta_{P,0}\,, \tag{2.15}$$

which includes both the temperature quadrupole and the contribution from polarisation. This term arises because Thomson scattering is not isotropic: it depends on the angular distribution of incoming radiation and generates linear polarisation.

The second Boltzmann equation we consider is that for neutrinos. The derivation closely follows that of photons, with a few differences. At zeroth order, neutrinos follow an equilibrium Fermi–Dirac distribution with temperature $T_\nu(t)$. After neutrino decoupling, all non-gravitational interactions become negligible. As a consequence, the collision term in the Boltzmann equation vanishes, and neutrinos evolve purely through free streaming in a perturbed spacetime. Expanding to first order the distribution function, and expressing all quantities in terms of conformal time and in Fourier space, one obtains

$$\mathcal{N}' + ik\mu\frac{p}{E_\nu(p)}\,\mathcal{N} - \mathcal{H}\,p\frac{\partial\mathcal{N}}{\partial p} = -\Phi' - ik\mu\frac{E_\nu(p)}{p}\,\Psi\,, \tag{2.16}$$

where $\mathcal{H} \equiv a'/a$ is the conformal Hubble parameter.

The left-hand side describes the free propagation of neutrinos. The term proportional to $ik\mu$ accounts for the streaming of perturbations and is modulated by the factor $p/E_\nu(p)$, which is the particle velocity. For relativistic neutrinos, $E_\nu \approx p$ and this factor reduces to unity, recovering the photon-like behaviour. At late times, however, when neutrinos become non-relativistic, $p/E_\nu \ll 1$, and their free streaming is suppressed. The term $-\mathcal{H}\,p\,\partial\mathcal{N}/\partial p$ arises from the redshifting of momenta due to the expansion of the Universe. It accounts for the fact that the neutrino energy is not simply proportional to $\hat{p}$ once they become non-relativistic.

The right-hand side describes the effect of gravitational perturbations. The term $-\Phi'$ corresponds to gravitational redshift effects, while the term proportional to Ψ

encodes the effect of spatial gradients of the gravitational potential on the neutrino trajectories.

The third Boltzmann equation we consider is that for CDM. By definition, CDM particles interact only gravitationally, so that the collision term vanishes. Moreover, CDM is highly non-relativistic at all times of interest. Thus dark matter can be described as a pressureless fluid: we retain the bulk motion of the particles, but neglect their velocity dispersion. In this framework, the relevant variables are the density contrast

$$\delta_\mathrm{c} \equiv \frac{\rho_\mathrm{c} - \bar{\rho}_\mathrm{c}}{\bar{\rho}_\mathrm{c}}\,, \tag{2.17}$$

and the bulk velocity u_c of the CDM fluid. At zeroth order, the Boltzmann equation implies that the number density scales as $n_\mathrm{c} \propto a^{-3}$, reflecting dilution due to the expansion of the Universe. At first order, working in Fourier space and using conformal time, the Boltzmann equation reduces to the continuity equation

$$\delta_\mathrm{c}' + iku_\mathrm{c} + 3\Phi' = 0\,. \tag{2.18}$$

The term iku_c describes the change in density due to the divergence of the velocity field: matter flowing out of a region decreases its density, while inflow increases it. The term $3\Phi'$ encodes the effect of the time evolution of the gravitational potential. However, Eq. (2.18) alone is not sufficient, since it involves both the density contrast and the velocity field. To close the system, one needs an additional equation describing the evolution of u_c. This is provided by the Euler equation, which, at linear order, reads

$$u_\mathrm{c}' + \mathcal{H}u_\mathrm{c} + ik\Psi = 0\,. \tag{2.19}$$

The Euler equation expresses momentum conservation for the CDM fluid. The term $\mathcal{H}u_\mathrm{c}$ represents Hubble friction: as the Universe expands, peculiar velocities are damped. The term $ik\Psi$ corresponds to gravitational acceleration, driving the motion of dark matter particles towards potential wells.

Finally, we consider the Boltzmann equation for baryons. As for CDM, the relevant variables are the density contrast

$$\delta_\mathrm{b} \equiv \frac{\rho_\mathrm{b} - \bar{\rho}_\mathrm{b}}{\bar{\rho}_\mathrm{b}}\,, \tag{2.20}$$

and the bulk velocity u_b of the baryon fluid. Baryons are composed of protons and electrons. At the epochs of interest, these two species are tightly coupled by Coulomb interactions, which are extremely efficient in enforcing local charge neutrality and in equalising their velocities. As a result, they can be treated as a single fluid characterised by a common density contrast and bulk velocity. Since baryons are non-relativistic, the same fluid approximation used for CDM applies. At zeroth order, their number density scales as $n_\mathrm{b} \propto a^{-3}$ due to the expansion of the Universe. At linear order, the Boltzmann equation reduces to the continuity equation

$$\delta_\mathrm{b}' + iku_\mathrm{b} + 3\Phi' = 0\,. \tag{2.21}$$

The structure of this equation is identical to that of CDM and contains the same information. The absence of a collision term reflects the fact that Coulomb and Thomson interactions conserve the number of baryons, so that they do not directly source or sink density perturbations. However, the situation is different for the evolution of the velocity field. Taking the first moment of the Boltzmann equation for electrons and protons and combining them, one obtains the Euler equation for the baryon fluid:

$$u_b' + \mathcal{H} u_b + ik\Psi = \tau' \frac{4\rho_\gamma}{3\rho_b} (u_b + 3i\Theta_1) \, , \tag{2.22}$$

where Θ_1 is the dipole of the photon temperature perturbation. The left-hand side of Eq. (2.22) is identical to the CDM case and describes the free evolution of baryon velocities under Hubble expansion and gravitational forces. The right-hand side, instead, encodes the effect of Thomson scattering between photons and electrons. Physically, this term represents a momentum exchange between the photon and baryon fluids. Since photons carry momentum, their anisotropic distribution (in particular the dipole Θ_1) exerts a drag force on the baryons. The factor $4\rho_\gamma/3\rho_b$ is the ratio of photon to baryon inertia and determines the efficiency of this coupling. When $|\tau'|$ is large, scattering is very efficient and baryons are tightly coupled to photons. In this regime, the baryon velocity is forced to follow the photon dipole, leading to a single tightly coupled photon–baryon fluid. At later times, as the Universe expands and the free electron density drops, $|\tau'|$ decreases and the coupling weakens. This marks the transition to photon decoupling, after which baryons evolve more similarly to CDM.

2.1.2 The Einstein–Boltzmann system

The Boltzmann equations derived in the previous section describe the evolution of the different particle species in a perturbed spacetime. However, metric perturbations are dynamically generated by the perturbations in the matter and radiation fields through the Einstein equations. The evolution of cosmological perturbations is therefore determined by a coupled Einstein–Boltzmann system.

Starting from the perturbed metric introduced in Eq. (2.1), the Christoffel symbols can be expanded as

$$\Gamma^\alpha_{\mu\nu} = \bar{\Gamma}^\alpha_{\mu\nu} + \delta\Gamma^\alpha_{\mu\nu} \, , \tag{2.23}$$

where $\bar{\Gamma}^\alpha_{\mu\nu}$ denotes the background FLRW contribution, while $\delta\Gamma^\alpha_{\mu\nu}$ contains only linear terms in the scalar perturbations Φ and Ψ. From these quantities one obtains the linear perturbations of the Ricci tensor and Ricci scalar,

$$R_{\mu\nu} = \bar{R}_{\mu\nu} + \delta R_{\mu\nu} \, , \tag{2.24}$$

$$R = \bar{R} + \delta R \, , \tag{2.25}$$

and therefore the perturbed Einstein tensor,

$$G_{\mu\nu} = \bar{G}_{\mu\nu} + \delta G_{\mu\nu} \, . \tag{2.26}$$

On the matter side, perturbations in the phase–space distribution functions induce corresponding perturbations in the energy–momentum tensor through Eq. (2.8),

$$T^{\mu}_{\nu} = \bar{T}^{\mu}_{\nu} + \delta T^{\mu}_{\nu} \; . \tag{2.27}$$

The different components of $\delta T^{\mu}_{\ \nu}$ describe perturbations in the energy density, pressure, bulk velocity, and anisotropic stress of the various cosmological species. Equating the linear perturbations of the Einstein and energy–momentum tensors yields the perturbed Einstein equations,

$$\delta G^{\mu}{}_{\nu} = \frac{8\pi G}{c^2}\, \delta T^{\mu}_{\nu} \; . \tag{2.28}$$

It is convenient to introduce total matter and radiation variables. For matter (CDM and baryons), we define

$$\rho_{\rm m}\, \delta_{\rm m} \equiv \rho_c\, \delta_c + \rho_b\, \delta_b \, , \tag{2.29}$$

$$\rho_{\rm m}\, u_{\rm m} \equiv \rho_c\, u_c + \rho_b\, u_b \, , \tag{2.30}$$

while for radiation (photons and neutrinos) we define

$$\rho_{\rm r}\, \Theta_{r,0} \equiv \rho_\gamma\, \Theta_0 + \rho_\nu\, \mathcal{N}_0 \, , \tag{2.31}$$

$$\rho_{\rm r}\, \Theta_{r,1} \equiv \rho_\gamma\, \Theta_1 + \rho_\nu\, \mathcal{N}_1 \, . \tag{2.32}$$

In Fourier space and using conformal time, the perturbed Einstein equations can be written as

$$k^2\Phi + 3\mathcal{H}\left(\Phi' + \mathcal{H}\Psi\right) = 4\pi G a^2 \left[\rho_{\rm m}\, \delta_{\rm m} + 4\rho_{\rm r}\, \Theta_{r,0}\right] \; , \tag{2.33}$$

$$k^2(\Phi + \Psi) = -32\pi G a^2 \rho_{\rm r}\, \Theta_{r,2} \; . \tag{2.34}$$

Equation (2.33) is the generalisation of the Poisson equation to an expanding Universe. It relates the gravitational potential Φ to the total energy density perturbation, including both matter and radiation. The additional terms involving $\mathcal{H}$ account for the expansion of the background spacetime. Equation (2.34) describes the effect of anisotropic stress. In the absence of anisotropic stress, one would have $\Phi = -\Psi$. However, relativistic species such as photons and neutrinos possess a non-vanishing quadrupole moment ($\Theta_{r,2}$), which generates anisotropic stress and leads to a difference between the two metric potentials. A useful combination of the perturbed Einstein equations is obtained by eliminating time derivatives of the potentials, yielding

$$k^2\Phi = 4\pi G a^2 \left[\rho_{\rm m}\, \delta_{\rm m} + 4\rho_{\rm r}\, \Theta_{r,0} + \frac{3\mathcal{H}}{k}\left(i\rho_{\rm m}\, u_{\rm m} + 4\rho_{\rm r}\, \Theta_{r,1}\right)\right] \; . \tag{2.35}$$

This expression makes explicit that the gravitational potential is sourced not only by density perturbations, but also by velocity perturbations. The latter contribution becomes particularly relevant on large scales, where velocity flows contribute significantly to the gravitational field.

Solving the Einstein–Boltzmann system is challenging and computationally demanding, as it involves a large number of coupled integro-differential equations describing the evolution of multiple interacting species across a wide range of scales. As a result, accurate solutions rely on sophisticated numerical implementations, such as the Boltzmann codes `CAMB` (Lewis et al. 2000) and `CLASS` (Lesgourgues 2011).

2.1.3 Growth of matter perturbations

Having established the Einstein–Boltzmann system governing the evolution of cosmological perturbations, we now focus on the growth of matter perturbations, which are responsible for the formation of LSS in the Universe. A key aspect of the problem is that the evolution of perturbations depends both on time and on spatial scale. In particular, the behaviour of a given Fourier mode is determined by the relation between its wavelength and the comoving horizon. Modes that enter the horizon during radiation domination evolve differently from those entering during matter domination. In the following, we focus on sub-horizon scales at late times, where matter perturbations dominate the dynamics. In this case, the energy density of radiation becomes negligible compared to that of matter, and the evolution of the gravitational potential is driven by matter perturbations. Moreover, on sub-horizon scales ($k \gg \mathcal{H}$), velocity terms are suppressed by a factor $\mathcal{H}/k \ll 1$. Under these conditions, Eq. (2.35) reduces to the Poisson equation

$$k^2\Phi = 4\pi G a^2 \rho_{\rm m}\, \delta_{\rm m}\,. \tag{2.36}$$

This equation relates the gravitational potential to the matter density contrast and represents the Newtonian limit of the Einstein equations in an expanding Universe. The evolution of matter perturbations is then determined by combining Eq. (2.36) with the continuity and Euler equations for CDM, Eq. (2.18) and Eq. (2.19). Eliminating the velocity field $u_{\rm c}$, one obtains a second-order differential equation for the density contrast $\delta_{\rm c}$. It is convenient to express time derivatives in terms of the logarithm of the scale factor, $x \equiv \ln a$. Furthermore, at late times and on linear scales, baryons fall into the same gravitational potential wells as CDM. As a result, their density contrasts become approximately equal and the equation can be written in terms of $\delta_{\rm m}$, taking the form

$$\frac{{\rm d}^2\delta_{\rm m}}{{\rm d}x^2} + \left(2 + \frac{{\rm d}\ln H}{{\rm d}x}\right)\frac{{\rm d}\delta_{\rm m}}{{\rm d}x} - \frac{3}{2}\Omega_{\rm m}(a)\,\delta_{\rm m} = 0\,. \tag{2.37}$$

Equation (2.37) is the fundamental equation governing the linear growth of matter perturbations. The first term describes the acceleration of the perturbation growth, the second term represents a friction term due to the expansion of the Universe, and the last term encodes the effect of gravitational instability. The factor $\Omega_{\rm m}(a)$ explicitly shows that the growth rate depends on the relative abundance of matter. When matter dominates, gravitational attraction efficiently amplifies perturbations, while in epochs where other components (such as radiation or dark energy) are important, the growth is suppressed. In the matter-dominated era, where $\Omega_{\rm m}(a) \approx 1$ and $H \propto a^{-3/2}$, Eq. (2.37) admits simple analytic solutions. In this limit, one finds

$$\delta_{\rm m}(a) \propto a \quad (\text{growing mode})\,, \qquad \delta_{\rm m}(a) \propto a^{-3/2} \quad (\text{decaying mode})\,. \tag{2.38}$$

The growing solution is called growth factor $D(a)$. It corresponds to the gravitational amplification of density perturbations and is responsible for structure formation, while the decaying mode rapidly becomes negligible.

2.1.4 The transfer function

While Eq. (2.37) describes the time evolution of matter perturbations, it does not capture their dependence on spatial scale. However, as discussed above, the evolution of perturbations depends on when a given mode enters the horizon and on the background composition of the Universe at that time. As a result, perturbations on different scales do not grow in the same way. In particular, modes that enter the horizon during radiation domination experience a suppression of growth compared to those entering during matter domination. This scale-dependent evolution is encoded in the transfer function. The transfer function $T(k)$ is defined as the ratio between the late-time amplitude of a perturbation mode and its primordial value. It therefore quantifies how different Fourier modes are processed by the physical evolution of the Universe. In general, the matter density contrast can be written as

$$\delta_{\rm m}(k, a) = T(k)\, D(a)\, \delta_{\rm prim}(k)\,, \tag{2.39}$$

where $\delta_{\rm prim}(k)$ encodes the primordial fluctuations, $T(k)$ describes the scale dependence imprinted by early-time physics, and $D(a)$ accounts for the subsequent time evolution. The transfer function arises from the different evolution of perturbations before matter domination. On large scales, modes enter the horizon after matter-radiation equality and grow unimpeded. On small scales, instead, modes enter the horizon during radiation domination, when the pressure of relativistic species prevents gravitational collapse. As a result, the growth of these modes is suppressed until matter domination, leading to a characteristic scale dependence in the matter distribution. At sufficiently late times, when the Universe is matter dominated or dark-energy dominated, the evolution of perturbations becomes approximately scale independent on linear scales. In this regime, all modes grow at the same rate, which is described by the linear growth factor $D(a)$ introduced above. A quantitative illustration of the scale dependence introduced by the transfer function will be presented in Sect. 2.2.1, in the context of the matter power spectrum decomposition.

2.1.5 The growth factor

The growing–mode solution of Eq. (2.37) defines the linear, scale-independent growth factor, $D(a)$, which describes the evolution of matter density perturbations in the linear regime. Its behaviour depends on the underlying cosmological model and, in particular, on the expansion history of the Universe. As discussed in Sect. 1.3, the latter is determined by the energy content of the Universe through the Hubble parameter $H(a)$. This dependence is explicitly reflected in Eq. (2.37), where the friction term is proportional to $H(a)$.

Defining $E(a) \equiv H(a)/H_0$, in the case of ΛCDM and $k\Lambda$CDM models, the growth factor can be written as (Heath 1977):

$$D(a) = \frac{5\Omega_{\rm m} E(a)}{2} \int_0^a \frac{{\rm d}a'}{[a' E(a')]^3}\,, \tag{2.40}$$

where the normalisation is typically chosen such that $D(a = 1) = 1$.

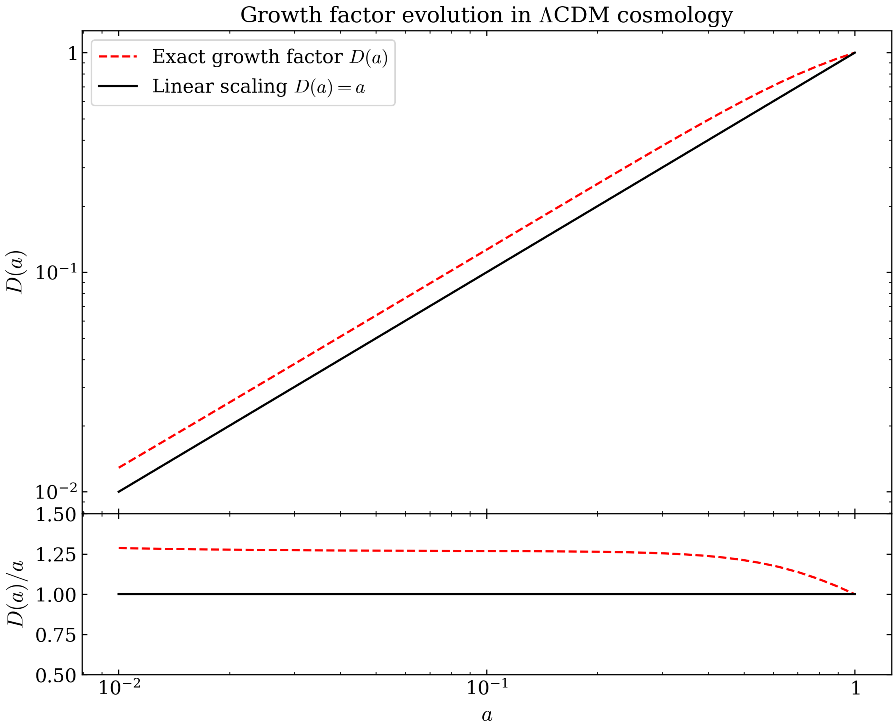

Figure 2.1: Evolution of the linear growth factor in a Planck18 ΛCDM cosmology (Planck collaboration 2020). The black solid lines show the linear scaling $D(a) = a$ expected during matter domination, while the red dashed lines show the exact growth factor. The upper panel displays $D(a)$ as a function of the scale factor, whereas the lower panel shows the ratio $D(a)/a$, highlighting deviations from the linear behaviour at late times.

For more general cosmological models, such as w_0w_aCDM, this integral expression is no longer valid, and Eq. (2.37) must be solved numerically. Moreover, Eq. (2.40) itself requires a numerical evaluation of the integral that, by construction, leads to a normalised growth factor. For consistency, and since later in this work we will require an un-normalised solution, we adopt a unified approach and compute $D(a)$ by directly solving Eq. (2.37) for all cosmological models. This choice ensures higher numerical accuracy (see Sect. 4.2). Figure 2.1 shows the evolution of the linear growth factor in a ΛCDM cosmology. During the matter-dominated era, density perturbations grow approximately proportionally to the scale factor due to gravitational instability. At later times, the accelerated expansion driven by dark energy increases the Hubble friction term in Eq. (2.37), suppressing the growth of structures. As a consequence, the growth factor departs from the linear scaling and the ratio $D(a)/a$ decreases towards the present epoch.

It is often convenient to introduce the growth rate

$$f(a) \equiv \frac{\mathrm{d} \ln D(a)}{\mathrm{d} \ln a} \,. \tag{2.41}$$

For a wide class of cosmological models, including those with dynamical dark energy, $f(a)$ is well approximated by the empirical relation (Heath 1977):

$$f(a) \approx [\Omega_{\mathrm{m}}(a)]^{\gamma} \,, \tag{2.42}$$

where γ is about 0.55 for GR.

2.2 Matter clustering statistics

In order to exploit the predictive power of linear perturbation theory and compare theoretical predictions with observations, it is necessary to describe the statistical properties of the density field. For simplicity, in the following we omit the subscript m and denote the matter density contrast as $\delta(\boldsymbol{x}, t)$, which we model as a stochastic field. On large scales, a common assumption is that the density contrast follows a Gaussian random field. In this case, the value of δ_i at a given comoving position $\boldsymbol{x}_i$ and time t is described by a Gaussian probability distribution function:

$$p(\delta_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\delta_i - \bar{\delta})^2}{2\sigma^2}\right] , \tag{2.43}$$

where $\bar{\delta} = 0$ due to statistical homogeneity, and $\sigma^2 = \langle\delta^2\rangle$ is the variance.

A particular realisation of the density field is specified by a set of N values $\{\delta_1, \delta_2, \ldots, \delta_N\}$, whose joint probability distribution is given by a multivariate Gaussian:

$$p(\delta_1, \ldots, \delta_N) = \frac{1}{\sqrt{|2\pi C|}} \exp\left[-\frac{1}{2}\delta_i C_{ij}^{-1} \delta_j\right] , \tag{2.44}$$

where C_{ij} is the covariance matrix, which quantifies the correlations between density fluctuations at different positions.

We now introduce two assumptions. The first is statistical homogeneity and isotropy (or stationarity). In the cosmological context it implies that the statistical properties of the field are invariant under spatial translations and under rotations, so that the covariance depends only on the modulus of the separation between points, $r = |\boldsymbol{x}_i - \boldsymbol{x}_j|$:

$$C_{ij} = C(r) . \tag{2.45}$$

The second is ergodicity, which implies that ensemble averages can be replaced by spatial averages over a sufficiently large volume. Physically, this means that the statistical properties of the Universe can be inferred from a single realisation, provided that it is large enough to be representative. In the limit $N \to \infty$, the covariance matrix becomes a continuous function of the separation r, and defines the 2PCF:

$$\xi(r) \equiv \langle\delta(\boldsymbol{x}, t)\, \delta(\boldsymbol{x} + \boldsymbol{r}, t)\rangle . \tag{2.46}$$

The 2PCF provides a direct measure of matter clustering. In particular, $\xi(r) > 0$ indicates an excess probability, with respect to a random distribution, of finding pairs of matter elements separated by a distance r, while $\xi(r) < 0$ corresponds to a deficit of such pairs. Therefore, the 2PCF quantifies the strength and scale dependence of spatial correlations in the matter distribution.

Owing to the fact that each Fourier mode corresponds to a well-defined physical scale and that spatial derivatives reduce to algebraic operators, it is convenient to express the density contrast as a superposition of plane waves:

$$\delta(\boldsymbol{x}, t) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3k \, \delta(\boldsymbol{k}, t) \, \exp\left[-i\boldsymbol{k} \cdot \boldsymbol{x}\right] \, , \tag{2.47}$$

where $\boldsymbol{k}$ is the comoving wavenumber.

The power spectrum $P(k)$ is defined as

$$\langle \delta(\boldsymbol{k}) \delta^*(\boldsymbol{k}') \rangle = (2\pi)^3 P(k) \, \delta_D(\boldsymbol{k} - \boldsymbol{k}') \, , \tag{2.48}$$

where δ_D is the Dirac delta function. It can then be shown that the 2PCF is the Fourier transform of the power spectrum:

$$\xi(r) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3k \, P(k) \, \exp\left[-i\boldsymbol{k} \cdot \boldsymbol{r}\right] \, . \tag{2.49}$$

Under the assumption of statistical isotropy, the power spectrum depends only on the magnitude $k = |\boldsymbol{k}|$.

2.2.1 Components of the matter power spectrum

In the standard cosmological framework, the matter power spectrum can be interpreted as the result of three main contributions: the primordial fluctuations, the early–time evolution encoded in the transfer function, and the late–time growth of structures.

Many inflationary models predict a primordial power spectrum of the form (Guth & Pi 1982)

$$P_{\mathrm{prim}}(k) = A_s \left(\frac{k}{k_p} \right)^{n_s - 1} \, , \tag{2.50}$$

where A_s is the scalar amplitude, typically of order 10^{-9}, n_s is the spectral index, which is close to unity, and k_p is a pivot scale, conventionally fixed at $0.05\,\mathrm{Mpc}$. This spectrum arises from quantum fluctuations generated during inflation and provides the initial conditions for the formation of cosmic structures. The subsequent evolution of these perturbations is affected by several physical processes occurring before matter domination. These effects modify the primordial spectrum in a scale-dependent way and are encoded in the transfer function $T(k)$ (see also Eq. 2.39), which determines the shape of the power spectrum. Fig. 2.2 shows the transfer function computed with `CAMB` for a ΛCDM cosmology, obtained from the decomposition of the linear matter power spectrum by removing the contribution of the primordial spectrum and the linear growth factor. On large scales, perturbation modes enter the horizon after matter–radiation equality and therefore preserve their primordial amplitude. On small scales, modes enter the horizon during radiation domination and experience a suppression of growth due to the relativistic pressure of radiation, producing the characteristic scale dependence of $T(k)$. The overlap between the curves at different redshifts shows that, in the linear regime, the transfer function is approximately time independent and encodes only the scale dependence imprinted by early–time physics.

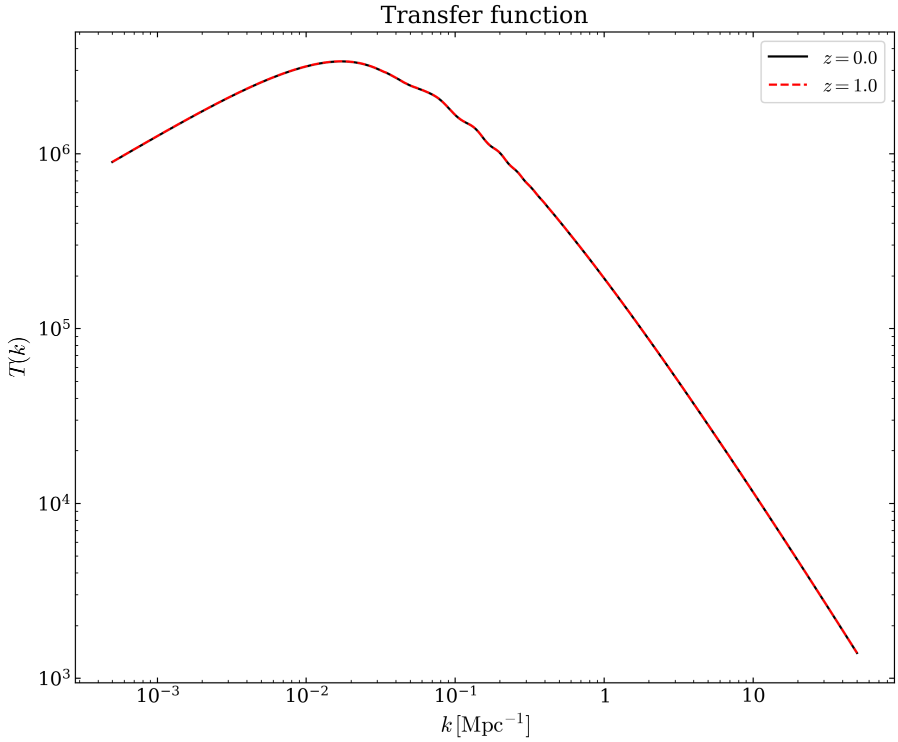

Figure 2.2: Transfer function computed from the linear matter power spectrum generated with `CAMB` for a Planck18 ΛCDM cosmology at redshifts $z = 0$ and $z = 1$. The overlap of the curves demonstrates the approximate time independence of the transfer function in the linear regime.

At late times, in the linear regime, the growth of density perturbations is driven by gravity and is independent on scale. As a result, each Fourier mode evolves independently, and the time dependence of the power spectrum can be described by the linear growth factor $D(t)$. Combining these contributions, the matter power spectrum can be written as

$$P(k,t) = A_s \left(\frac{k}{k_p}\right)^{n_s-1} T^2(k)\, D^2(t) \; . \tag{2.51}$$

This expression shows that the transfer function adjusts the scale dependence and thus the shape of the power spectrum, while the growth factor controls its time evolution. In the linear regime, the shape of the power spectrum remains unchanged, and only its amplitude evolves with time, so the power spectrum scales with the growth factor. Figure 2.3 shows the evolution of the linear matter power spectrum with redshift in a ΛCDM cosmology.

This expression shows that the transfer function adjusts the scale dependence, and therefore the shape, of the power spectrum, while the growth factor governs its time evolution. In the linear regime, the growth factor is scale independent, so the shape of the power spectrum remains approximately unchanged and only its amplitude evolves with time. Figure 2.3 illustrates this behaviour for a ΛCDM cosmology: the decrease of the power spectrum amplitude at increasing redshift

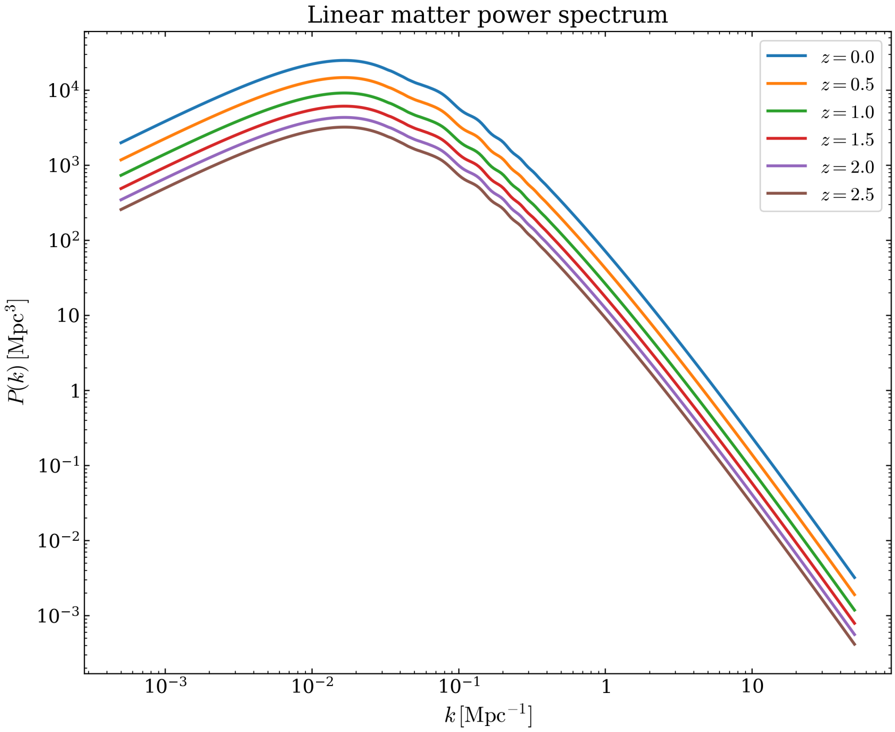

Figure 2.3: Linear matter power spectrum computed with `CAMB` for a Planck18 ΛCDM cosmology at different redshifts.

reflects the linear growth of matter perturbations, while the near-invariant shape confirms that the time evolution is primarily driven by the scale-independent growth factor.

2.2.2 The mass variance

The variance of the density field is defined as

$$\sigma^2(t) = \langle \delta(\boldsymbol{x}, t)^2 \rangle \; . \tag{2.52}$$

It can be shown that the variance is related to the power spectrum through

$$\sigma^2(t) = \frac{1}{2\pi^2} \int \mathrm{d}k \, k^2 \, P(k, t) \; . \tag{2.53}$$

This relation shows that the power spectrum quantifies the contribution to the variance of the density field per unit interval in wavenumber. Since the wavenumber is inversely related to the spatial scale, small (large) values of k correspond to large (small) physical scales. It is often useful to consider the density contrast field smoothed over a spherical region of radius R:

$$\delta_R(\boldsymbol{x}, t) = \int \mathrm{d}^3 y \, \delta(\boldsymbol{y}, t) \, W(|\boldsymbol{x} - \boldsymbol{y}|) \; , \tag{2.54}$$

where W is a window function, which weights the density contrast at position $\boldsymbol{y}$ based on its distance from $\boldsymbol{x}$. The variance of the smoothed density field over a sphere of radius R, commonly referred to as the mass variance, is given by

$$\sigma_R^2(t) = \frac{1}{2\pi^2} \int \mathrm{d}k\, k^2\, P(k,t)\, |\tilde{W}(kR)|^2 \ , \tag{2.55}$$

where $\tilde{W}$ is the Fourier transform of the window function. The window function acts as a low-pass filter, such that fluctuations with $k \gtrsim R^{-1}$ are suppressed, while only modes with $k \lesssim R^{-1}$ contribute to the variance. The quantity σ_R represents the root-mean-square amplitude of density fluctuations smoothed over a scale R. Physically, it quantifies the typical level of inhomogeneity of the matter distribution within spherical regions of radius R. Values $\sigma_R \ll 1$ correspond to the linear regime, where perturbations are small, while $\sigma_R \gtrsim 1$ indicates the onset of nonlinear structure formation. A commonly used quantity is σ_8, defined as the root-mean-square fluctuation of the matter density field smoothed over a sphere of radius $R = 8\, h^{-1}\,\mathrm{Mpc}$. It is widely used as a normalisation parameter of the matter power spectrum in cosmological analyses. An alternative definition of the mass variance, independent of the parameter h, is σ_{12} (see Sect. 3.3), corresponding to the mass variance smoothed on a physical scale of $12\,\mathrm{Mpc}$ (Sánchez 2020). This definition is equivalent to expressing σ_8 in physical units assuming the Planck best fit value h around 0.67 (Planck collaboration 2020).

2.2.3 Observational effects on the power spectrum

In order to compare theoretical predictions of the matter power spectrum with real data sets, several observational effects must be taken into account. In practice, galaxy redshift surveys do not directly probe the total matter density contrast $\delta(\boldsymbol{x}, z)$, but rather the galaxy density contrast $\delta_g(\boldsymbol{x}, z)$, which is related to the underlying matter field through a bias relation. In the simplest approximation, this relation is assumed to be linear and scale-independent:

$$\delta_g(\boldsymbol{x}, z) = b(z)\, \delta(\boldsymbol{x}, z) \ , \tag{2.56}$$

where $b(z)$ is the galaxy bias, which depends on redshift and on the properties of the galaxy sample. As a consequence, in Fourier space the galaxy power spectrum is related to the matter power spectrum by a factor b^2.

A second important effect arises from the fact that galaxy positions are inferred from their observed redshifts, which include contributions from both the Hubble flow and the peculiar velocities of galaxies. Since peculiar velocities are sourced by the underlying matter distribution, they introduce anisotropies in the observed clustering pattern. This effect is known as redshift–space distortions (RSD). In linear theory, the effect of RSD on the density contrast in Fourier space can be written as

$$\delta_{\mathrm{g,RSD}}(\boldsymbol{k}, z) = \left[b(z) + f(z)\, \mu^2\right] \delta(\boldsymbol{k}, z) \ , \tag{2.57}$$

where $\mu \equiv \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{n}}$ is the cosine of the angle between the wavevector and the line of sight, and $f(z)$ is the linear growth rate. As a result, the observed power spectrum becomes anisotropic and depends on both k and μ.

A further geometrical effect arises when converting redshifts into distances, which requires the assumption of a fiducial cosmology. If the assumed cosmology differs from the true one, the inferred distances are distorted. This is known as the Alcock–Paczynski (AP) effect (Alcock & Paczynski 1979). Let us denote by $\bar{z}$ the observed redshift, by $\chi_{\rm fid}(\bar{z})$ the comoving distance in the fiducial cosmology, and by $H_{\rm fid}(\bar{z})$ the corresponding Hubble parameter. The distortions can be parametrised through the scaling factors

$$\alpha_\perp(\bar{z}) \equiv \frac{\chi(\bar{z})}{\chi_{\rm fid}(\bar{z})} \ , \qquad \alpha_\parallel(\bar{z}) \equiv \frac{H_{\rm fid}(\bar{z})}{H(\bar{z})} \ , \tag{2.58}$$

which relate transverse and radial distances in the true and fiducial cosmologies. As a consequence, the observed wavevector components are rescaled as

$$k_\perp = \frac{k_{\perp,\rm obs}}{\alpha_\perp} \ , \qquad k_\parallel = \frac{k_{\parallel,\rm obs}}{\alpha_\parallel} \ . \tag{2.59}$$

Combining these effects, the observed galaxy power spectrum in redshift space can be written as

$$P_{\rm g,RDS}(k_{\rm obs}, \mu, \bar{z}) = \frac{1}{\alpha_\perp^2 \alpha_\parallel} \left[b(\bar{z}) + f(\bar{z})\,\mu^2\right]^2 P(k, \bar{z}) + P_N \ , \tag{2.60}$$

where $P_N = 1/\bar{n}_g$ is the shot noise, assumed to be scale-independent.

2.3 Cosmological inference with the matter power spectrum

The statistical description of matter clustering developed in Sect. 2.2 provides the foundation for connecting cosmological models to observations. In particular, the matter power spectrum plays a central role in this framework, as it not only characterises the distribution of matter on large scales, but also acts as a fundamental building block for several cosmological observables. In practice, measurements of LSS are not limited to the power spectrum itself, but are often expressed in terms of summary statistics, such as the multipoles of the galaxy power spectrum, or related probes like the halo mass function (HMF) and the void size function (VSF), which will be discussed more in detail in Sect. 2.3.1. These observables are all sensitive, directly or indirectly, to the underlying matter power spectrum, which therefore provides a unified link between cosmological parameters and observational data. The comparison between theory and observations requires accurate predictions for these quantities, which are obtained by modelling the evolution of cosmological perturbations. At the same time, cosmological inference relies on exploring a high-dimensional parameter space through statistical methods, making computational efficiency a crucial aspect of the analysis. In the following, we describe the role of the power spectrum in cosmological data analysis and how statistical methods are used to extract cosmological information from clustering measurements.

2.3.1 Summary statistics based on the matter power spectrum

On sufficiently large scales, where density fluctuations remain small, the evolution of cosmic structures can be accurately described within linear perturbation theory. In this regime, the matter power spectrum is fully characterised by the linear growth of density perturbations and can be predicted by solving the Einstein–Boltzmann system. However, as structures evolve, gravitational collapse induces mode coupling and drives the growth of nonlinear features, causing the linear approximation to break down on progressively smaller scales. As a consequence, a distinction must be made between the linear matter power spectrum, which describes the statistics of the density field in the linear regime, and the nonlinear matter power spectrum, which accounts for the effects of nonlinear gravitational evolution.

Modern cosmological analyses often exploit information from mildly nonlinear scales in order to improve parameter constraints. This requires accurate predictions of the nonlinear matter power spectrum. Rather than solving the fully nonlinear dynamics directly, Boltzmann solvers such as `CAMB` and `CLASS` typically employ fitting prescriptions calibrated on N-body simulations, the most widely used being `Halofit` (Smith et al. 2003). The resulting nonlinear power spectrum plays a central role in weak lensing models and can be integrated in full-shape analyses of galaxy clustering, where nonlinear gravitational evolution significantly affects the observed signal. Nevertheless, as will be discussed below, several cosmological probes remain fundamentally connected to the linear matter power spectrum through quantities derived from the variance of the density field. Fig. 2.4 shows the difference between the linear and nonlinear power spectrum. On large scales, the two predictions coincide, reflecting the validity of linear perturbation theory. On progressively smaller scales, nonlinear gravitational evolution enhances the clustering amplitude, causing the nonlinear power spectrum to deviate from the linear prediction.

A primary application of the matter power spectrum is the analysis of galaxy clustering in redshift space. As we saw in Sect. 2.2.3, the presence of RSD introduces an angular dependence in the galaxy power spectrum. It is then convenient to decompose it into multipoles using Legendre polynomials:

$$P_{\mathrm{g,RDS}}(k_{\mathrm{obs}}, \mu) = \sum_{\ell} P_\ell(k)\, \mathcal{L}_\ell(\mu) \ , \tag{2.61}$$

where $\mathcal{L}_\ell(\mu)$ are the Legendre polynomials, and the multipole moments are defined as

$$P_\ell(k) = \frac{2\ell+1}{2} \int_{-1}^{1} \mathrm{d}\mu\, P_{\mathrm{g,RDS}}(k_{\mathrm{obs}}, \mu)\, \mathcal{L}_\ell(\mu) \ . \tag{2.62}$$

In practice, only the first few multipoles are relevant. The monopole ($\ell = 0$) corresponds to the angle-averaged power spectrum, while the quadrupole ($\ell = 2$) and hexadecapole ($\ell = 4$) encode the anisotropic signal induced by RSD. This decomposition provides a convenient way to compress the angular information of the power spectrum and allows cosmological information to be extracted from the full clustering signal. In modern analyses, the theoretical modelling of these multipoles relies on the linear matter power spectrum and includes nonlinear corrections

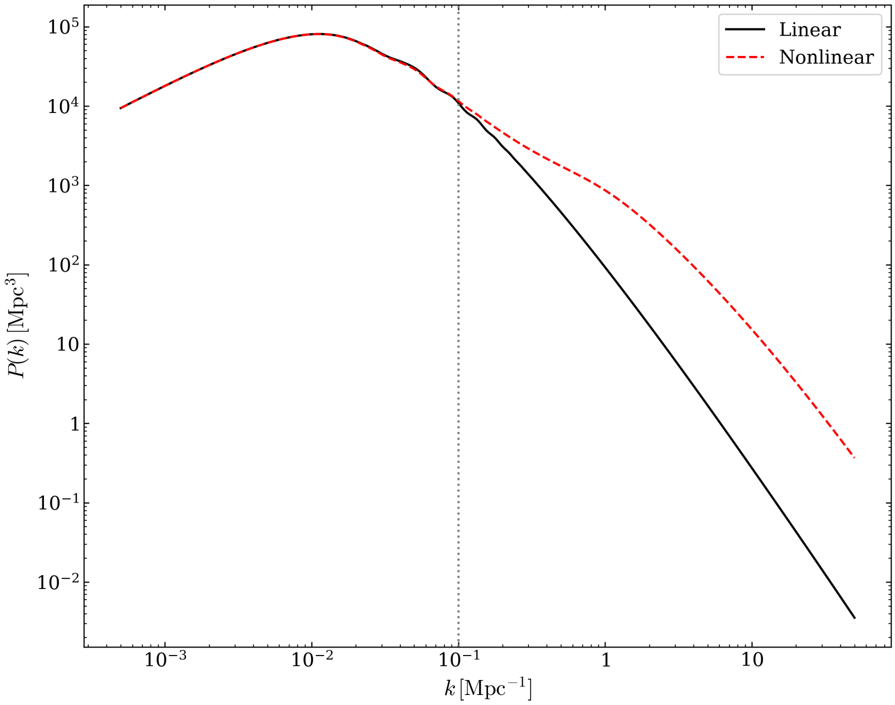

Figure 2.4: Comparison between the linear (solid black line) and nonlinear (dashed red line) matter power spectra at $z = 0$, computed with `CAMB`. The vertical dotted line at $k = 0.1\,\mathrm{Mpc}^{-1}$ approximately marks the transition beyond which nonlinear effects become increasingly important.

to both matter clustering and RSD, particularly on intermediate and small scales where the linear approximation is no longer adequate. In particular, the amplitude of the monopole is sensitive to the combination $b^2P(k)$, while the anisotropic signal encoded in higher–order multipoles provides direct information on the growth of structures through the parameter $f(z)$. However, due to degeneracies between the bias and the amplitude of matter fluctuations, galaxy clustering measurements typically constrain the combination $f\sigma_8(z)$ rather than $f(z)$ alone. Furthermore, the AP effect introduces distortions that depend on the expansion history of the Universe, allowing measurements of the Hubble parameter $H(z)$ and the comoving distance $\chi(z)$. Figure 2.5 shows the monopole and quadrupole measurements obtained from the Baryon Oscillation Spectroscopic Survey (BOSS) data (Alam et al. 2017). The detection of a non–zero quadrupole provides direct evidence for anisotropy induced by RSD and allows constraints on the growth rate of cosmic structures through measurements of $f\sigma_8(z)$.

A complementary probe of cosmology is provided by the HMF, whose theoretical modelling is fundamentally connected to the mass variance derived from the linear matter power spectrum. Dark matter haloes host the formation of galaxies and galaxy clusters, and their abundance as a function of mass and redshift provides a powerful probe of cosmology. In particular, the number density of massive haloes is highly sensitive to key cosmological parameters such as the matter density parameter

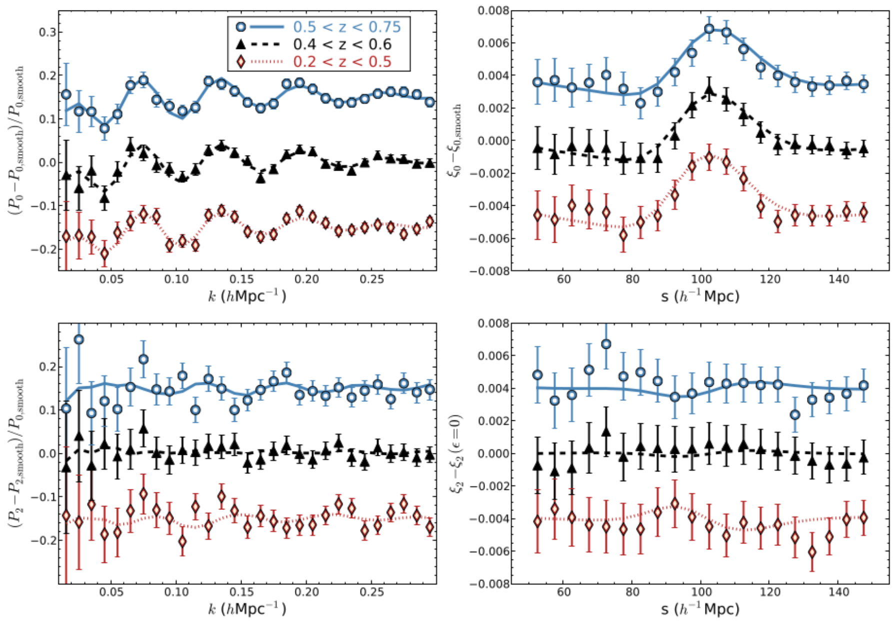

Figure 2.5: Measurements of the monopole and quadrupole of the galaxy clustering signal from the BOSS survey in different redshift bins. The left panels show the power spectrum multipoles, while the right panels show the corresponding 2PCF multipoles in configuration space. The monopole, $P_0(k)$ and $\xi_0(s)$, represents the angle–averaged clustering signal, whereas the quadrupole, $P_2(k)$ and $\xi_2(s)$, quantifies the anisotropy generated by RSD. The quantities are plotted relative to smooth reference models in order to highlight the oscillatory BAO feature and the anisotropic clustering signal. Credit: Alam et al. (2017).

$\Omega_{\rm m}$ and the amplitude of matter fluctuations σ_8. The connection between the HMF and the linear matter power spectrum arises through the variance of the density field. Denoting the comoving number density of dark matter haloes of mass M at redshift z as $n(M, z)$, the HMF can be written as (Bocquet et al. 2015)

$$\frac{{\rm d}n}{{\rm d}M} = f(\sigma)\,\frac{\bar{\rho}_{\rm m}}{M}\,\frac{{\rm d}\ln\sigma^{-1}}{{\rm d}M}\ , \tag{2.63}$$

where $\bar{\rho}_{\rm m}$ is the present-day mean matter density, $\sigma^2(M, z)$ is the variance of the matter density field smoothed on a scale corresponding to the mass M (see Eq. 2.55), and $f(\sigma)$ is the multiplicity function. The latter encodes the fraction of matter that collapses into haloes and is typically calibrated using numerical simulations (see e.g. Tinker et al. 2008; Fiorilli et al. 2025). Eq. (2.63) shows that the abundance of haloes is directly determined by the amplitude of density fluctuations: large values of $\sigma(M)$ correspond to more efficient structure formation, leading to a higher number of massive haloes. Conversely, rare massive clusters correspond to the high-mass tail of the distribution, where the variance is small and the abundance is exponentially suppressed. Since the quantity $\sigma(M, t)$ is obtained by smoothing the linear matter power spectrum on a given scale, the cosmological dependence of the HMF is fundamentally encoded in the statistics of the linear density field.

In addition to collapsed structures, the large-scale distribution of matter also contains extended underdense regions known as cosmic voids. These are large regions, typically with sizes in the range 10–100 h^{-1} Mpc, that are nearly devoid of galaxies, approximately spherical, and occupy a significant fraction of the volume of the Universe (Contarini et al. 2026). Voids originate from the evolution of the negative initial density fluctuations and expand over time, becoming a prominent component of the LLS. The first-order statistical properties of voids are commonly described by the VSF, which quantifies their abundance as a function of size. Similarly to the HMF, the VSF is connected to the statistical properties of the underlying linear density field and depends on the variance σ^2 of the density fluctuations, making the VSF fundamentally linked to the linear matter power spectrum (Sheth & van de Weygaert 2004). As a consequence, it is sensitive to both the growth of structures and the expansion history of the Universe, through quantities such as the growth factor and the Hubble parameter. Owing to their low-density environments, voids are less affected by nonlinear gravitational evolution and galaxy bias, and therefore provide a relatively clean probe of the matter distribution. In addition, dark energy influences the late-time expansion rate, thereby affecting the evolution and characteristic sizes of voids.

Together, these probes provide a broad picture of LSS. The multipoles of the galaxy power spectrum constrain both the growth of cosmic structures and the geometry of the Universe, whereas the HMF links the statistical properties of the density field to the abundance of collapsed objects. Complementarily, the VSF traces the distribution of underdense regions, adding sensitivity to the growth of structures and the properties of dark energy.

2.3.2 Cosmological costraints

Cosmological inference aims at constraining the parameters of a given cosmological model by comparing observational data with theoretical predictions. In the context of LSS, this is typically achieved by modelling clustering observables and confronting them with measurements from galaxy surveys and other cosmological probes. This procedure requires both an accurate theoretical description of the evolution of cosmological perturbations and a statistical framework to assess the agreement between models and data.

Boltzmann solvers are numerical tools designed to compute cosmological observables from a given set of parameters. In particular, they provide high-precision predictions for quantities such as the matter power spectrum and the angular power spectrum of the CMB. These predictions are obtained by solving the coupled system of Einstein and Boltzmann equations (see Sect. 2.1.2). As anticipated in the introduction, widely used Boltzmann solvers include `CAMB` (Lewis et al. 2000) and `CLASS` (Lesgourgues 2011), which are standard tools in modern cosmology.

To quantify the agreement between models and data, a statistical framework is required. The cornerstone of modern cosmological analyses is Bayesian inference. Given a data set D and a set of model parameters $\boldsymbol{\theta}$, the likelihood is defined as the probability of observing the data given the model, $\mathcal{L}(D|\boldsymbol{\theta})$. Within a Bayesian framework, the posterior probability distribution of the parameters is obtained through Bayes' theorem,

$$\mathcal{P}(\boldsymbol{\theta}|D) = \frac{\mathcal{L}(D|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\mathcal{E}(D)}\,, \tag{2.64}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution and $\mathcal{E}(D)$ is the evidence, which is independent of the model parameters. An estimate of the model parameters can be obtained by maximising the posterior distribution. The associated uncertainties are determined by the shape of the posterior around its maximum. Assuming a Gaussian approximation, the covariance matrix of the parameters is given by the inverse of the Fisher information matrix (Fisher 1925),

$$\mathcal{F}_{ij} \equiv -\left\langle \frac{\partial^2 \ln \mathcal{P}}{\partial\theta_i\,\partial\theta_j} \right\rangle\,, \qquad \mathbf{C} = \mathcal{F}^{-1}\,, \tag{2.65}$$

from which parameter uncertainties and correlations can be derived as $\sigma_i^2 = \mathcal{F}_{ii}^{-1}$. However, in realistic applications the posterior distribution is rarely Gaussian, and it often must be explored numerically.

This is typically achieved through sampling techniques such as MCMC methods, which allow one to reconstruct the full posterior distribution and to derive robust constraints on cosmological parameters (see Speagle 2020, for a review). MCMC methods provide an efficient way to sample complex posterior distributions in high-dimensional parameter spaces. Rather than evaluating the posterior on a regular grid, which rapidly becomes computationally prohibitive as the number of parameters increases, MCMC algorithms generate a sequence of points, known as a Markov chain, whose distribution converges to the target posterior. At each step, a new set of parameters is proposed and accepted or rejected according to a criterion that depends on the ratio of posterior probabilities between the proposed and current states. After an initial burn-in phase, during which the chain approaches the equilibrium distribution, the sampled points can be used to reconstruct the posterior probability density and to derive parameter estimates, confidence intervals, and parameter degeneracies. Widely used algorithms in cosmology include the Metropolis–Hastings (Metropolis et al. 1953; Hastings 1970) sampler. In this framework, the likelihood function must be evaluated a large number of times, often of the order of 10^5–10^6 evaluations, in order to efficiently sample the parameter space. Each likelihood evaluation requires the computation of the relevant theoretical observables, such as the matter power spectrum, for a given set of cosmological parameters. When these predictions are obtained using Boltzmann solvers, the computational cost becomes a major limitation, significantly increasing the time required for cosmological analyses. This issue is particularly relevant when combining multiple probes or when exploring extended cosmological models with a large number of free parameters.

The framework described above is widely employed in modern cosmological analyses. For instance, the full-shape modelling of the galaxy power spectrum derived from DESI (DESI Collaboration 2016) observations has been used to extract cosmological information from LSS data, providing constraints on the matter density, the amplitude of matter fluctuations, the dark energy equation of state, neutrino masses, and modified gravity (Adame et al. 2025a). Similarly, the information encoded in the matter power spectrum can be accessed through the cluster mass function, whose evolution traces the growth of density perturbations. For example, cluster abundances measured from the first eROSITA All-Sky Survey (Merloni et al. 2024) have been used to derive competitive constraints on cosmological parameters, including $\Omega_{\rm m}$, σ_8, neutrino masses, and the dark energy equation-of-state parameter (Ghirardini et al. 2024). Finally, the VSF measured from the BOSS DR12 galaxy catalogue has been used to constrain cosmological parameters such as $\Omega_{\rm m}$, σ_8, and the dark energy equation-of-state parameter, providing an independent test of the standard cosmological model (Contarini et al. 2023).

Chapter 3

Machine learning and evolution mapping framework

As we saw in Sect. 2.3.2, despite their accuracy, Boltzmann solvers are computationally expensive, as they require the numerical integration of a system of coupled differential equations for each set of cosmological parameters. This represents a major limitation for cosmological inference, where the theoretical model must be evaluated repeatedly in order to efficiently explore the parameter space within the Bayesian framework. To overcome this limitation, fast surrogate models, commonly referred to as emulators, have been developed. These approaches aim at accurately reproducing the output of Boltzmann solvers while drastically reducing the computational cost, thus enabling efficient exploration of the parameter space within Bayesian analyses. In recent years, machine learning (ML) techniques have emerged as a powerful tool for the construction of such emulators. In particular, supervised learning algorithms and, specifically, neural networks (NN) have proven to be highly effective in capturing the complex dependence of cosmological observables on the underlying model parameters. Several cosmological emulators have been developed to predict quantities such as the matter power spectrum with high accuracy. However, a major challenge in their construction is the high dimensionality of the cosmological parameter space, which typically requires a large number of training simulations or restricts the range of validity of the emulator.

In this context, alternative approaches can be employed to reduce the effective dimensionality of the problem by exploiting physical insights into the structure of the power spectrum. One such approach is the evolution mapping (Sánchez et al. 2022), which separates the dependence on different cosmological parameters into distinct contributions related to the shape and time evolution of the power spectrum. This method allows one to extend the applicability of emulators trained on a limited parameter space, while maintaining a high level of accuracy.

3.1 Introduction to machine learning

A ML algorithm is an algorithm designed to learn from data (Goodfellow et al. 2016). It enables the solution of tasks that are difficult to address using fixed programs explicitly designed by human beings. ML tasks are typically described in terms of how a system processes an example, where an example is defined as a collection of features that quantitatively describe a given object or event. ML algorithms can be broadly classified as unsupervised or supervised, depending on the type of information available during the learning process:

- Unsupervised learning algorithms are provided with a data set containing only input features and aim at identifying patterns or structures within the data.
- Supervised learning algorithms are trained on data sets in which each example is associated with a label or target value, with the goal of learning a mapping between inputs and outputs.

In the following, we focus on supervised learning, which is the approach we adopted in the construction of our emulator.

3.1.1 Linear regression

In its simplest form, a supervised learning problem can be described as the approximation of a function that maps an input vector $\boldsymbol{x} \in \mathbb{R}^n$ to an output $y \in \mathbb{R}$. A simple example is given by linear regression, where the model prediction can be written as

$$\hat{y} = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + b \,, \tag{3.1}$$

where $\boldsymbol{w} \in \mathbb{R}^n$, the vector of weights, and $b \in \mathbb{R}$, the bias term, are the parameters of the model. The quality of the prediction is quantified through a loss function, which measures the discrepancy between the predicted and true values. A commonly used choice is the mean squared error (MSE),

$$\mathrm{MSE} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \,, \tag{3.2}$$

evaluated over a data set of m examples. In this context, y is referred to as the label, $\boldsymbol{x}$ as the feature vector, and the set $D = \{\boldsymbol{x}_i, y_i\}$ as the training data set.

Although linear regression provides a simple and analytically tractable framework, it is inherently limited to linear mappings between inputs and outputs. This restriction makes it inadequate for modelling highly nonlinear functions, such as cosmological observables, which motivates the introduction of more flexible models in Sect. 3.1.3.

3.1.2 Training and validation

The training process consists in optimising the model parameters in order to minimise the loss function over the training data set, while ensuring that the resulting model generalises well to previously unseen data. In practice, this is achieved through iterative optimisation algorithms. Optimisation is the process of finding the set of parameters that minimises the loss function, which can be, for example, the MSE. A commonly used optimisation algorithm is gradient descent, which updates the model parameters along the direction of decreasing loss. Starting from an initial guess for the weights, $\boldsymbol{w}_1$, the loss function is evaluated over the training data set and its gradient is computed to determine the direction of steepest increase. The parameters are then updated as

$$\boldsymbol{w}_2 = \boldsymbol{w}_1 - \eta \nabla \mathrm{MSE} \,, \tag{3.3}$$

where η is a hyperparameter, called the learning rate, which controls the magnitude of the update. This procedure is iterated multiple times, and the full set of iterations is referred to as the training process, typically organised in epochs.

A central challenge in ML is that the algorithm must perform well on new, previously unseen inputs, and not only on the data used during training. This ability is referred to as generalisation. To quantify it, different types of errors are typically considered:

- the training error, defined as the discrepancy between the model predictions and the true values evaluated on the training data set;
- the test error, defined as the discrepancy between the model predictions and the true values evaluated on an independent data set that is not used during training.

A model is said to generalise well when the training error is sufficiently small and the gap between the training and test errors is limited. If the training error is large, the model fails to capture the underlying relation between inputs and outputs, a situation known as underfitting. Conversely, if the training error is small but the test error is significantly larger, the model is said to be affected by overfitting, meaning that it has learned the specific features of the training data rather than the underlying general relation. In practice, the available data set is typically split into training and validation sets. The training set is used to optimise the model parameters, while the validation set provides an independent assessment of the model performance and allows one to monitor overfitting. This procedure is generally referred to as model validation.

3.1.3 Neural network architecture

Generally, a simple linear regression is not sufficient to capture the complexity of cosmological observables such as the matter power spectrum. Therefore, a more flexible and nonlinear mapping is required in order to model these quantities accurately. A natural choice in this context is a deep NN. A NN consists of a sequence of layers, each composed of a set of neurons. Each layer takes as input the outputs of the previous layer and produces a new representation through a nonlinear transformation, defined by an activation function. The first layer corresponds to the input data, followed by a number of hidden layers, while the final layer provides the desired output. A schematic example of such an architecture is shown in Fig. 3.1. More specifically, the transformation performed by the l-th layer can be written as

$$\boldsymbol{h}^{(l)} = \sigma\left(W^{(l)}\boldsymbol{h}^{(l-1)} + \boldsymbol{b}^{(l)}\right) \,, \tag{3.4}$$

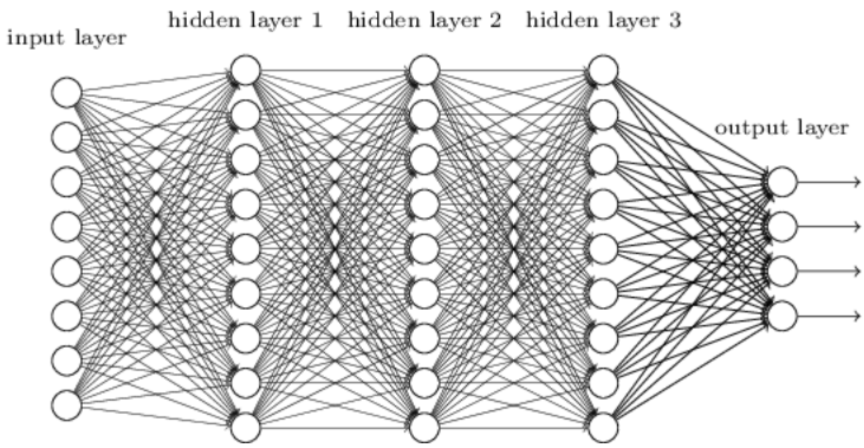

Figure 3.1: Schematic representation of a fully connected feed-forward NN. The input layer encodes the model features, which are propagated through multiple hidden layers via weighted connections and nonlinear activation functions. The final output layer provides the predicted quantities. Figure adapted from Nielsen (2015).

where $\boldsymbol{h}^{(l)}$ is the output of the layer, $W^{(l)}$ and $\boldsymbol{b}^{(l)}$ are the weight matrix and bias vector, respectively, and σ denotes a nonlinear activation function. The activation function is a key element of the network, since it introduces nonlinearity and enhances the expressive power of the model. In the context of cosmological emulators, a custom activation function is employed, in which additional trainable parameters are associated with each neuron. Different types of activation functions can be used in ML models. We employ the function defined in Spurio Mancini et al. (2022):

$$\sigma\left(\boldsymbol{h}^{(l)}\right) = \left[\boldsymbol{\beta} + (1-\boldsymbol{\beta}) \odot \left(1 + \exp\left[-\boldsymbol{\alpha} \odot \boldsymbol{h}^{(l)}\right]\right)^{-1}\right] , \tag{3.5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are trainable vector parameters of the activation function that allow the network to adapt the nonlinear mapping directly from the data, and the symbol $\odot$ is a element-wise product.

In this work, we consider fully connected feed-forward NN, where each neuron in a given layer is connected to all neurons in the adjacent layers. The number of hidden layers and the number of neurons in each layer define the architecture of the NN, given that the sizes of the input and output layers are fixed by the problem. The choice of architecture is crucial, as it determines the model complexity. An excessively complex architecture, with too many layers or neurons, may lead to overfitting, that is, the model fails to generalise to unseen data. Conversely, an overly simple architecture may result in underfitting, where the model is unable to capture the underlying structure of the data.

The training of a NN consists in optimising all trainable parameters, including weights, biases, and activation function coefficients, by minimising a loss function. However, the final performance of the model depends not only on these parameters, but also on a set of hyperparameters that define the training procedure. These include, for example, the learning rate, the batch size, and the number of training epochs, as well as the architecture of the network itself. In practice, these hyperparameters are tuned in order to achieve a balance between accuracy and generalisation.

These techniques provide a flexible and efficient framework to approximate the mapping between cosmological parameters and the theoretical quantities to be considered in a Bayesian framework, making them particularly well suited for the construction of fast and accurate emulators.

3.2 Cosmological emulators

In cosmological parameter inference pipelines based on two-point statistics, a major computational bottleneck arises from the repeated evaluation of theoretical matter power spectrum, which requires running Boltzmann solvers over a large set of cosmological parameters combinations. Cosmological observables such as the matter power spectrum are, however, smooth functions of the underlying cosmological parameters. This property makes them particularly well suited for emulation.

Cosmological emulators are surrogate models that approximate the mapping between cosmological parameters and observables. This approach is especially advantageous in Bayesian inference, where the posterior distribution is sampled through a large number of model evaluations. By replacing direct Boltzmann computations, emulators can reduce the computational cost by several orders of magnitude, while maintaining a level of accuracy sufficient for reliable parameter estimation. In practice, emulators are typically constructed by training ML models, such as NN, on a finite set of precomputed theoretical predictions, consisting of cosmological parameters as inputs and the corresponding observables as outputs. Once trained, the emulator can be efficiently interpolated within the parameter space, enabling rapid evaluations of the target observable. Despite these advantages, the construction of accurate emulators is not trivial, particularly due to the high dimensionality of the cosmological parameter space, which poses significant challenges for training and generalisation, as will be described in the following. Many cosmological emulators have nevertheless been developed in the literature, employing different techniques and targeting a variety of observables, as discussed in Sect. 3.2.2.

3.2.1 The curse of dimensionality

To illustrate the issue of high dimensionality, let us consider a function $f(\theta_1, \ldots, \theta_N)$ of N cosmological parameters, which we aim to approximate with an emulator. The integer N defines the dimensionality of the parameter space, where each parameter θ_i spans a finite interval $[\theta_{i,\mathrm{min}}, \theta_{i,\mathrm{max}}]$. The construction of an emulator requires evaluating the target function at a set of points sampling this parameter space. If each parameter range is discretised into m intervals, the total number of sampling points scales as m^N. Therefore, the volume of the parameter space grows exponentially with the number of free parameters, making a dense and uniform sampling rapidly impractical as N increases. In practice, an accurate training of the emulator requires a sufficiently dense coverage of the parameter space to avoid large interpolation errors and poor generalisation. This implies that the number of training samples cannot be too small, further exacerbating the computational cost.

Moreover, increasing the dimensionality of the parameter space generally increases the complexity of the mapping that the emulator must learn. Variations along a larger number of independent directions can induce more intricate dependencies of the target function on the input parameters, making the approximation problem intrinsically more challenging. As a consequence, achieving a given level of accuracy may require NN architectures with a larger number of layers, neurons, or trainable parameters. The optimisation of such models is typically more demanding and may require larger training sets to avoid overfitting and maintain good generalisation performance. These issues becomes even more severe when the training set is constructed from computationally expensive numerical evaluations, such as Boltzmann solvers. In these cases, each training sample requires a full numerical run, so that the total cost of generating the training set increases rapidly with the dimensionality of the parameter space, often becoming prohibitive.

A common strategy to mitigate this problem is to restrict the parameter space to a limited region of interest, for example by narrowing the allowed parameter ranges or fixing a subset of parameters. While this approach reduces the number of required training samples, it also limits the validity of the emulator, which can only be reliably used within the domain covered by the training data. These limitations motivate the development of alternative approaches, such as the evolution mapping method, which is discussed in Sect. 3.3.

3.2.2 Available cosmological emulators

A variety of emulators have been developed and applied to a wide range of astrophysical and cosmological problems. Among all emulators, we focus in particular on two representative suites, `CosmoPower` (Spurio Mancini et al. 2022) and `BACCO` (Aricò et al. 2022), which are particularly relevant for our analysis. Both provide emulators for the matter power spectrum, the key observable considered in this work.

`BACCO` is a suite of cosmological emulators designed to model LSS observables with high accuracy and computational efficiency. It provides predictions for the linear and nonlinear matter power spectrum and it includes emulators for related quantities, such as Lagrangian perturbation theory spectra and baryonic effects, providing a comprehensive description of structure formation. The models are trained over an extended cosmological parameter space comprising eight cosmological parameters: the baryon and CDM density parameters, $\Omega_{\rm b}$ and $\Omega_{\rm cdm}$, the dimensionless Hubble parameter h, the sum of neutrino masses M_ν, expressed in eV, and the parameters w_0 and w_a describing the time evolution of the dark energy equation of state through the CPL parametrisation. The parameter ranges span approximately 20 times the Planck uncertainty region around the best-fitting cosmology. The primordial amplitude $A_{\rm s}$ and spectral index $n_{\rm s}$ are instead incorporated through an analytical rescaling, avoiding the need to include these parameters in the NN training (see Sect. 3.3.1). The linear matter power spectrum is sampled at 600 wavenumbers in the range $k \in [10^{-4}, 50]\, h\,{\rm Mpc}^{-1}$. For 68% of the validation set, the emulator achieves an accuracy at the 0.1%–0.5% level over redshifts z smaller than nine. Evaluation times are of the order of milliseconds, making the emulator suitable for fast large-scale cosmological inference analyses.

`CosmoPower` is a suite of NN-based cosmological power spectrum emulators designed to accelerate parameter estimation from two-point statistics analyses of LSS and CMB surveys. For the matter power spectrum, the emulation error is below 0.4% over the range $k \in [10^{-5}, 10]\,\mathrm{Mpc}^{-1}$ and $z \in [0, 5]$. The emulators can recover fiducial cosmological constraints with differences in the posterior distributions smaller than the sampling noise, while achieving a speed-up of up to $\mathcal{O}(10^4)$ with respect to standard Boltzmann-based computations. In addition to pre-trained models, `CosmoPower` provides the tools to construct and train custom NN emulators over user-defined parameter spaces. This allows for flexible adaptation to specific applications, as demonstrated in this work, where a dedicated emulator has been developed (see Sect. 4).

A recent example based on the evolution-mapping framework is `Aletheia` (Sánchez et al. 2025), an emulator for the nonlinear matter power spectrum. The emulator is trained on a reduced parameter space and incorporates the effects of different expansion histories through a dedicated correction based on the integrated growth history. This approach enables accurate predictions over the range $0.006\,\mathrm{Mpc}^{-1} < k < 2\,\mathrm{Mpc}^{-1}$, achieving sub-percent accuracy with respect to N-body simulations and a prediction variance of approximately 0.2%. Other notable examples are `EuclidEmulator` (Knabenhans et al. 2019) and its successor `EuclidEmulator2` (Knabenhans et al. 2021), developed within the Euclid collaboration (Laureijs et al. 2011) to provide fast predictions of the nonlinear matter power spectrum. The former emulates a six-parameter cosmological model and achieves percent-level accuracy, whereas the latter extends the parameter space to include a time-dependent dark-energy equation of state and massive neutrinos, reaching sub-percent to percent-level accuracy over the range of scales and redshifts relevant for LSS analyses.

Finally, for a comprehensive list of available power spectrum emulators, we refer the reader to Tanaka et al. (2026).

3.3 Evolution mapping

As discussed in Sect. 3.2.1, the performance of emulators is fundamentally limited by the dimensionality of the parameter space explored during the training phase. In this work, we adopt the evolution mapping approach (Sánchez et al. 2022; Esposito et al. 2024; Fiorilli et al. 2025), to reduce the effective dimensionality of the problem by exploiting physical degeneracies among cosmological parameters. In order to build a linear power spectrum emulator, this is the optimal approach.

A key requirement of this method is the use of physical units. In standard practice, theoretical predictions are often expressed in units of $h^{-1}\,\mathrm{Mpc}$. However, in these units, parameters such as the dimensionless Hubble constant h enter both the shape and the amplitude of the power spectrum, breaking otherwise exact degeneracies. By expressing all quantities in physical units (i.e. Mpc), these degeneracies are restored. For example, parameters such as the scalar amplitude A_s and h contribute only to the overall normalisation of the linear power spectrum. A direct

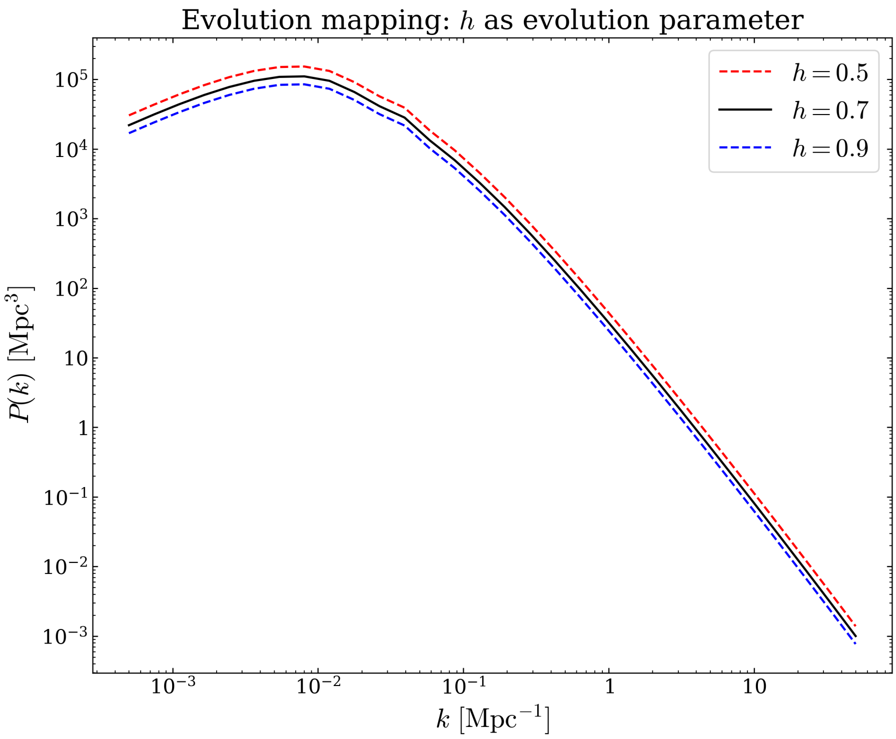

Figure 3.2: Linear matter power spectra computed with `CAMB` in physical units for different values of the dimensionless Hubble parameter h, within a ΛCDM cosmology. All other cosmological parameters are fixed to the Planck collaboration (2020) best-fit values. The spectra exhibit identical shapes and differ only by an overall normalisation, illustrating that h acts as an evolution parameter.

illustration of the evolution-mapping concept is shown in Fig. 3.2, where power spectra computed for different values of h share the same shape and differ only by a scale-independent normalisation. Following the same reasoning, the usage of physical energy densities, ω_s (see Eq. 1.21), instead of Ω_s, is of key importance to avoid including the dependency on h in these parameters. A similar consideration applies to the definition of the mass variance. The commonly used quantity σ_8 is defined within a radius $R = 8\,h^{-1}\,\mathrm{Mpc}$, and therefore inherits an explicit dependence on h. To avoid this, we instead define the variance within a fixed physical scale $R = 12\,\mathrm{Mpc}$, denoted as σ_{12} (see Sect. 2.2.2), which provides a comparable measure of the clustering amplitude while preserving the degeneracy with respect to h. The pivot scale k_p is also expressed in Mpc^{-1}.

Within this framework, cosmological parameters can be divided into two classes:

- Shape parameters $\mathbf{\Theta}_s = (\omega_\gamma, \omega_\mathrm{b}, \omega_\mathrm{cdm}, n_s, \dots)$, which determine the shape of the primordial power spectrum and the transfer function. Once these parameters are fixed, the shape of the linear matter power spectrum is fully specified.

- Evolution parameters $\mathbf{\Theta}_e = (A_s, \omega_k, \omega_\mathrm{DE}, w(a), h, \dots)$, which only affect the overall amplitude of the power spectrum at a given redshift.

In the presence of massive neutrinos, the growth of perturbations becomes scale dependent already at linear order, and this separation no longer strictly holds. In this work, we restrict ourselves to models with $\omega_\nu = 0$, and defer the extension to massive neutrinos to future work.

Since evolution parameters only modify the amplitude of the power spectrum, their effect is degenerate. For fixed values of the shape parameters, all cosmological models that share the same clustering amplitude yield identical linear power spectra, even for different choices of evolution parameters. This property is captured by the evolution-mapping relation (Sánchez et al. 2022), which forms the basis of the methodology described below:

$$P(k|z, \mathbf{\Theta}_s, \mathbf{\Theta}_e) = P\Big(k|z, \mathbf{\Theta}_s, \sigma_{12}(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)\Big)\,. \tag{3.6}$$

At linear order, the time evolution of the power spectrum can therefore be described as a relabelling of the redshift corresponding to a given value of σ_{12}. In other words, different cosmological models characterised by the same shape parameters but different evolution parameters can be mapped onto each other by matching their clustering amplitude. This degeneracy is exact only when quantities are expressed in physical units. In linear theory, the variance of the density field evolves proportionally to the linear growth factor, $\sigma_{12}(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e) \propto D(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)$. As a result, the clustering amplitude can be parametrised in terms of $D(z)$, allowing one to rewrite the evolution-mapping relation as

$$P(k|z, \mathbf{\Theta}_s, \mathbf{\Theta}_e) = P\Big(k|z, \mathbf{\Theta}_s, D(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)\Big)\,. \tag{3.7}$$

This formulation makes explicit that, at linear order, the dependence on the evolution parameters enters through the growth factor. Let us consider then two power spectra with identical shape parameters but different evolution parameters and redshifts, namely $P(k, z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)$ and $P(k, z', \mathbf{\Theta}_s, \mathbf{\Theta}'_e)$. Denoting by $D(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)$ and $D(z', \mathbf{\Theta}_s, \mathbf{\Theta}'_e)$ the corresponding linear growth factors, the two spectra are related by a simple rescaling of the amplitude:

$$P(k, z, \mathbf{\Theta}_s, \mathbf{\Theta}_e) = P(k, z', \mathbf{\Theta}_s, \mathbf{\Theta}'_e) \times \left[\frac{D(z, \mathbf{\Theta}_s, \mathbf{\Theta}_e)}{D(z', \mathbf{\Theta}_s, \mathbf{\Theta}'_e)}\right]^2 \equiv P(k, z', \mathbf{\Theta}_s, \mathbf{\Theta}'_e)\ \times D_{\rm corr}\,. \tag{3.8}$$

An explicit example of this mapping is shown in Fig. 3.3. We compare a reference power spectrum computed at redshift $z = 1$ with a second spectrum evaluated at $z = 0$ for a different value of the evolution parameter h, while keeping all shape parameters fixed. The latter is then rescaled by the square of the ratio of the corresponding growth factors $D_{\rm corr}$. The rescaled spectrum accurately reproduces the reference one over a wide range of scales, demonstrating the effectiveness of the evolution-mapping relation. To quantify the agreement, we define the relative residuals as

$$\Delta(k) = \frac{P_{\rm map}(k) - P_{\rm ref}(k)}{P_{\rm ref}(k)} \times 100\,, \tag{3.9}$$

where $P_{\rm ref}$ denotes the reference spectrum and $P_{\rm map}$ the rescaled one. As shown in the lower panel of Fig. 3.3, the residuals remain at the sub-percent level over most of the k range considered.

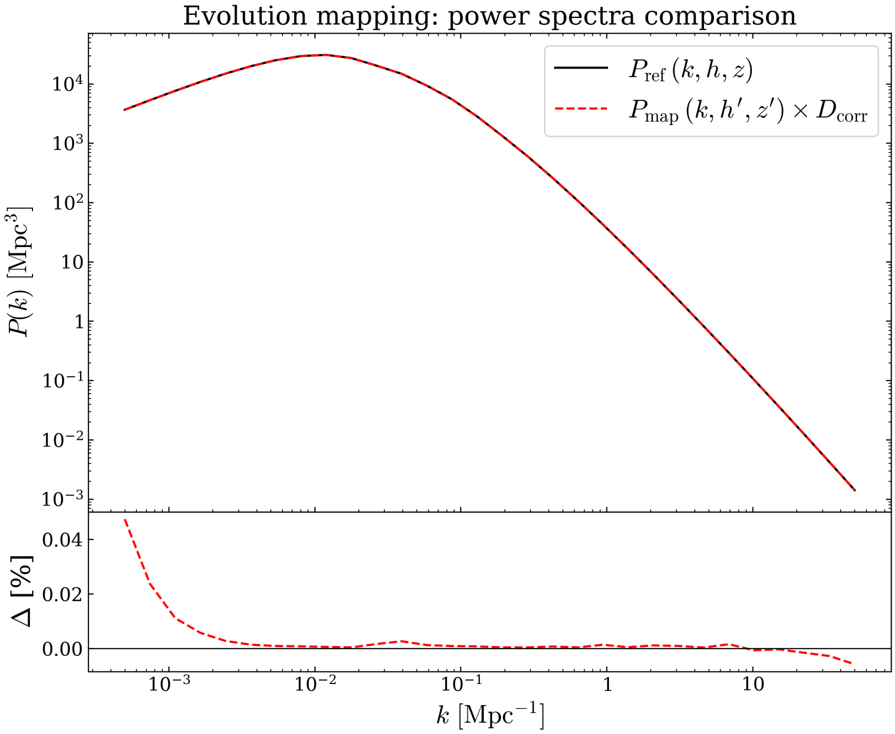

Figure 3.3: Illustration of the evolution-mapping relation. The black solid line shows the reference linear matter power spectrum computed with `CAMB` at $z = 1$ within a ΛCDM cosmology, with all parameters fixed to the Planck collaboration (2020) best-fit values. The red dashed line is obtained by rescaling a spectrum computed at $z = 0$ for a different value of the dimensionless Hubble parameter ($h = 0.65$), while keeping all other cosmological parameters fixed to the same reference values. The rescaling is performed using the square of the ratio of the corresponding linear growth factors $D_{\rm corr}$. The lower panel shows the relative residuals, demonstrating sub-percent agreement over a wide range of scales. The residual increase at the largest scales reflects the fact that the exact linear evolution computed by `CAMB` is not perfectly described by a scale-independent growth factor, leading to a weak scale dependence of the growth on scales approaching the horizon.

Deviations are observed on the largest scales. These arise because the exact linear evolution computed by `CAMB` is not perfectly described by a scale-independent growth factor over the full range of wavenumbers. On scales approaching the horizon, relativistic effects introduce a weak scale dependence in the growth of matter perturbations, which is not captured by the evolution-mapping approximation based on a fully scale-independent growth factor. As a result, the rescaled spectrum slightly overestimates the exact `CAMB` prediction at the lowest wavenumbers. At the highest wavenumbers considered, a much smaller deviation is also visible. This reflects the residual imprint of radiation on the evolution of small-scale perturbations: modes that entered the horizon during the radiation-dominated era experienced a slightly suppressed growth relative to larger-scale modes, an effect that is not fully captured by the scale-independent evolution-mapping approximation.

Despite these limitations, the agreement remains excellent across the entire range. The residual deviations never exceed the 0.05%, which is more than an order of magnitude smaller than the precision typically achieved by current clustering measurements and well below the percent-level accuracy commonly adopted as a benchmark for precision cosmology (see Sect. 3.2.2). It is also worth noting that such small deviations are confined to very extreme scales, well outside the range typically probed by LSS analyses. These results show that the errors introduced by evolution mapping are negligible for practical LSS applications, while retaining a substantial computational advantage.

A key consequence of the evolution mapping is a substantial reduction in the effective dimensionality of the problem. Once the power spectrum has been computed for a given set of shape parameters, a large family of cosmological models with different evolution parameters and redshifts can be obtained through a simple rescaling of the amplitude, proportional to σ_{12}^2 or, equivalently, D^2. As a result, when constructing an emulator for the matter power spectrum, the dependence on the evolution parameters does not need to be explicitly learned. In practice, this implies that only the parameters controlling the shape of the transfer function must be sampled during the training phase. In our implementation, these correspond to the physical baryon and CDM densities, $\omega_{\rm b}$ and $\omega_{\rm cdm}$, which therefore constitute the only computational bottleneck (see Sect. 4.2).

3.3.1 Parameter rescaling and curvature correction

The evolution-mapping relation can be complemented by additional analytical treatments that account for the dependence of the linear power spectrum on specific cosmological parameters, as well as by a correction factor that captures the modification of its shape induced by the geometric curvature of spacetime.

The scalar amplitude A_s and the spectral index n_s can be recovered exactly through a rescaling of the power spectrum. At linear order, the primordial power spectrum is parametrised as $P_{\rm prim}(k) = A_s(k/k_p)^{n_s-1}$, so that variations in A_s and n_s correspond to a multiplicative rescaling and a scale-dependent tilt, respectively. As a consequence, two linear matter power spectra that differ only in these parameters are related by (Aricò et al. 2022)

$$P(k, A_s', n_s') = \frac{A_s'}{A_s}\left(\frac{k}{k_p}\right)^{n_s'-n_s} P(k, A_s, n_s)\,. \tag{3.10}$$

This relation can be verified explicitly by comparing power spectra computed for different values of A_s and n_s and applying the corresponding rescaling, as shown in Fig. 3.4. This relation allows one to account for the dependence on A_s and n_s without explicitly including them in the training of the emulator. When combined with the evolution-mapping relation, which absorbs the dependence on the remaining evolution parameters into the growth factor, this further reduces the dimensionality of the parameter space that must be explored numerically.

Finally, the presence of spatial curvature modifies the form of the primordial power spectrum and must be explicitly taken into account. In non-flat geometries, the relevant wavenumber is no longer the comoving Fourier mode k, but the

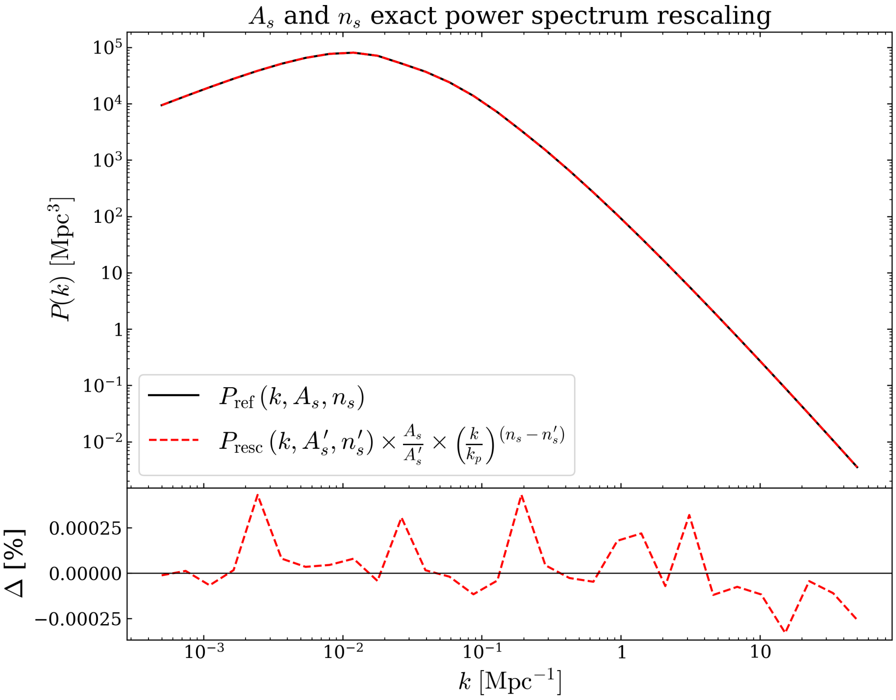

Figure 3.4: Exact rescaling of the linear matter power spectrum with respect to the parameters A_s and n_s. The black solid line shows the reference spectrum computed with `CAMB` at $z = 0$ within a ΛCDM cosmology, with all parameters fixed to the Planck collaboration (2020) best-fit values. The red dashed line is obtained by rescaling a spectrum computed with the same setup but adopting $A_s = 1.9 \times 10^{-9}$ and $n_s = 0.9$. The rescaling is performed using Eq. (3.10). The lower panel shows the relative residuals.

curvature-dependent quantity

$$q = \sqrt{k^2 + K}\,, \tag{3.11}$$

where $K = -H_0^2\,\Omega_k$. The primordial power spectrum can then be written as (Mitra et al. 2019)

$$P_{\rm prim}(q) \propto \frac{(q^2 - 4K)^2}{q(q^2 - K)} \left(\frac{q}{k_p}\right)^{n_s - 1} . \tag{3.12}$$

Since $q \to k$ at large k, the impact of spatial curvature is negligible on small scales, while it becomes relevant on large scales. Within the evolution-mapping framework, which is naturally formulated for flat geometries, the impact of spatial curvature can be effectively incorporated through a multiplicative correction factor. However, Eq. (3.12) only describes the primordial spectrum, while the full matter power spectrum also depends on the transfer function and on the growth of perturbations, both of which are modified in curved geometries in a non-trivial way and do not admit a simple analytical description. For this reason, rather than attempting a fully analytical derivation, we introduce a phenomenological correction inspired by Eq. (3.12) and calibrated against Boltzmann solver results. The adopted correction factor is

$$\mathrm{fact}(k) = \frac{(q^2 - 4K)^2}{(q^2 - K)\,k^2}\,, \tag{3.13}$$

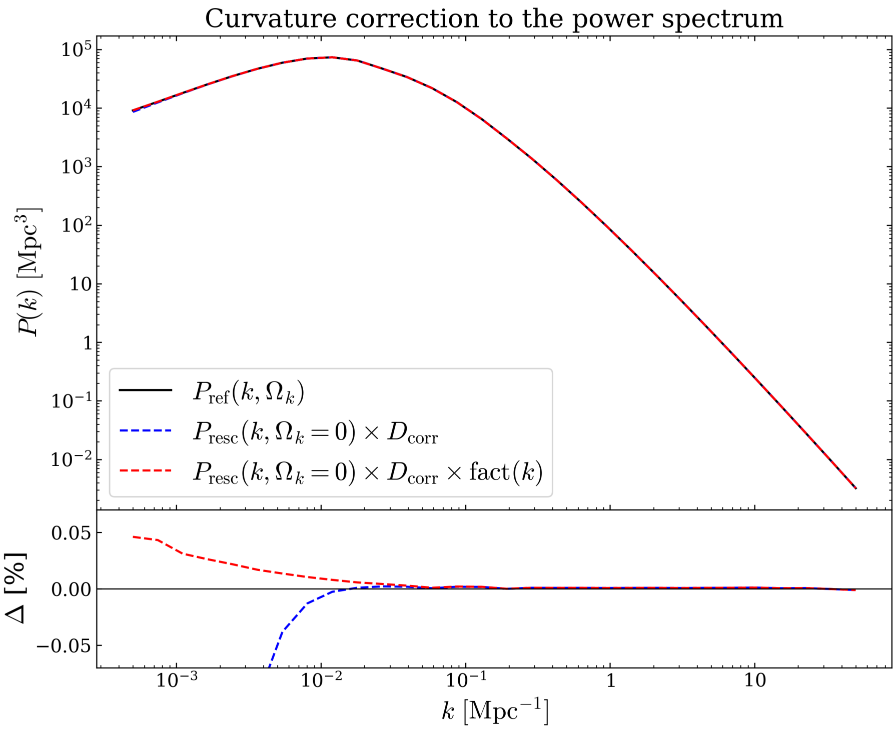

Figure 3.5: Curvature correction to the linear matter power spectrum. The black solid line shows the reference spectrum computed with `CAMB` at $z = 0$ in a $k\Lambda$CDM cosmology with parameters fixed to the Planck collaboration (2020) best-fit values. The blue dashed line corresponds to a spectrum computed with the same setup but with $\Omega_k = 0$, rescaled by the growth-factor correction $D_{\rm corr}$, which depends on Ω_k as an evolution parameter. The red dashed line is obtained by further applying the curvature correction factor fact(k) (Eq. 3.13). The lower panel shows the residuals with respect to the reference spectrum. The vertical range is restricted to highlight the improved agreement on large scales, where curvature effects are significant. Without this correction, deviations reach around −6%.

so that the curved-space power spectrum is approximated as

$$P_{\rm curved}(k) \simeq \mathrm{fact}(k)\, P_{\rm flat}(k)\,. \tag{3.14}$$

This factor accounts for the modification of the large-scale behaviour of the power spectrum induced by spatial curvature, ensuring a more consistent extension of the method to non-flat cosmologies. In Fig. 3.5 we show an explicit example, where a spectrum obtained in a flat geometry is first rescaled using the growth factor and then further corrected by applying the curvature-dependent factor in Eq. (3.13). The comparison with the reference curved model highlights the importance of the curvature correction, especially on large scales.

3.3.2 Limits of the method

Despite its effectiveness, the evolution-mapping approach is subject to a number of limitations that ultimately arise from the assumptions underlying linear perturbation theory. In particular, the method relies on the separability of the evolution of matter perturbations in scale and time, such that $\delta(k, z) \propto D(z)$, allowing the entire redshift dependence of the linear matter power spectrum to be described through a scale-independent growth factor.

As discussed in Sect. 3.3, this approximation is not exact over the full range of scales. Small deviations arise at the largest and smallest wavenumbers considered, approximately corresponding to k smaller about $10^{-3}\,\mathrm{Mpc}^{-1}$ and greater about $10\,\mathrm{Mpc}^{-1}$. These departures are associated with physical effects that are not fully captured by a simple scale-independent growth rescaling, including relativistic corrections on scales approaching the horizon and the residual impact of radiation on very small scales. Nevertheless, the resulting discrepancies remain below 0.05%, thus an *ad hoc* correction is not required as they are negligible for practical applications.

A further limitation is encountered in cosmological models with a time-dependent dark energy equation of state. In these scenarios, the time-varying equation of state alters not only the background expansion history and the Hubble rate, but also the perturbation dynamics. Specifically, the matter perturbation evolution, Eq. (2.37), includes additional contributions from the dark energy density contrast, $\delta_{\mathrm{DE}}(k, z)$, which opposes gravitational collapse. While strongly suppressed on sub-horizon scales with a negligible impact over most of the wavenumber range, these perturbations contribute significantly on near- and super-horizon scales. This introduces an explicit scale dependence in the growth factor, i.e. $D(z) \to D(k, z)$. As a result, a scale-independent mapping fails to perfectly reproduce the exact linear evolution computed by `CAMB`, introducing small deviations on scales approaching the horizon. Figure 3.6 illustrates this behaviour by comparing the exact linear matter power spectrum computed with `CAMB` in a w_0w_aCDM cosmology with the spectrum obtained through the evolution-mapping procedure. While the agreement remains excellent over most of the wavenumber range, residuals reach the 0.05% level for wavelenghts smaller than $10^{-3}\,\mathrm{Mpc}^{-1}$. Throughout this work, rather than introducing an empirical correction for these deviations, we adopt a conservative approach and discard the range of the emulator to scales where the required accuracy is not guaranteed. Since the affected scales lie well beyond the region probed by current LSS surveys, we emphasise that this limitation has negligible practical impact on the analyses presented in this Thesis (see Chapter 5).

Another limitation arises in the presence of massive neutrinos. Due to their large thermal velocities, massive neutrinos free-stream out of overdense regions below a characteristic scale, suppressing the growth of structure in a scale-dependent way. As a consequence, the growth of total matter perturbations is no longer scale independent even at linear order, and the neutrino density parameter ω_ν cannot be cleanly classified as either a shape or an evolution parameter. Although cosmologies with massive neutrinos are not considered in this Thesis, several strategies have been proposed to incorporate their effects within the evolution-mapping framework. For instance, Pezzotta et al. (2025) introduced an emulator for the suppression of clustering induced by massive neutrinos using a parametrisation based on A_s and σ_{12}. An alternative approach consists of replacing the standard growth factor with a scale-dependent quantity, $D(k, z)$, explicitly accounting for neutrino free-streaming. While naturally compatible with the architecture developed in this work, the latter

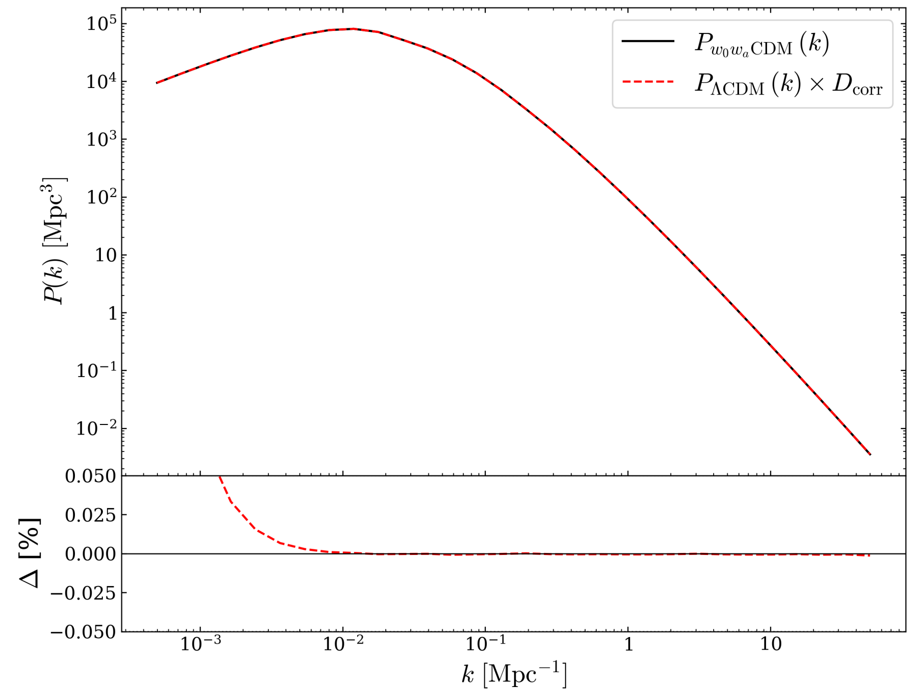

Figure 3.6: Limits of the evolution-mapping approximation in a $w_0 w_a$CDM cosmology. The black solid line shows the reference linear matter power spectrum computed with `CAMB` within a $w_0 w_a$CDM model, while the red dashed line corresponds to the spectrum reconstructed through the evolution-mapping procedure from a ΛCDM model. The lower panel shows the relative residuals. The vertical range is restricted to $\pm 0.05\%$ in order to highlight the wavenumber interval over which the target accuracy is achieved.

solution requires the numerical evaluation of significantly more complex evolution equations (see Blas et al. 2014), thereby reducing the computational efficiency of the method. A detailed investigation of these extensions is therefore left for future work.

Similarly, modified-gravity models generally induce a scale-dependent growth of matter perturbations (e.g. Hu & Sawicki 2007), breaking the fundamental assumption on which the evolution-mapping framework is built. The inclusion of these scenarios would therefore require dedicated corrections and is left to future developments.

Finally, although the evolution-mapping formalism cannot be directly extended to the nonlinear matter power spectrum, a promising strategy exists to incorporate nonlinear gravitational evolution. As shown by Sánchez et al. (2025), the nonlinear correction retains memory of the growth history of cosmic structures and can therefore be modelled separately from the linear power spectrum. This suggests a natural extension of the present framework, in which a second NN emulator is trained to predict the nonlinear correction as a function of the growth history. Such an approach would preserve the efficiency of the evolution-mapping formalism while extending its applicability to the nonlinear regime.

In summary, we discussed how the methodology presented in this Thesis, despite its intrinsic limitations, simultaneously accommodates a wide range of potential extensions. While our primary goal remains the accurate emulation of the linear matter power spectrum for the most relevant cosmological scenarios, we stress that the modular structure of the framework facilitates the inclusion of dedicated corrections, providing a natural path towards extensions for alternative scenarios characterised by a higher degree of complexity.

Chapter 4

A new power spectrum emulator

In Chapter 2 and in Chapter 3 we introduced the physical ingredients that determine the evolution of the matter power spectrum, together with the ML techniques and the evolution-mapping strategy adopted to reduce the dimensionality of the problem. We now describe how these concepts are combined to construct a new cosmological emulator capable of reproducing the linear matter power spectrum over a wide range of cosmological models and redshifts.

The implementation presented in this work has been developed within the framework of `CosmoBolognaLib` (`CBL`), a publicly available set of `C++` numerical libraries designed for cosmological calculations and LSS analyses (Marulli et al. 2015). The libraries provide a common numerical environment for handling cosmological catalogues, modelling clustering statistics, and performing Bayesian parameter inference. In particular, the `CBL` already interfaces with external Boltzmann solvers such as `CAMB` and `CLASS`, enabling the computation of theoretical matter power spectra and derived observables. However, as discussed in the previous chapters, the repeated evaluation of these theoretical predictions within Bayesian inference pipelines represents a major computational bottleneck, especially with large parameter spaces. To overcome this limitation, previous work by Nicolosi (2022) introduced a general emulator framework within the `CBL`, based on NN architectures implemented through `CosmoPower`. That implementation demonstrated the possibility of accurately reproducing cosmological correlation functions while achieving speed-up of the analyses pipeline of the interesting quantities. Building upon this framework, the goal of the present work is to develop a new emulator specifically designed for the linear matter power spectrum, extending its applicability to non-standard cosmological scenarios and multiple redshifts through the evolution-mapping approach described above.

In this chapter, we first provide an overview of the `CBL` structure and discuss the emulator framework already available within the libraries. We then present the implementation strategy adopted for the new emulator, focusing on the model used to reconstruct the matter power spectrum from the emulated quantities. Particular attention is devoted to the numerical treatment of the growth factor, whose computation has been modified in order to achieve a non-normalized and highly accurate evolution across different cosmological models. Finally, we describe the construction of the training and validation data sets, together with the training procedure employed to optimize the NN and assess its accuracy over the explored parameter space.

4.1 CosmoBolognaLib

CBL is a *free software* set of modular C++ libraries developed for cosmological calculations and LSS analyses (Marulli et al. 2015). The libraries provide a unified computational framework for theoretical modelling, statistical inference, and the analysis of cosmological data sets. Python bindings are also available, allowing the main functionalities of the libraries to be accessed within Python environments.

The CBL are organised into several specialised modules dedicated to different aspects of cosmological analyses. In particular, the Cosmology module provides routines for the computation of cosmological distances, background quantities, growth functions, matter power spectra, correlation functions, halo statistics, and RSD models. Linear and nonlinear matter power spectra can be computed through interfaces to external Boltzmann solvers, including CAMB and CLASS. Additional modules are devoted to the management of observational and simulated catalogues, the measurement and modelling of clustering observables, and Bayesian statistical analyses based on likelihood sampling techniques. The current implementation of the libraries also includes emulator-based infrastructures designed to accelerate the evaluation of cosmological observables. In particular, the Emulator framework introduced by Nicolosi (2022) provides a general interface for NN emulators based on the CosmoPower architecture. The emulator implementation adopted in this work has been developed within this framework.

In the present analysis, the CBL were used both to generate the data sets employed for the emulator training and to integrate the trained model into the cosmological inference pipeline. The training and validation data sets were produced by repeatedly evaluating the linear matter power spectrum with CAMB over the cosmological parameter space explored in this work (see Sect. 4.2). The NN was trained externally through the CosmoPower environment. The resulting trained weights were then imported into the CBL and used through the Emulator class, allowing the emulator to replace the direct Boltzmann-solver evaluations within the cosmological pipeline. In this way, the computation of the linear matter power spectrum could be performed through the emulator while preserving compatibility with the existing modelling and statistical infrastructures of the libraries. A dedicated implementation of the evolution-mapping strategy described in Sect. 3.3 was introduced within the routines responsible for evaluating the cosmological evolution of the power spectrum. The final implementation allows the emulator-based power spectrum to be directly employed in theoretical modelling and MCMC parameter inference analyses performed within the CBL framework.

4.1.1 The Emulator class

The emulator infrastructure implemented within the CBL is designed to provide a general interface for evaluating NN emulators directly inside the cosmological pipeline. The implementation adopted in this work is based on the emulator framework introduced by Nicolosi (2022), which allows pre-trained NN to be integrated into the CBL environment without relying on external ML libraries during the inference stage. The Emulator class implements the forward propagation of a fully-connected feed-forward NN trained externally through a generic ML framework. In the present work, the networks adopted for the emulator construction were trained using the CosmoPower framework. The network parameters are stored after the training procedure and subsequently imported into the CBL. In particular, the class loads the weights and bias vectors of each layer, the parameters defining the activation functions, the normalization coefficients adopted for the input and output quantities, and the discrete set of modes on which the target function is evaluated.

Given a set of parameters as input, the emulator evaluates the corresponding cosmological observable through the forward propagation of the NN. The input quantities are propagated through the successive network layers, where a sequence of linear transformations and nonlinear activations is applied according to the activation function defined in Eq. (3.5). The network finally returns the emulated quantity on the grid adopted during the training procedure. Whenever the emulated function is required at user-defined data points (e.g. an arbitrary set of wavevectors), the network first predicts the function on the grid adopted during training and subsequently interpolated onto the requested points.

The implementation of the forward propagation directly in C++ allows the emulator to be employed within the native CBL cosmological pipeline without external dependencies. This approach enables fast evaluations of cosmological observables while preserving compatibility with the modelling and statistical infrastructures of the libraries.

4.2 Implementation strategy

The main goal of this work is to construct a cosmological emulator capable of accurately reproducing the linear matter power spectrum computed with CAMB over a broad range of cosmological parameters, redshifts, and cosmological scenarios. In particular, the implementation developed in this work has been tested on the ΛCDM, $k\Lambda$CDM, w_0CDM, w_0w_aCDM, kw_aCDM, and kw_0w_aCDM cosmologies. In the following, however, we focus primarily on the ΛCDM, $k\Lambda$CDM, and w_0w_aCDM models (see Sect. 1.3.2). These three cases encompass the standard cosmological scenario together with two of its most widely studied extensions. The ΛCDM model, which currently represents the standard model of cosmology, serves as the starting point of the present analysis and as the reference framework against which all extended models are evaluated. The $k\Lambda$CDM model relaxes the assumption of spatial flatness, allowing the impact of a non-zero curvature parameter on the emulator performance to be investigated. Finally, the w_0w_aCDM model replaces the cosmological constant with a time-dependent dark energy equation of state. This model is relevant in light of recent DESI constraints, which show a mild preference for dynamical dark energy models when combined with CMB and supernova data (Adame et al. 2025b).

As discussed in Sect. 2.2.1, in linear regime the matter power spectrum can be written as

$$P(k,a) = A_s \left(\frac{k}{k_p}\right)^{n_s-1} T^2(k) D^2(a) \ , \tag{4.1}$$

where A_s sets the primordial amplitude of scalar perturbations, n_s determines the primordial spectral tilt, $T(k)$ encodes the scale dependence generated by early-Universe physical processes, and $D(a)$ describes the late-time growth of matter perturbations. The transfer function contains the physical imprint of processes occurring before matter domination. In particular, perturbation modes entering the horizon during the radiation-dominated era grow less efficiently with respect to modes entering during matter domination. This effect generates the characteristic scale dependence of the matter power spectrum and is encoded in $T(k)$. At late times, instead, linear perturbations evolve with an approximately scale-independent growth, described by the growth factor $D(a)$, which therefore controls the overall evolution of the power-spectrum amplitude.

As shown in Sect. 3.3, this decomposition naturally separates the cosmological dependence of the power spectrum into two distinct contributions: a scale-dependent component, mainly determined by the shape parameters, and an evolution component associated with the background expansion history. In particular, once the shape parameters are fixed, the power spectrum corresponding to different cosmological evolutions can be reconstructed through suitable rescalings involving the growth factor. This property provides the basis of the evolution-mapping strategy adopted in this work.

A crucial aspect of the implementation concerns the computation of the growth factor. As discussed in Sect. 2.1.5, the evolution of linear matter perturbations is governed by Eq. (2.37), whose solution depends explicitly on the Hubble expansion rate $H(a)$. Since the latter changes with the cosmological model, different cosmologies are characterised by different growth histories. In particular, models including spatial curvature or dynamical dark energy modify the evolution of the Hubble friction term entering the perturbation equation (see Sect. 1.3.2), leading to a different evolution of the linear growth factor. For this reason, the growth-factor computation within the `CBL` framework was modified in order to provide a numerically accurate non-normalised solution for all cosmological models explored in this work. The growth factor is computed directly through a different numerical integration of Eq. (2.37) for each cosmological model. This choice is required because the evolution mapping relies on the absolute amplitude ratio between different cosmological evolutions. A normalised growth factor would remove the amplitude information necessary to correctly reconstruct the redshift evolution of the power spectrum.

Furthermore, the dependence on the primordial parameters A_s and n_s can be treated analytically through exact rescaling relations. As discussed in Sect. 3.3.1, the contribution of the primordial spectrum can be factorised analytically, allowing one to reconstruct the dependence on A_s and n_s without including these quantities among the parameters learned by the emulator. Combined with the evolution mapping, this considerably reduces the effective dimensionality of the parameter space.

Putting these ingredients together, it is possible to construct a mapping between power spectra corresponding to cosmological models sharing the same shape parameters $\omega_{\rm b}$ and $\omega_{\rm cdm}$. Let us consider a reference ΛCDM cosmology characterised by the parameters $A_s^{\rm ref}$, $n_s^{\rm ref}$, $h^{\rm ref}$, $\omega_{\rm b}$, $\omega_{\rm cdm}$, with corresponding linear matter power spectrum $P^{\rm ref}(k, z^{\rm ref}) \equiv P(k, z^{\rm ref}, A_s^{\rm ref}, n_s^{\rm ref}, h^{\rm ref}, \omega_{\rm b}, \omega_{\rm cdm})$. We then consider a generic target cosmology described by the parameters A_s, n_s, h, $\omega_{\rm b}$, $\omega_{\rm cdm}$, ω_k, w_0, $w_a, \ldots,$ with power spectrum evaluated at redshift z. The corresponding mapping can be written as

$$P(k, z) = \frac{A_s}{A_s^{\rm ref}} \left(\frac{k}{k_p}\right)^{n_s - n_s^{\rm ref}} \left[\frac{D^{\rm model}(z)}{D^{\rm ref}(z^{\rm ref})}\right]^2 \mathrm{fact}(k)\, P^{\rm ref}(k, z^{\rm ref}) \;, \tag{4.2}$$

where $D^{\rm model}(z)$ is the growth factor associated with the target cosmological model, while $D^{\rm ref}(z^{\rm ref})$ corresponds to the reference cosmology (used for the emulator training, see Sect. 4.2.1). The factor $\mathrm{fact}(k)$ represents the curvature correction discussed in Sect. 3.3.1 and accounts for residual scale-dependent curvature effects that are not fully captured by the growth-factor rescaling alone. In the ΛCDM and w_0w_aCDM scenarios this contribution is null, while in non-flat cosmologies a residual curvature-dependent correction to the power-spectrum shape must be included.

The parameters $\omega_{\rm b}$ and $\omega_{\rm cdm}$ determine several characteristic scales of the matter distribution, including the horizon scale at matter–radiation equality and the BAO features. Consequently, they produce a non-trivial modification of the shape of the power spectrum that cannot be accurately reproduced through simple analytical rescalings.

Equation (4.2) shows that most cosmological dependencies can be reconstructed through analytical rescalings and fast numerical computations. The only remaining quantity that still requires the evaluation of a Boltzmann solver is the reference power spectrum $P^{\rm ref}(k, z^{\rm ref})$. This computation represents the main computational bottleneck in repeated posterior evaluations. The strategy adopted in this work therefore consists in constructing an emulator for the reference quantity $P^{\rm ref}(k, z^{\rm ref}, \omega_{\rm b}, \omega_{\rm cdm})$, leaving only the dependence on $\omega_{\rm b}$ and $\omega_{\rm cdm}$ to be learned by the NN. Denoting the emulator prediction by $P_{\rm emu}(k, \omega_{\rm b}, \omega_{\rm cdm})$ and using $z^{\rm ref} = 0$, Eq. (4.2) becomes

$$P(k, z) = \frac{A_s}{A_s^{\rm ref}} \left(\frac{k}{k_p}\right)^{n_s - n_s^{\rm ref}} \left[\frac{D^{\rm model}(z)}{D^{\rm ref}(z=0)}\right]^2 \mathrm{fact}(k)\, P_{\rm emu}(k, z=0, \omega_{\rm b}, \omega_{\rm cdm}) \;. \tag{4.3}$$

Equation (4.3) summarises the implementation strategy adopted in this work. The primordial rescaling terms account exactly for the dependence on A_s and n_s, the growth-factor ratio reconstructs the cosmological evolution of the perturbation amplitude, and the emulator reproduces the residual scale dependence associated with $\omega_{\rm b}$ and $\omega_{\rm cdm}$. This approach mitigates the curse-of-dimensionality problem, considerably simplifying the training procedure.

Within the `CBL` framework, all the ingredients entering Eq. (4.3) are internally combined within a single function call. In particular, the analytical primordial rescalings, the numerical computation of the growth-factor ratio, the curvature correction term, and the NN evaluation of the reference power spectrum are all performed automatically inside the pipeline. From the user perspective, the target linear matter power spectrum can therefore be obtained through a single function evaluation, without explicitly handling the individual rescaling contributions. This implementation strategy provides a computationally efficient and user-friendly interface for cosmological analyses. In particular, the emulator can be seamlessly integrated within existing `CBL` likelihood pipelines, allowing one to replace the direct `CAMB` evaluation without modifying the structure of the statistical analysis.

4.2.1 Data set creation

The construction of the emulator requires a training data set composed of input–output pairs. In the present case, the emulator is designed to reproduce the linear matter power spectrum $P(k, \omega_{\mathrm{b}}, \omega_{\mathrm{cdm}})$ as a function of the wavenumber k for different values of the physical baryon and CDM densities ω_{b} and ω_{cdm}. As discussed in Sect. 3.3, Eq. (4.3) allows most cosmological dependencies of the linear matter power spectrum to be reconstructed through analytical rescalings and growth-factor evolution. Consequently, only the residual shape dependence associated with ω_{b} and ω_{cdm} must be learned by the NN emulator. All remaining cosmological parameters can therefore be fixed to arbitrary reference values, provided that the corresponding rescaling factors entering Eq. (4.3) are consistently reconstructed. In particular, the primordial parameters A_s and n_s are fixed to reference values, while the growth factors associated with both the reference and target cosmologies are computed numerically for each cosmological model. Since the growth-factor evolution depends explicitly on the background cosmology, it still retains an implicit dependence on ω_{b}, ω_{cdm} and other cosmological parameters and therefore cannot be treated as a fixed contribution.

The reference cosmology adopted for the emulator training corresponds to the Planck best-fit ΛCDM model (Planck collaboration 2020). The cosmological parameters used for the data set generation are summarised in Table 4.1. Although the evolution-mapping formalism needs to be expressed in terms of the physical density parameters $\omega_i \equiv \Omega_i h^2$, the `CBL` interface to `CAMB` requires the cosmological model to be specified through the corresponding density parameters Ω_i. Consequently, the parameter values adopted during the generation of the training and validation data sets are specified consistently in the same convention.

The sampling of the wavenumber range was chosen to provide a higher resolution in the region containing the BAO features. Since the oscillatory structure of the power spectrum in this regime is more difficult for the NN to reproduce accurately, a denser sampling was adopted in the corresponding interval. More precisely:

- the interval $1.5 \times 10^{-4}\,\mathrm{Mpc}^{-1} \leq k < 10^{-2}\,\mathrm{Mpc}^{-1}$ was sampled with 100 points;
- the region $10^{-2}\,\mathrm{Mpc}^{-1} \leq k < 1\,\mathrm{Mpc}^{-1}$, containing the BAO features, was sampled with 200 points;
- the interval $1\,\mathrm{Mpc}^{-1} \leq k \leq 75\,\mathrm{Mpc}^{-1}$ was sampled with an additional 100 points.

Parameter	Value / range
z	0
$k\,[\mathrm{Mpc}^{-1}]$	$[1.5 \times 10^{-4},\ 75]$
Ω_b	$[0.03,\ 0.06]$
Ω_cdm	$[0.07,\ 0.70]$
h	0.6736
A_s	2.1×10^{-9}
n_s	0.9649
τ	0.0544
Ω_ν	0
Ω_γ	9.26535×10^{-5}
Ω_DE	$1 - \Omega_\mathrm{b} - \Omega_\mathrm{cdm} - \Omega_\gamma$
Ω_k	0
w_0	-1
w_a	0

Table 4.1: Cosmological parameters adopted for the reference cosmology used in the emulator training. The emulator is trained on the dependence of the linear matter power spectrum on ω_b and ω_cdm, while all remaining parameters are fixed to the Planck best-fit ΛCDM cosmology (Planck collaboration 2020). The wavenumber range is expressed in physical units.

This choice leads to a total of 400 sampled wavenumbers and therefore to an emulator output layer composed of 400 neurons, each corresponding to one evaluation of the power spectrum at a given k.

The parameter ranges for ω_b and ω_cdm were chosen in order to cover a broad region of the physically relevant parameter space, allowing the emulator to be applied to a wide variety of cosmological scenarios. The reduction of the effective parameter-space dimensionality obtained through the evolution mapping makes it possible to explore relatively large parameter ranges. The adopted wavenumber and cosmological parameter ranges are inspired by the ranges considered in `BACCO` (Aricò et al. 2022).

The training and testing data sets were generated by sampling the $(\omega_\mathrm{b}, \omega_\mathrm{cdm})$ parameter space. After several preliminary tests balancing emulator accuracy and computational cost, we adopted a data set composed of 80 000 training samples and 20 000 testing samples. The sampling procedure was designed to maximise the coverage uniformity of the parameter space through the maxmin space filling strategy (Wessing 2015). Operationally, an initial set of random points was generated from uniform distributions within the parameter ranges, followed by an iterative optimisation procedure aimed at maximising the mutual distances between neighbouring samples. This strategy improves the homogeneity of the parameter-space coverage and reduces the presence of undersampled regions compared to a standard Latin Hypercube Sampling design (Virtanen et al. 2020).

Once the wavenumber sampling and the pairs of cosmological parameters (ω_b, ω_cdm) were generated, the corresponding linear matter power spectra were computed with CAMB, since the goal of the emulator is to reproduce its linear predictions. The computation of the training and testing data sets, as well as the NN training and the statistical analyses, was performed on the Open Physics Hub high-performance computing cluster at the University of Bologna. The infrastructure consists of heterogeneous computing nodes organised in the BladeRunner and Matrix partitions, managed through the SLURM workload manager (OPH 2020). While the BladeRunner partition is optimised for low-parallel workloads, the Matrix partition is designed for highly parallel computations. In the present case, the generation of the power-spectrum data sets was performed on a single Matrix node employing 112 virtual CPUs running in parallel at a clock frequency of 2.2 GHz. The full data set generation required more than seven hours of computation time.

Before the training phase, the power spectra were preprocessed by taking their logarithm. This transformation significantly reduces the dynamic range of the training targets and improves the numerical stability of the optimisation procedure, allowing the NN to learn more accurately also the small-scale features where the power spectrum amplitude becomes smaller. Consequently, the NN is trained to predict $\log P(k, \omega_{\rm b}, \omega_{\rm cdm})$ rather than the power spectrum itself. Within the final CBL implementation, the emulator output is automatically exponentiated after all rescalings have been correctly applied, reconstructing the physical linear matter power spectrum.

4.2.2 Training and validation

The NN emulator was trained using the learning pipeline implemented within the CosmoPower framework. The training procedure optimises the weights connecting the neurons of consecutive layers in order to reproduce the mapping between the cosmological input parameters $(\omega_{\rm b}, \omega_{\rm cdm})$ and the corresponding linear matter power spectrum. Once trained, the resulting set of weights is imported within the CBL framework, as discussed in Sect. 4.1.1.

The adopted NN architecture consists of four hidden layers with 128 neurons each. The architecture was tuned in order to achieve a balance between model complexity and generalisation capability, since, as discussed in Sect. 3.1.3, excessively complex architectures may lead to overfitting of the training set, while architectures with insufficient complexity may fail to reproduce accurately the structure of the target function, leading to underfitting. Once the NN architecture is fixed, the learning procedure is determined by the choice of the hyperparameters regulating the optimisation process. In the present implementation, the training was divided into six consecutive learning stages, each characterised by a different hyperparameter configuration. This progressive strategy allows the optimisation procedure to explore efficiently the training set during the initial stages of the training and subsequently refine the solution close to the minimum of the loss function. The adopted hyperparameters are:

- learning rate: regulates the magnitude of the optimisation steps during the gradient-descent procedure;

Hyperparameter	Values for the six learning stages
Learning rate	$[10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}]$
Batch size	$[10^2, 5\times10^2, 10^3, 5\times10^3, 10^4, 2\times10^5]$
Gradient accumulation steps	[1, 1, 1, 1, 1, 1]
Patience	[20, 20, 20, 20, 20, 20]
Maximum epochs	$[2\times10^3, 2\times10^3, 2\times10^3, 2\times10^3, 2\times10^3, 2\times10^3]$

Table 4.2: Hyperparameter configurations adopted during the six consecutive learning stages of the emulator training. The progressive reduction of the learning rate and the simultaneous increase of the batch size allow the optimisation procedure to transition from an initial broad exploration of the weight space to a more stable fine-tuning phase close to the minimum of the loss function.

- batch size: number of training samples processed before updating the network weights;
- gradient accumulation steps: number of batches over which the gradients are accumulated before performing a weight update;
- patience values: maximum number of epochs without improvement in the loss function before interrupting the current learning stage;
- max epochs: maximum number of complete passes through the training data set allowed during each learning stage.

The choice of the hyperparameters strongly affects both the convergence stability and the final accuracy of the model. Larger learning rates allow the optimisation algorithm to explore the weight space more rapidly, although excessively large values may lead to unstable convergence. Conversely, smaller learning rates improve the precision of the optimisation close to the minimum of the loss function, at the cost of slower convergence. Similarly, larger batch sizes provide a more stable estimate of the gradients, while smaller batches introduce additional stochasticity that may help the optimisation escape shallow local minima. After several preliminary tests, the best performance was obtained through a progressive reduction of the learning rate combined with an increase of the batch size during the successive learning stages, as shown in Table 4.2. This strategy allows the optimisation procedure to explore rapidly the parameter space at the beginning of the training and subsequently perform a finer adjustment of the network weights during the final stages.

The training procedure was performed on a single `Matrix` node of the Open Physics Hub cluster, employing 112 virtual CPUs running in parallel at a clock frequency of 2.2 GHz. The full training required approximately 25 minutes of computation time.

The accuracy of the trained emulator was evaluated through both the training and testing data sets, as discussed in Sect. 3.1.2. Denoting by $P_{\mathtt{CAMB}}(k)$ the linear matter power spectrum computed with `CAMB` and by $P_{\rm emu}(k)$ the corresponding emulator prediction for the same cosmological parameters, the adopted performance

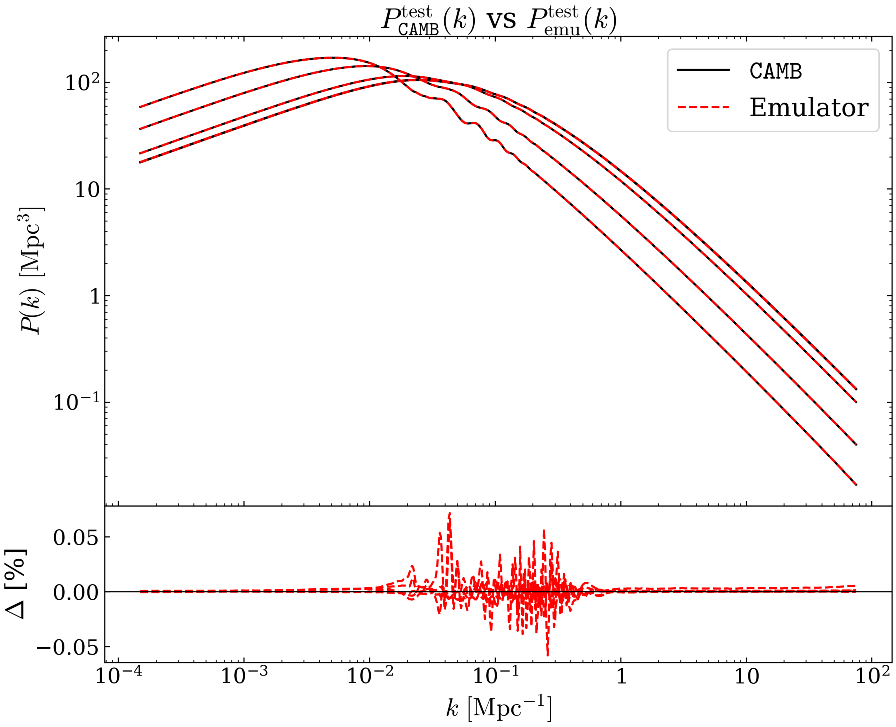

Figure 4.1: Comparison between the linear matter power spectra computed with `CAMB` and the corresponding emulator predictions for four representative cosmological models characterised by different values of $\omega_{\rm b}$ and $\omega_{\rm cdm}$. The upper panel shows the `CAMB` spectra (black solid curves) and the emulator predictions (red dashed curves). The lower panel displays the relative residuals with respect to the `CAMB` spectra.

metric is the Root Mean Square Error (RMSE), defined as

$$\mathrm{RMSE} \equiv \sqrt{\frac{1}{m}\sum_{i=1}^{m}\Big[P_{i,\,\mathrm{emu}}(k) - P_{i,\,\texttt{CAMB}}(k)\Big]^2}\ , \tag{4.4}$$

where m denotes the number of spectra belonging to either the training or testing set. The resulting values are

$$\mathrm{RMSE}_{\rm train} = 9.64 \times 10^{-5}\ , \qquad \mathrm{RMSE}_{\rm test} = 9.67 \times 10^{-5}\ ,$$

corresponding to an absolute difference of the order of 10^{-7}. The close agreement between the training and testing errors indicates that the emulator achieves a balanced fit. A direct comparison between the linear matter power spectra computed with `CAMB` and the corresponding emulator predictions is shown in Fig. 4.1.

In addition to the RMSE, the emulator accuracy was quantified through the relative residuals defined in Sect. 3.3, evaluated over the entire testing set as a function of the wavenumber k. The target accuracy of the implementation is to maintain the residuals below the 0.1% level across the considered wavenumber range. The resulting residual distribution is shown in Fig. 4.2. The red solid curve represents the median residual computed over the testing set, while the shaded region corresponds to the central 68% interval of the distribution, namely the region between

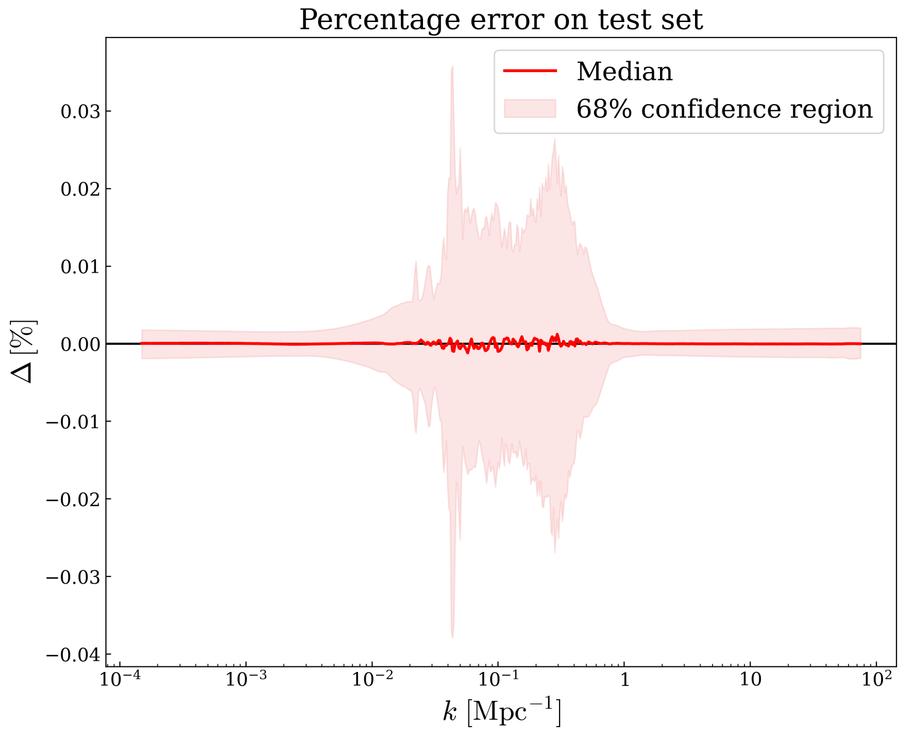

Figure 4.2: Relative residuals between the emulator predictions and the corresponding `CAMB` linear matter power spectra evaluated over the testing set as a function of the wavenumber k. The red solid curve represents the median residual distribution, while the shaded region corresponds to the central 68% interval. The residuals remain below the $\pm 0.05\%$ level over the full wavenumber range and the largest deviations are observed around the BAO scale, as expected.

the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles. The largest discrepancies between the emulator and `CAMB` predictions are observed in correspondence with the BAO region, where the oscillatory structure of the power spectrum is intrinsically more difficult for the NN to reproduce accurately. Nevertheless, the residuals remain well below the target accuracy threshold, oscillating approximately between -0.04% and 0.03%. This level of agreement demonstrates that the emulator is sufficiently accurate to be employed within standard cosmological inference analyses.

The emulator implementation can also be evaluated in comoving units. Figure 4.3 shows a comparison between the linear matter power spectra computed with `CAMB` and with the final emulator implementation for representative ΛCDM, $k\Lambda$CDM, and w_0w_aCDM cosmologies. In all cases, the spectra are reconstructed through the complete mapping of Eq. (4.3), combining the NN prediction, the A_s and n_s rescaling factors, the growth factor evolution, and the curvature correction term. The residuals shown in Fig. 4.3 demonstrate that the emulator is capable of reproducing accurately the linear `CAMB` predictions over a broad range of scales and cosmological models. In all cases, the largest residuals are observed in correspondence with the BAOs region, where the oscillatory structure of the power spectrum makes the learning procedure intrinsically more challenging.

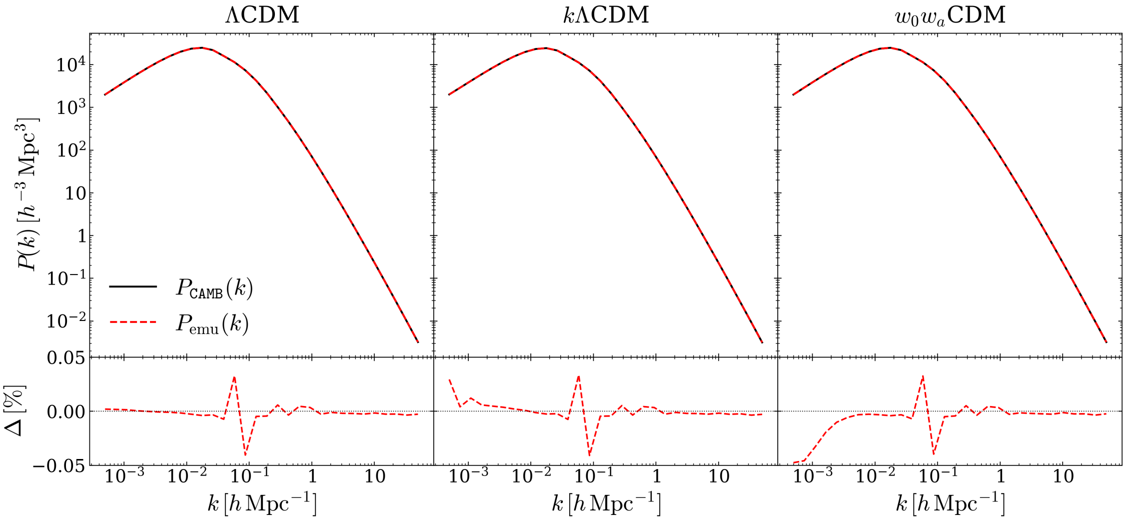

Figure 4.3: Comparison between the linear matter power spectra computed with `CAMB` and with the emulator implementation for ΛCDM, kΛCDM, and w_0w_aCDM cosmologies. All spectra are evaluated at $z = 0$ using the Planck best fit cosmological parameters (Planck collaboration 2020), with the exception of $\Omega_k = 0.01$ for the kΛCDM model and $w_0 = -0.99$, $w_a = -0.02$ for the w_0w_aCDM model. The upper panels show the linear matter power spectra, where the black solid curves correspond to the `CAMB` predictions and the red dashed curves to the emulator reconstruction obtained through Eq. (4.3). The lower panels show the corresponding percentage residuals relative to the `CAMB` spectra.

In the kΛCDM case, the increase of the residuals at large scales is associated with the effect of spatial curvature on the shape of the power spectrum. This contribution is mitigated by the curvature correction factor introduced in Sect. 3.3.1, which improves the reconstruction of the large-scale behaviour. Nevertheless, at large scales the residuals tend to increase for increasingly large deviations of Ω_k from zero. In the w_0w_aCDM model, instead, the accuracy deteriorates at low wavenumber, because of the effects of the dark energy perturbation on super-horizon and near-horizon scales (see Sect. 3.3.2), and sub-percent accuracy guaranteed only for wavenumbers higher than 10^{-3}.

Chapter 5

Results

In the previous chapters, we described the construction of the emulation framework adopted in this work and discussed its implementation within the `CBL` environment. The main motivation for introducing the emulator is the high computational cost associated with repeated evaluations of Boltzmann solvers such as `CAMB`, especially in the context of Bayesian parameter inference analyses, where the theoretical model must be evaluated multiple of times. The purpose of this chapter is therefore to assess the accuracy and the computational performance of the emulator in cosmological analyses. In particular, the emulator is validated by comparing its predictions and the corresponding parameter constraints against those obtained using `CAMB` as the reference theoretical model.

The validation procedure is based on mock matter power spectrum measurements generated from theoretical cosmological models with fixed parameter values. For each cosmological scenario considered in this work, synthetic data sets of the form k, $P(k)$, $\sigma_{P(k)}$ were constructed, where k denotes the wavenumber, $P(k)$ the matter power spectrum, and $\sigma_{P(k)}$ the associated uncertainty. Bayesian parameter estimation analyses were then performed through MCMC techniques. For each data set, the analysis was carried out twice: first using `CAMB` to model the theoretical matter power spectrum, and subsequently replacing it with the emulator. This comparison makes it possible to evaluate whether the emulator is able to reproduce not only the theoretical spectra themselves, but also the posterior distributions and parameter constraints inferred from the MCMC analyses.

The results are presented for three different cosmological scenarios. We first consider the standard flat ΛCDM model, which represents the baseline cosmology adopted throughout modern cosmological analyses. We then extend the comparison to two non-standard scenarios, namely $k\Lambda$CDM and w_0w_aCDM, in order to test the robustness of the emulator in the presence of additional cosmological degrees of freedom.

Finally, the computational performance of the emulator is discussed by directly comparing its execution time with that of `CAMB`. This allows us to quantify the speed-up achieved by the emulation framework and to evaluate its suitability for large-scale cosmological inference applications.

Parameter	Value	Prior range
z	[0, 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75, 2]	–
$k\,[\mathrm{Mpc}^{-1}]$	$[5\times10^{-4}, 50]$	–
Ω_b	0.04	[0.031, 0.069]
Ω_cdm	0.26	[0.061, 0.69]
h	0.69	[0.5, 1]
$A_s\,[10^{-9}]$	1.9	[1.5, 2.6]
n_s	1.1	[0.8, 1.2]
τ	0.0544	–
Ω_γ	9.26535×10^{-5}	–
Ω_DE	$1-\Omega_\mathrm{b}-\Omega_\mathrm{cdm}-\Omega_\gamma$	–

Table 5.1: Cosmological parameters adopted to generate the synthetic matter power spectrum for the ΛCDM scenario, together with the prior ranges used in the MCMC analysis for the free parameters. The parameter A_s and its prior range are expressed in units of 10^{-9}, while Ω_DE is a derived parameter.

5.1 The standard scenario

In the standard cosmological scenario, the synthetic data set was constructed by computing the matter power spectrum with `CAMB` for a fixed set of cosmological parameters, wavenumbers, and redshifts. In particular, the power spectrum was evaluated on 40 logarithmically spaced k bins in the range $5\times10^{-4} \leq k/\mathrm{Mpc}^{-1} \leq 50$. The cosmological parameters adopted to generate the synthetic data set were chosen to be slightly different from the Planck best-fit values (Planck collaboration 2020). This choice was made to show that the emulator is capable of accurately reproducing theoretical predictions also for cosmologies that differ from the reference one adopted during the training procedure. The complete set of parameters employed to generate the mock data is reported in Table 5.1. To mimic observational uncertainties, an artificial Gaussian noise contribution was added to each power spectrum point. In particular, the uncertainty associated with each measurement was assumed to be equal to 5% of the corresponding value of $P(k)$, namely $\sigma_{P(k)} = 0.05\,P(k)$. Under this assumption, the covariance matrix is diagonal, implying that the uncertainties associated with different k bins and redshifts are treated as statistically independent.

Since several cosmological parameters produce partially degenerate effects on the matter power spectrum, a single-redshift analysis would not be sufficient to efficiently break these degeneracies. In particular, variations of parameters such as A_s and h can induce the same change in the amplitude of the power spectrum. To improve the constraining power of the analysis, the synthetic data set was therefore generated at nine different redshifts uniformly distributed in the interval $0 \leq z \leq 2$. The redshift evolution of the matter power spectrum provides additional information that helps disentangle the impact of the different cosmological parameters on both the shape and amplitude of $P(k)$.

The MCMC analysis was performed by varying the cosmological parameters that are most strongly constrained by the matter power spectrum, namely Ω_b, Ω_cdm, A_s,

n_s, and h. These parameters mainly affect either the overall amplitude or the shape of the matter power spectrum, and can therefore be efficiently constrained through clustering observables. On the other hand, parameters such as the optical depth τ and the photon density parameter Ω_γ have a significantly weaker impact on the late-time matter power spectrum and are instead primarily constrained by CMB observations. For this reason, they were kept fixed throughout the analysis.

Uniform priors were adopted for all the free parameters. The prior intervals were selected to be sufficiently broad to fully contain the posterior distributions obtained from the MCMC sampling, while at the same time remaining well entirely within the validity range of the emulator. Under the assumption of independent Gaussian uncertainties, the likelihood function can be written as

$$\mathcal{L} \propto \exp\left(-\frac{\chi^2}{2}\right) ,$$

where

$$\chi^2 = \sum_i \frac{\left[P_i^{\mathrm{data}} - P_i^{\mathrm{model}}\right]^2}{\sigma_i^2} .$$

Here, P_i^{data} denotes the synthetic power spectrum data points, P_i^{model} the theoretical prediction computed either with `CAMB` or with the emulator, and σ_i the corresponding uncertainty. Since the covariance matrix is assumed to be diagonal, no correlation term between different measurements appears in the likelihood expression. The posterior probability distribution is then obtained through Bayes' theorem as the product of the likelihood and the prior probability distribution.

To sample the posterior distribution, we employed a MCMC algorithm using 56 parallel walkers and chains of length 6000 steps each. The use of multiple walkers allows for a more efficient exploration of the parameter space and improves the robustness of the sampling procedure.

The results of the MCMC sampling performed using `CAMB` and the emulator are shown in Fig. 5.1. The figure compares the posterior distributions obtained with the two approaches for the set of cosmological parameters considered in the ΛCDM analysis. The diagonal panels show the one-dimensional marginalised posterior distributions for each parameter, while the off-diagonal panels display the corresponding two-dimensional confidence regions. A first important result emerging from Fig. 5.1 is the excellent agreement between the posterior distributions reconstructed with `CAMB` and those obtained using the emulator. The median values inferred from the two analyses are nearly identical for all the fitted parameters, and the widths of the marginalised distributions are also highly consistent. In particular, the 16^{th}–84^{th} percentile regions obtained with the emulator closely match those reconstructed using `CAMB`. The agreement is not limited to the one-dimensional posterior distributions. The two-dimensional confidence contours also exhibit a very similar shape, orientation, and extent in the two analyses. This aspect is particularly relevant because it shows that the emulator is able to correctly reproduce the parameter degeneracies encoded in the matter power spectrum. The fact that the emulator reproduces both the orientation and the size of these contours shows that

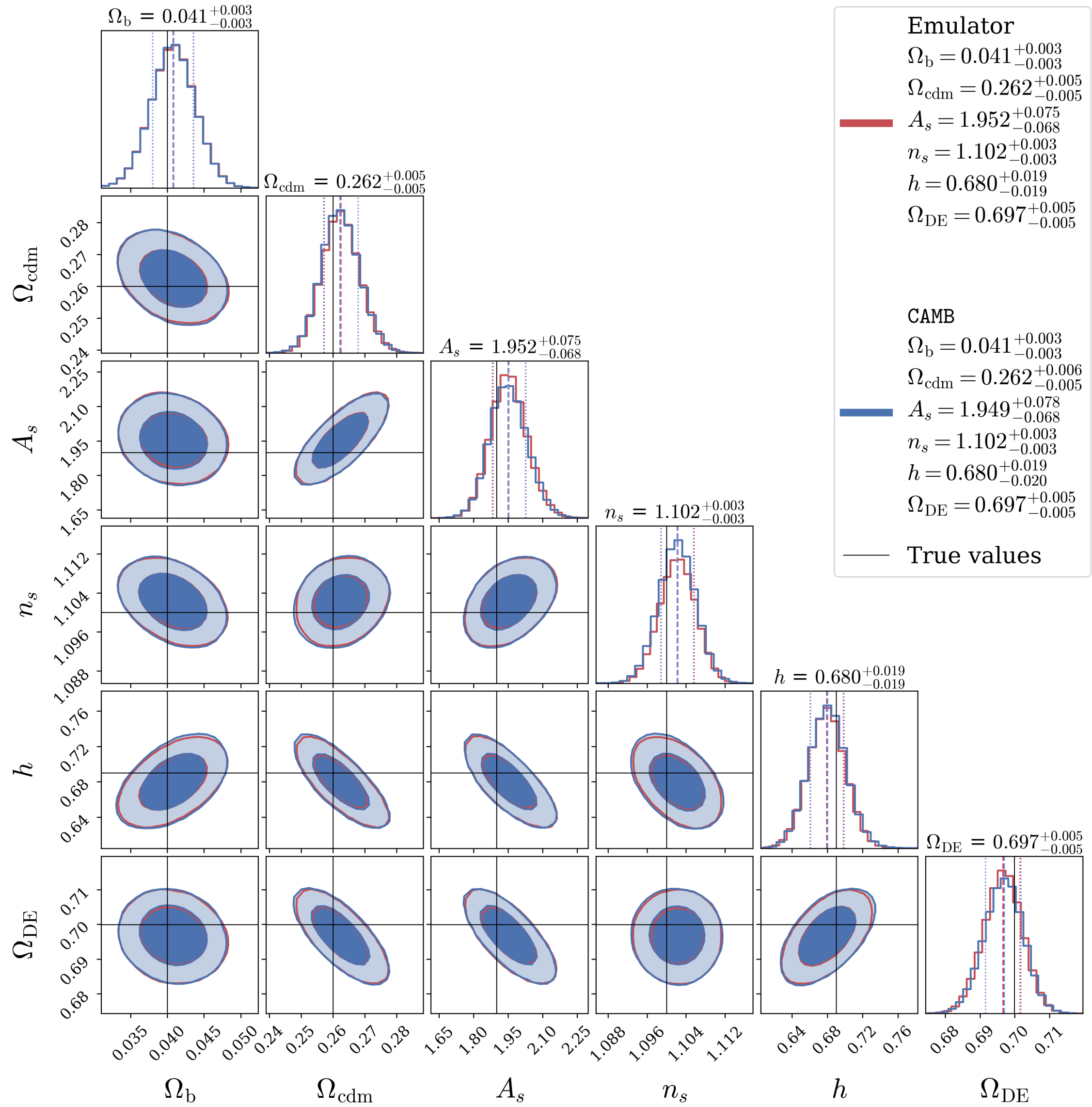

Figure 5.1: Comparison between the posterior distributions obtained using `CAMB` and the emulator in the ΛCDM analysis. The diagonal panels show the one-dimensional marginalised posterior distributions of the fitted cosmological parameters, while the off-diagonal panels display the corresponding two-dimensional confidence contours. The blue distributions correspond to the `CAMB` analysis, whereas the red distributions are obtained using the emulator. Dashed lines indicate the median values of the posterior distributions, while dotted lines mark the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles. The black solid lines indicate the fiducial parameter values used to generate the mock dataset.

it accurately captures the response of the matter power spectrum to variations of the cosmological parameters across the explored parameter space. Another important aspect visible in Fig. 5.1 is that the true cosmological values used to generate the synthetic data set always lie well within the reconstructed posterior regions. This provides a consistency check for both the mock-data generation procedure and the modelling pipeline. At the same time, the posterior distributions are not expected to be perfectly centred on the input values. Indeed, random noise was added to the synthetic power-spectrum measurements in order to mimic realistic observational uncertainties, and therefore the maximum-posterior and best-fit parameter values are expected to fluctuate around the underlying cosmology.

Overall, the comparison presented in Fig. 5.1 shows that the emulator can reliably reproduce the results obtained with the Boltzmann solver not only at the level of the theoretical matter power spectrum, but also within a full Bayesian parameter estimation pipeline. This validates the emulator as a robust and accurate alternative to `CAMB` for cosmological inference applications in ΛCDM models.

5.2 The non-standard scenarios

We now extend the analysis to non-standard cosmological models. In particular, we consider two extensions of the ΛCDM framework, namely the $k\Lambda$CDM and the $w_0 w_a$CDM scenarios. These models introduce additional cosmological degrees of freedom with respect to the flat ΛCDM case and therefore provide a more stringent test for the emulation framework.

5.2.1 $k\Lambda$CDM

We first focus on the $k\Lambda$CDM model, in which the flatness assumption is relaxed by allowing for a non-vanishing curvature density parameter Ω_k. In this scenario, the MCMC analysis was performed by varying Ω_b, Ω_cdm, and Ω_k, while the remaining cosmological parameters were kept fixed to the values reported in Table 5.2. The latter choice was adopted to suppress degeneracies with other cosmological parameters, which could otherwise lead to poorly constrained directions in parameter space and consequently to non-closed confidence regions. The synthetic data set was generated following the same procedure adopted for the ΛCDM analysis. However, in this case an additional consideration concerning the low-k behaviour of the emulator becomes relevant. As discussed in Sect. 3.3.1, the emulation procedure for non-flat cosmologies includes a phenomenological curvature correction factor calibrated against `CAMB` predictions. Although this correction substantially improves the agreement with the Boltzmann solver, the reconstruction of the spectrum at very low wavenumbers remains more challenging than in the flat case. In particular, in the low-k regime, the residual inaccuracy associated with the correction factor adopted in the rescaling procedure generates a small departure from the `CAMB` prediction. When the posterior distribution is sampled through the MCMC algorithm, different values of Ω_k are explored around the maximum of the posterior. As a consequence, this effect on the large-scale evolution can generate residuals that exceed the sub-percent accuracy

Parameter	Value	Prior range
z	[0, 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75, 2]	–
$k\,[\mathrm{Mpc}^{-1}]$	$[8 \times 10^{-4}, 50]$	–
Ω_b	0.04	[0.031, 0.069]
Ω_cdm	0.26	[0.061, 0.69]
Ω_k	0	$[-0.5, 4]$
h	0.69	–
$A_s\,[10^{-9}]$	1.9×10^{-9}	–
n_s	1.1	–
τ	0.0544	–
Ω_γ	9.26535×10^{-5}	–
Ω_DE	$1 - \Omega_\mathrm{b} - \Omega_\mathrm{cdm} - \Omega_\gamma - \Omega_k$	–

Table 5.2: Cosmological parameters adopted to generate the synthetic matter power spectrum for the $k\Lambda$CDM scenario, together with the prior ranges used in the MCMC analysis for the free parameters. Compared to the ΛCDM case, the minimum wavenumber was increased in order to exclude the low-k region where the combined effect of the weak departure of the exact linear evolution from the scale-independent growth-factor and curvature corrections produces larger residuals in the emulated spectrum.

level requested at very low k. In practice, we find that maintaining residuals below approximately 0.05% throughout the sampled parameter space requires excluding the lowest-k region from the analysis. For this reason, unlike the ΛCDM case, the synthetic power spectrum was evaluated on 40 logarithmically spaced wavenumber bins in the range $8 \times 10^{-4} \leq k/\mathrm{Mpc}^{-1} \leq 50$. This choice does not reflect a limitation of the MCMC procedure itself, but rather a conservative selection adopted to ensure that the emulator remains highly accurate over the full parameter region explored during the sampling. Moreover, the excluded scales lie well beyond the range typically employed in current cosmological analyses and therefore have a negligible impact on the practical applicability of the method.

As in the previous case, a Gaussian uncertainty equal to 5% of the corresponding value of the power spectrum was assigned to each data point. The synthetic data set was generated at nine different redshifts uniformly distributed in the interval $0 \leq z \leq 2$, in order to improve the constraining power of the analysis and reduce parameter degeneracies through the redshift evolution of the matter power spectrum.

Uniform priors were adopted for all the free parameters, while the likelihood function was constructed assuming statistically independent Gaussian uncertainties, consistently with the ΛCDM analysis. The posterior distribution was sampled using the same MCMC configuration discussed in the previous section, namely 56 parallel walkers and chains of length 6000 steps each.

The posterior distributions obtained using `CAMB` and the emulator are shown in Fig. 5.2. As in the ΛCDM analysis, the diagonal panels display the one-dimensional marginalised posterior distributions, while the off-diagonal panels show the corresponding two-dimensional confidence contours. Also in this non-standard scenario, the agreement between the emulator and `CAMB` remains very good throughout the

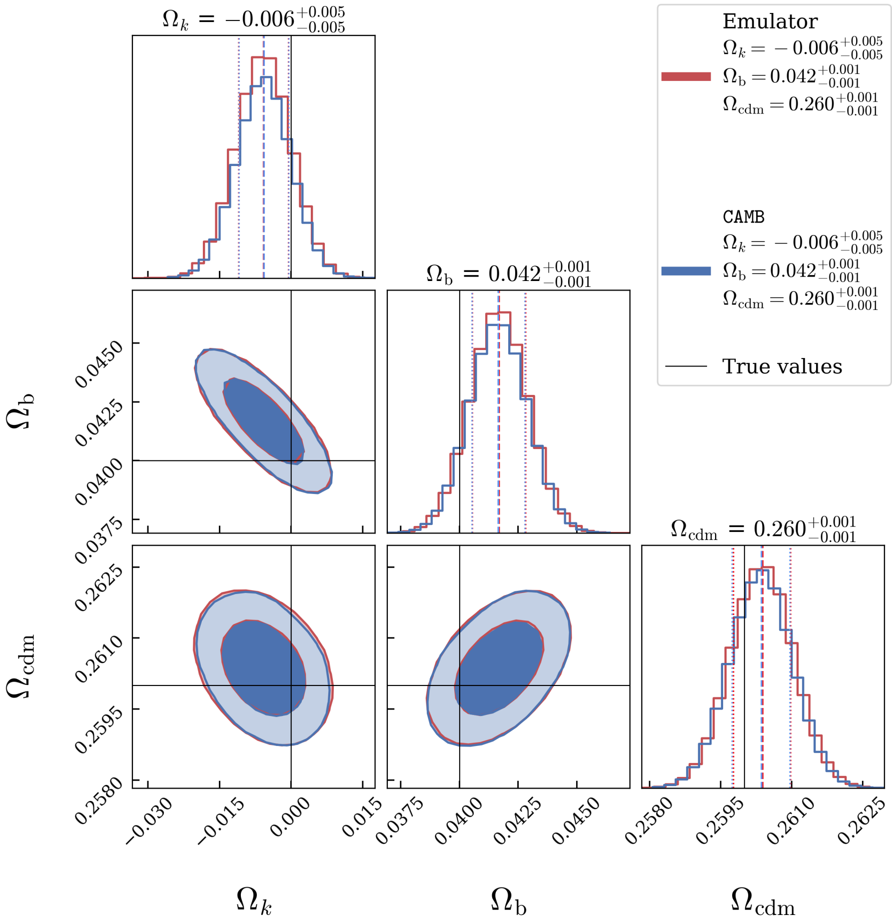

Figure 5.2: Comparison between the posterior distributions obtained using `CAMB` and the emulator for the $k\Lambda$CDM scenario. The diagonal panels show the one-dimensional marginalised posterior distributions of the fitted cosmological parameters, while the off-diagonal panels display the corresponding two-dimensional confidence contours. The blue distributions correspond to the `CAMB` analysis, whereas the red distributions are obtained using the emulator. Dashed lines indicate the median values of the posterior distributions, while dotted lines mark the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles. The black solid lines indicate the fiducial parameter values used to generate the mock dataset.

explored parameter space. The posterior contours reconstructed with the emulator closely follow those obtained with the Boltzmann solver, reproducing both the extent and the orientation of the parameter degeneracies. This confirms that the emulator is able to accurately capture the dependence of the matter power spectrum on the curvature parameter together with the standard matter density components within the selected wavenumber range.

Compared to the flat ΛCDM case, the posterior distributions exhibit a slightly broader degeneracy in the Ω_{cdm}–Ω_k plane, as visible in Fig. 5.2. A small displacement of the marginalised constraints with respect to the fiducial cosmological parameters is also observed. Such a behaviour is expected when fitting noisy mock data. The slight offset may additionally reflect the treatment adopted out work, which relies on a scale-independent growth factor, although it is more plausible that it originates from ordinary fluctuations associated with the posterior sampling. In either case, the effect is negligible for the purposes of the present analysis. More importantly, the two-dimensional confidence contours remain fully consistent with the fiducial cosmological parameters within the 68% credible regions. This confirms that the mock data vector represents a statistically plausible realisation of the underlying cosmological model and that the recovered parameter constraints are consistent with the expected level of statistical fluctuations. Furthermore, the posterior distributions obtained with the emulator closely reproduce those obtained with `CAMB`, with nearly indistinguishable confidence contours across the full parameter space. This agreement demonstrates that the emulation framework provides reliable parameter inference also in the presence of spatial curvature within the range of scales considered in this work.

5.2.2 $w_0 w_a$CDM

In the $w_0 w_a$CDM scenario, we consider a dynamical dark energy model in which the dark energy equation of state evolves with redshift according to the parametrisation introduced in Sect. 1.3.2. In this case, the MCMC analysis was performed by varying the parameters w_0, w_a, Ω_{b}, and Ω_{cdm}, while all the remaining cosmological quantities were kept fixed. The synthetic data set was generated following the same procedure adopted for the previous cosmological scenarios. However, as discussed in Sect. 3.3.2, the low-k regime is affected by additional scale-dependent effects, as the presence of dark energy perturbations on large scales, that complicate its accurate modelling. As a consequence, accurately reproducing the behaviour of the `CAMB` matter power spectrum at very low wavenumbers becomes more challenging than in both the ΛCDM and kΛCDM cases. Within the parameter region explored during the MCMC sampling, we find that residuals can exceeding the target subpercent accuracy threshold at low k. For this reason, in order to maintain residuals below approximately 0.05% over the full sampled parameter space, the analysis was restricted to a more conservative wavenumber interval with respect to the previous scenarios. The synthetic matter power spectrum was therefore computed on 40 logarithmically spaced wavenumber bins in the range $3 \times 10^{-3} \leq k/\mathrm{Mpc}^{-1} \leq 50$. Compared to the previous analyses, the larger minimum wavenumber adopted in this

Parameter	Value	Prior range
z	[0, 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75, 2]	–
$k\,[\mathrm{Mpc}^{-1}]$	$[3 \times 10^{-3}, 50]$	–
Ω_{b}	0.04	[0.031, 0.069]
Ω_{cdm}	0.26	[0.061, 0.69]
w_0	-1	$[-2, 0.5]$
w_a	0	$[-2.5, 1]$
h	0.69	–
$A_s\,[10^{-9}]$	1.9×10^{-9}	–
n_s	1.1	–
τ	0.0544	–
Ω_γ	9.26535×10^{-5}	–

Table 5.3: Cosmological parameters adopted to generate the synthetic matter power spectrum for the w_0w_aCDM scenario, together with the prior ranges used in the MCMC analysis for the free parameters. Compared to the previous cosmological scenarios, the minimum wavenumber was increased in order to maintain residuals small in the emulated spectrum.

case reflects the increased difficulty in accurately modelling the large-scale behaviour of the spectrum. As in the previous scenarios, the adopted k range is determined by the requirement of maintaining a high level of emulation accuracy across the entire parameter space explored by the MCMC sampler, and the impact of this restriction on practical applications is minimal, since the excluded scales are far larger than those probed by current cosmological observations.

The cosmological parameters adopted to generate the synthetic data set are summarised in Table 5.3, together with the prior ranges employed in the MCMC analysis. Apart from the different choice of the k range, both the synthetic data generation and the MCMC configuration are identical to those adopted in the previous analyses.

The posterior distributions reconstructed using `CAMB` and the emulator are shown in Fig. 5.3. Also in this case, the emulator successfully reproduces the posterior structure obtained with the Boltzmann solver throughout the explored parameter space. The confidence contours reconstructed with the emulator remain in very good agreement with those obtained using `CAMB`, both in terms of their orientation and their overall extent. Compared to the previous cosmological scenarios, the posterior distributions exhibit stronger degeneracy directions. This behaviour is expected, since variations of w_0 and w_a can partially mimic changes in the matter density parameters through their impact on the expansion history and on the growth of cosmic structures. As in the $k\Lambda$CDM analysis, the true cosmological parameters lie slightly outside the central 68% confidence interval of the corresponding one-dimensional marginalised posterior distributions. However, this behaviour is common to both the `CAMB` and emulator analyses, and expected, since the synthetic data set includes a random noise realisation that causes the recovered parameter values to fluctuate around the fiducial cosmology. Indeed, the fiducial cosmological parameters remain

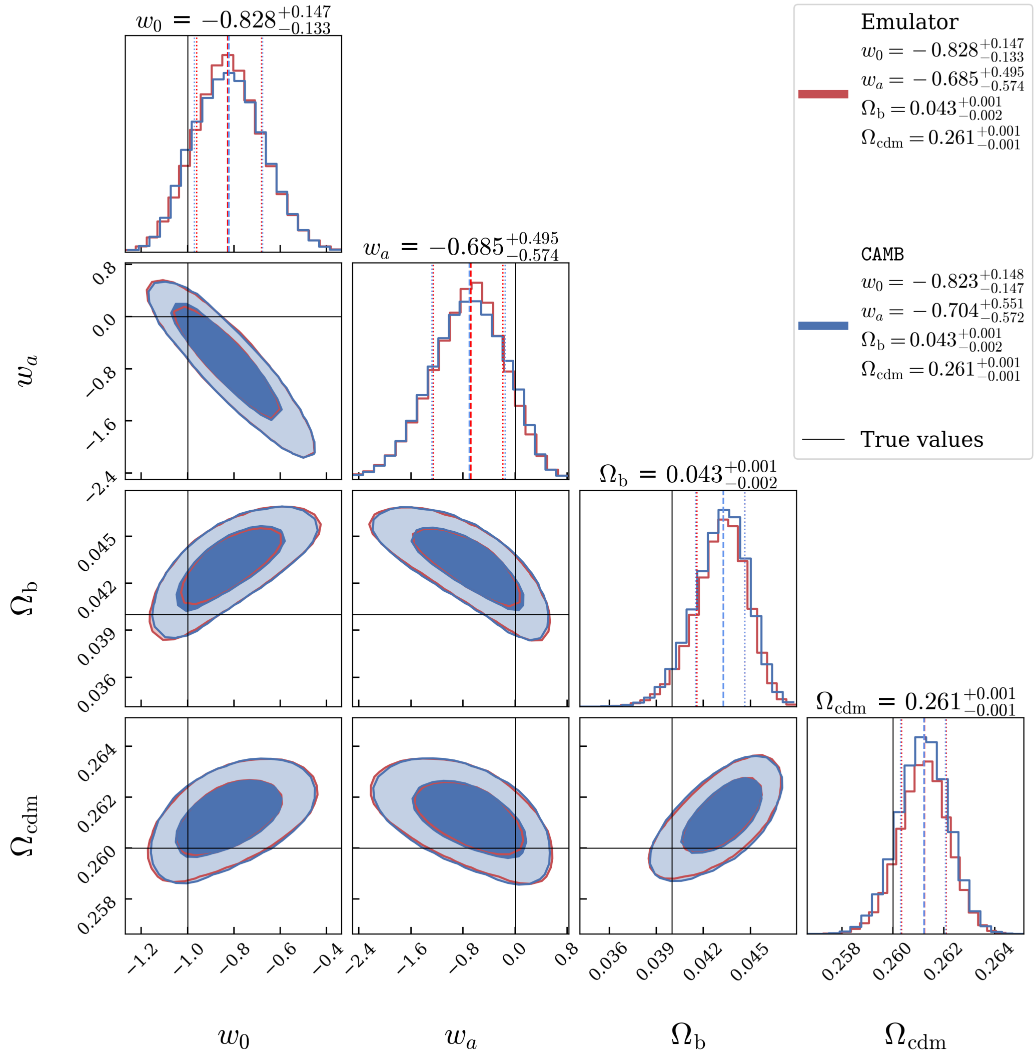

Figure 5.3: Comparison between the posterior distributions obtained using `CAMB` and the emulator for the $w_0 w_a$CDM scenario. The diagonal panels show the one-dimensional marginalised posterior distributions of the fitted cosmological parameters, while the off-diagonal panels display the corresponding two-dimensional confidence contours. The blue distributions correspond to the `CAMB` analysis, whereas the red distributions are obtained using the emulator. Dashed lines indicate the posterior median values, while dotted lines mark the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles. The black solid lines indicate the fiducial parameter values used to generate the mock dataset.

consistent with the reconstructed two-dimensional confidence contours, while apparent discrepancies can appear in one-dimensional marginalised distribution, as the strong parameter degeneracies and non-Gaussian posterior structure can lead to projection effects.

The comparison presented in Fig. 5.3 shows that the emulator remains reliable also in the presence of dynamical dark energy models within the selected wavenumber range, despite the stronger parameter degeneracies and the increased complexity of the large-scale behaviour of the matter power spectrum.

The analyses presented in this chapter also provide information on the effective validity range of the emulation framework in the different cosmological scenarios considered in this work. In the standard ΛCDM case, the emulator remains highly accurate over a very broad wavenumber interval extending down to the whole emulation range. The effect of spatial curvature on very low-k regime makes the spectrum more difficult to reproduce accurately (see Sect. 3.3.1), requiring the minimum wavenumber to be increased to $k = 8 \times 10^{-4}\,\mathrm{Mpc}^{-1}$. Finally, in the $w_0 w_a$CDM scenario, a more conservative validity range is adopted, and sub-percent residuals are guaranteed only for wavelenghts higher than $k = 3 \times 10^{-3}\,\mathrm{Mpc}^{-1}$. Nevertheless, even in the most challenging scenario considered in this work, the emulator remains accurate over a wavenumber range that is significantly broader than the one typically adopted in observational LSS analyses.

5.3 Speed-up time

Having verified the accuracy of the emulator with respect to `CAMB` in the three cosmological scenarios considered, we now discuss the corresponding computational improvement obtained through the emulation framework. As discussed in Sect. 3.2.2, Bayesian parameter estimation techniques based on MCMC sampling require a very large number of posterior evaluations. Since each evaluation involves the computation of the theoretical matter power spectrum, direct calls to Boltzmann solvers become computationally expensive, especially when large multidimensional parameter spaces are explored. One of the main motivations of the present work is therefore the development of an emulation framework capable of accurately reproducing the `CAMB` predictions while substantially reducing the computational cost associated with cosmological inference analyses.

All the Bayesian analyses presented in this chapter, both those performed using `CAMB` and those based on the emulator, were executed on the Open Physics Hub computing cluster `Matrix`. In particular, the computations were performed on a single node equipped with 112 CPUs running in parallel at a clock frequency of 2.2 GHz. The same computational setup, including the same MCMC configuration and hardware environment, was adopted for both approaches in order to ensure a consistent and reliable comparison of the execution times. The computational improvement provided by the emulator can be quantified through the speed-up factor. If $T_{\texttt{CAMB}}$ denotes the execution time required to complete a Bayesian analysis using `CAMB`, and T_{emu} the corresponding execution time obtained using the emulator, the

Scenario	`CAMB` time [h:m:s]	Emulator time [h:m:s]	Speed-up factor
ΛCDM	25:44:25	0:01:41	922
$k\Lambda$CDM	20:03:55	0:01:13	986
w_0w_aCDM	30:58:05	0:00:44	2513

Table 5.4: Total execution times required to complete the MCMC analyses using `CAMB` and the emulator for the ΛCDM, $k\Lambda$CDM, and w_0w_aCDM cosmological scenarios. The same computational setup and MCMC configuration were adopted in all cases. The last column reports the corresponding speed-up factors obtained through Eq. 5.1.

speed-up factor S is defined as

$$S \equiv \frac{T_{\texttt{CAMB}}}{T_{\rm emu}} \,. \tag{5.1}$$

This quantity therefore measures how many times faster the emulator is with respect to the direct Boltzmann solver evaluation under identical computational conditions. The execution times measured for the three cosmological scenarios considered in this work are reported in Table 5.4. The reported timings correspond to the total execution time required to complete the full MCMC analyses. The comparison was intentionally designed to favour the Boltzmann solver. While the posterior requires power-spectrum evaluations at nine different redshifts, `CAMB` was configured to compute all of them within a single call, thereby solving the underlying Einstein–Boltzmann system only once for each cosmological model. In contrast, the emulator is evaluated independently at each redshift. The resulting speed-up factors should therefore be regarded as conservative estimates of the computational gain provided by the emulation framework.

A first important result emerging from Table 5.4 is that the emulator reduces the computational cost of the Bayesian analyses by almost three orders of magnitude in all the cosmological scenarios considered in this work. In the ΛCDM and $k\Lambda$CDM cases, the emulator provides a speed-up factor close to 10^3, reducing execution times from approximately one day to about one minute. An even larger improvement is obtained in the w_0w_aCDM scenario, where the emulator achieves a speed-up factor greater than 2.5×10^3. The larger computational gain observed in the dynamical dark energy scenario is mainly related to the increased complexity of the corresponding `CAMB` computations. In models with evolving dark energy equation of state, the Boltzmann solver requires a more computationally demanding treatment of the background evolution and of the perturbation equations, leading to longer execution times for each likelihood evaluation. Since the computational cost of the emulator remains nearly unchanged across the different cosmological scenarios, the relative improvement becomes even more significant in the w_0w_aCDM case.

The computational improvement obtained through the emulator is achieved while maintaining the level of accuracy discussed in the previous sections within the validated wavenumber ranges of each cosmological scenario. In particular, sub-percent agreement with `CAMB` is preserved over a wide range of modes, effectively covering all the physical scales probed in modern cosmological analyses.

Chapter 6

Conclusions

The matter power spectrum is one of the primary observables used to constrain cosmological models and infer the fundamental properties of the Universe. Modern cosmological analyses rely on Bayesian inference techniques that require repeated evaluations of its theoretical prediction over large parameter spaces. As a consequence, parameter estimation pipelines often become computationally expensive, since Boltzmann solvers such as `CAMB` and `CLASS` must be executed thousands or millions of times during the sampling process. In recent years, ML techniques have emerged as an effective solution to alleviate this computational burden while preserving the accuracy of first-principles physical calculations. Several emulators, including `BACCO`, `CosmoPower`, and `Aletheia` (see Sect 3.2), have demonstrated that NN-based approaches can accurately reproduce cosmological observables over wide cosmological parameter spaces while reducing computational costs by orders of magnitude.

6.1 Main results

The goal of this Thesis was to develop a new emulator of the linear matter power spectrum capable of surpassing the accuracy of publicly available codes, while remaining valid across broad parameter ranges and a variety of cosmological models. To achieve this ambitious objective, we relied on a state-of-the-art methodology that combines the capabilities of ML with a technique for reducing the effective dimensionality of the parameter space.

The main methodological developments and validation results presented in this work can be summarised as follows.

- We exploited the evolution mapping formalism to separately model the cosmological parameters regulating the shape and the amplitude of the linear matter power spectrum, and we used an analytical rescaling of A_s and n_s to parametrise the primordial matter power spectrum (see Sect. 3.3). The remaining shape parameters, ω_{b} and ω_{cdm}, were incorporated into the emulator pipeline using the `CosmoPower` framework to train a NN. Specifically, the network was trained to predict the linear matter power spectrum computed by CAMB as a function only of these two parameters. By exploiting physically motivated degeneracies between cosmological parameters, this approach proved effective in drastically reducing the dimensionality of the parameter space that must be sampled during the training phase, and allowed us to build a high-precision emulator.

- To infer the amplitude of the power spectrum, we instead used a rescaling based on the unnormalised growth factor, for which we developed a new implementation characterised by high accuracy and computational efficiency. This formulation provides a unified description of the linear growth of matter perturbations and allows the same emulator architecture to be employed across cosmological scenarios characterised by different expansion histories. Finally, to complete the emulator pipeline, we implemented a phenomenological correction to account for spatial curvature, extending the validity of the method to non-flat cosmologies. The full pipeline was then integrated into the CBL framework extending the pre-existing `Emulator` class (see Sect. 4.1), ensuring straightforward utilisation by the user and direct applicability to several cosmological inference pipelines already present in the library.

- The validation of the emulator was then performed with a testing set, and its accuracy was also proven in different scenarios, assuming ΛCDM, kΛCDM, and w_0w_aCDM cosmologies (see Sect. 4.2). The results showed that the emulator accurately reproduces the linear matter power spectrum predicted by CAMB over the full range of scales relevant for current LSS analyses. Within the adopted validity range, residuals remain well below the 0.1% accuracy level, demonstrating its competitiveness against existing publicly available emulators, which generally yield residuals on the order of $1-2\%$.

- The effective performance of the emulator was then assessed through Bayesian analyses performed on synthetic data sets (see Sect. 5.1 and Sect. 5.2). The inferred cosmological parameters and corresponding confidence intervals were found to be statistically indistinguishable from those obtained using CAMB, demonstrating that the emulator accuracy is sufficient for parameter-estimation applications and that the remaining residuals do not introduce measurable biases in the recovered constraints. Beyond validating the achieved accuracy, employing the emulator within the context of Bayesian inference—specifically for sampling the posterior distribution via MCMC—allowed us to assess the computational time savings. In particular, the speed-up compared to CAMB is approximately three orders of magnitude, reducing the time required for MCMC runs from the order of one day to about one minute.

However, the methodology presented in this Thesis is subject to certain limitations, which are inherently tied to the assumptions underlying the evolution mapping formalism. Specifically, the approach relies on the scale independence of the linear growth factor. Small departures from this approximation arise on scales approaching the horizon, where relativistic effects and dark energy perturbations induce a scale-dependent growth of matter perturbations. Rather than introducing empirical corrections, we adopted a conservative validity range. Because the excluded scales lie well outside the region probed by current observations, these theoretical restrictions have a negligible impact on practical LSS analyses.

Despite these limitations, the adopted physically motivated mapping offers a highly effective strategy for extending the predictive power of NN emulators beyond the parameter space directly sampled during training. Within the accepted validity range, the approximation remains exceptionally robust, ensuring an accuracy at the sub-0.05% level over almost the entire wavenumber range considered.

Overall, this work demonstrates that combining NN with physically motivated rescaling techniques yields an accurate, computationally efficient, and highly flexible framework for cosmological parameter inference. The achieved balance between predictive precision, execution speed, and extensibility makes the proposed methodology a highly promising tool for the analysis of current and forthcoming LSS data sets.

6.2 Future perspectives

A natural extension of the present work concerns the application of the emulator to the modelling of additional LSS observables. Although this Thesis focused on the linear matter power spectrum, as discussed in Sect. 2.3.1, this quantity constitutes a fundamental ingredient in the theoretical description of several cosmological probes.

In particular, the emulator could be employed in analyses based on the 2PCF, whose theoretical prediction is obtained from the matter power spectrum through a Fourier transform. Similarly, both the theoretical prediction of the HMF and the VSF require repeated evaluations of the variance of matter density fluctuations, which is computed through integrals of the linear matter power spectrum. Consequently, cosmological analyses based on these probes will greatly benefit from the matter power spectrum emulator presented in this Thesis. The high computational efficiency achieved by our emulator can therefore be exploited within modelling pipelines for any observable whose theoretical prediction relies on the matter power spectrum, significantly reducing the computational cost associated with repeated model evaluations during parameter inference analyses. To this end, our future efforts will focus on extending the current analysis to thoroughly assess the emulator's impact on these applications, paving the way for its adoption in a much broader range of LSS studies.

Another promising direction concerns the extension of the evolution-mapping formalism to cosmological models characterised by scale-dependent growth. As discussed in Sect. 3.3.2, the current implementation relies on the approximation that the growth of matter perturbations can be described by a scale-independent growth factor, $D(z)$. This assumption is responsible for both the simplicity and the computational efficiency of the method. However, this condition is violated in cosmologies with massive neutrinos and in specific modified gravity models, where the growth of structure becomes scale-dependent, denoted as $D(k, z)$. A natural extension of the present framework would therefore consist in replacing the standard growth factor with a scale-dependent generalisation. Such an approach would preserve the underlying structure of evolution mapping while extending its applicability to a broader class of cosmological models. Nevertheless, calculating $D(k, z)$ requires the numerical solution of significantly more complex perturbation equations, which falls outside the scope of this Thesis work. Furthermore, directly computing an accurate scale-dependent growth factor could compromise the efficiency of the emulator pipeline due to the increased computational cost. An alternative promising solution for future work is therefore the implementation of a correction function driven by a second trained NN. Since this correction is expected to be smooth and exhibit a clean dependence on cosmology, the second NN could efficiently emulate the ratio between the scale-independent prediction and the true power spectrum, thereby extending the emulator's validity without sacrificing performance.

A further extension concerns the inclusion of nonlinear gravitational evolution. Although the emulator developed in this work is designed to reproduce the linear matter power spectrum, the nonlinear correction retains information about the whole growth history of cosmic structures and can therefore be modelled from the linear power spectrum. Following this idea, the emulator developed in this Thesis could be combined with a second NN trained to predict the nonlinear correction factor, $P_{\rm NL}(k, z)/P_{\rm L}(k, z)$. The nonlinear matter power spectrum could then be reconstructed by multiplying the emulated linear prediction by the corresponding nonlinear correction, as shown in Sánchez et al. (2025). Since this ratio is expected to exhibit a smoother dependence on cosmological parameters than the full nonlinear spectrum, such a strategy could preserve the computational efficiency of the evolution-mapping framework while substantially extending its range of applicability.

Taken together, these possible developments highlight the modular nature of the methodology presented in this Thesis. Although the current implementation is focused on the accurate and efficient emulation of the linear matter power spectrum, the framework naturally accommodates extensions to additional observables, alternative cosmological scenarios, and nonlinear structure formation, providing a clear path towards increasingly realistic cosmological applications.

Bibliography

Adame, A., Aguilar, J., Ahlen, S., et al. 2025a, Journal of Cosmology and Astroparticle Physics, 2025, 028

Adame, A., Aguilar, J., Ahlen, S., et al. 2025b, Journal of Cosmology and Astroparticle Physics, 2025, 021

Adame, A. G., Aguilar, J., Ahlen, S., et al. 2025, Journal of Cosmology and Astroparticle Physics, 2025, 021

Alam, S., Ata, M., Bailey, S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617–2652

Alam, S., Aubert, M., Avila, S., et al. 2021, Physical Review D, 103, 083533

Alcock, C. & Paczynski, B. 1979, Nature, 281, 358

Aricò, G., Angulo, R. E., & Zennaro, M. 2022, arXiv e-prints, arXiv:2104.14568

Blas, D., Garny, M., Konstandin, T., & Lesgourgues, J. 2014, Journal of Cosmology and Astroparticle Physics, 2014, 039–039

Bocquet, S., Saro, A., Dolag, K., & Mohr, J. J. 2015, Monthly Notices of the Royal Astronomical Society, 456, 2361–2373

Brout, D., Scolnic, D., Popovic, B., et al. 2022, The Astrophysical Journal, 938, 110

Chevallier, M. & Polarski, D. 2001, International Journal of Modern Physics D, 10, 213–223

Clifton, T., Ferreira, P. G., Padilla, A., & Skordis, C. 2012, Physics Reports, 513, 1–189

Contarini, S., Pisani, A., Hamaus, N., et al. 2023, The Astrophysical Journal, 953, 46

Contarini, S., Verza, G., & Pisani, A. 2026, The Astronomy and Astrophysics Review, 34, 1

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, The Astrophysical Journal, 302, L1

DES Collaboration, Abbott, T. M. C., Acevedo, M., et al. 2024, The Astrophysical Journal, 973, L14

DESI Collaboration. 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. 2021a, Astroparticle Physics, 131, 102605

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. 2021b, Astroparticle Physics, 131, 102604

Di Valentino, E., Anchordoqui, L. A., Akarsu, Ö., et al. 2021c, Astroparticle Physics, 131, 102607

Dodelson, S. & Schmidt, F. 2021, Modern Cosmology, 2nd edn. (Academic Press)

Esposito, M., Sánchez, A. G., Bel, J., & Ruiz, A. N. 2024, Monthly Notices of the Royal Astronomical Society, 534, 3906–3915

Fiorilli, A., Ruiz, A. N., Sanchez, A. G., & Esposito, M. 2025, Evolution mapping III: A new recipe for the halo mass function

Fisher, R. A. 1925, Proceedings of the Cambridge Philosophical Society, 22, 700

Fixsen, D. J. 2009, The Astrophysical Journal, 707, 916–920

Ghirardini, V., Bulbul, E., Artis, E., et al. 2024, Astronomy & Astrophysics, 689, A298

Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press), http://www.deeplearningbook.org

Guth, A. H. & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110

Hastings, W. K. 1970, Biometrika, 57, 97

Heath, D. J. 1977, Monthly Notices of the Royal Astronomical Society, 179, 351

Hu, W. & Sawicki, I. 2007, Physical Review D, 76

Hubble, E. 1929, Proceedings of the National Academy of Sciences, 15, 168

Knabenhans, M., Stadel, J., Marelli, S., et al. 2019, Monthly Notices of the Royal Astronomical Society, 484, 5509–5529

Knabenhans, M., Stadel, J., Potter, D., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 2840–2869

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, Euclid Definition Study Report

Lesgourgues, J. 2011, arXiv e-prints, arXiv:1104.2932

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Linder, E. V. 2003, Physical Review Letters, 90

Madhavacheril, M. S., Qu, F. J., Sherwin, B. D., et al. 2024, The Astrophysical Journal, 962, 113

Mangano, G., Miele, G., Pastor, S., et al. 2005, Nuclear Physics B, 729, 221–234

Marulli, F., Veropalumbo, A., & Moresco, M. 2015, Astronomy and Computing, 14, 35–42

Merloni, A., Lamer, G., Liu, T., et al. 2024, Astronomy & Astrophysics, 682, A34

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, The Journal of Chemical Physics, 21, 1087

Mitra, S., Park, C.-G., Choudhury, T. R., & Ratra, B. 2019, Monthly Notices of the Royal Astronomical Society, 487, 5118–5128

Nicolosi, P. 2022, Master's thesis, Alma Mater Studiorum – Università di Bologna

Nielsen, M. A. 2015, Neural Networks and Deep Learning (Determination Press), last accessed: 2026

OPH. 2020, OPH Computing Cluster User Guide, internal documentation

Pezzotta, A., Eggemeier, A., Gambardella, G., et al. 2025, Physical Review D, 112

Planck collaboration. 2020, Astronomy & Astrophysics, 641, A6

Planck Collaboration. 2020, Astronomy & Astrophysics, 641, A1

Rovelli, C. 2021, General Relativity: The Essentials (Cambridge University Press)

Sánchez, A. G., Ruiz, A. N., Rodriguez, F., et al. 2025, arXiv e-prints, arXiv:2511.13826

Sheth, R. K. & van de Weygaert, R. 2004, Monthly Notices of the Royal Astronomical Society, 350, 517–538

Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, Monthly Notices of the Royal Astronomical Society, 341, 1311–1332

Speagle, J. S. 2020, A Conceptual Introduction to Markov Chain Monte Carlo Methods

Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, Mon. Not. R. Astron. Soc., 511, 1771

Sánchez, A. G. 2020, Physical Review D, 102

Sánchez, A. G., Ruiz, A. N., Jara, J. G., & Padilla, N. D. 2022, Monthly Notices of the Royal Astronomical Society, 514, 5673–5685

Tanaka, S., Nishimichi, T., & Kobayashi, Y. 2026, Dark Quest II: A Wide-Coverage Neural Network Emulator of the Nonlinear Matter Power Spectrum Across Extended Cosmologies

Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, The Astrophysical Journal, 688, 709–728

Virtanen, P. et al. 2020, Nature Methods, 17, 261

Vittorio, N. 2018, Cosmology (CRC Press)

Wessing, S. 2015, PhD thesis, Technische Universität Dortmund

Yadav, J., Bharadwaj, S., Pandey, B., & Seshadri, T. R. 2005, Monthly Notices of the Royal Astronomical Society, 364, 601–606