ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Astrofisica e Cosmologia

Deep Neural Networks to unveil the properties of the Cosmic Web

Relatore:
Dott. Federico Marulli

Correlatore:
Ch.mo Prof. Lauro Moscardini

Presentata da:
Jacopo Neri

Sessione IV
Anno Accademico 2019/2020

Indice

Abstract

The large-scale structure of the Universe provides a key probe to test the cosmological framework. To describe the spatial properties and dynamics of cosmic structures a statistical approach is required, to exploit the main observable statistics of the cosmic density field, like the power spectrum and bispectrum, which directly depend on the cosmological parameters. Current and upcoming projects, such as DESI, Euclid, LSST, PFS, SKA and WFIRST, will reconstruct the main properties of the Cosmic Web, at increasingly large volumes. These ground-based and space missions will provide a spectacularly large amount of data that will be used to discriminate among different cosmological scenarios. The *standard* way to exploit astronomical datasets is through statistical inference analayses that require huge computational resources and the assumption of a likelihood model for the considered statistics. An alternative, groundbreaking approach to the problem is to exploit Machine Learning algorithms and, in particular, the so-called Artificial Neural Networks. This represents a novel data analysis approach in Cosmology, that shall become increasingly attractive in the near future to fully exploit the huge amount of observed and simulated datasets that will become available soon. In fact, Artificial Neural Networks can approximate any continuous functions and can be used to represent complex connections between input data and model parameters, after a proper training phase.

The main goal of this Thesis work is to test Machine Learning techniques for cosmological analyses. We develop and validate new methods and numerical algorithms to constrain the main parameters of the standard cosmological model, that is Ω_m, Ω_b, h, n_s, σ_8, exploiting a likelihood-free inference analysis. The training dataset considered in this work consists of a huge set of second-order and third-order statistics of the dark matter density field, measured from the Quijote N-body simulations [Villaescusa-Navarro et al., 2019]. These are one of the largest sets of dark matter N-body simulations currently available, that span a significant range of the cosmological parameters of the standard model. We implement and train new Neural Networks that can take in input measurements of two-point correlation functions, power spectra and bispectra, and provide in output constraints on the main cosmological parameters. After the training and validation phases, we test the accuracy of our implemented Machine Learning algorithms by processing never-seen-before input datasets generated with cosmological parameters

comparable with Planck18 ones [Planck Collaboration et al., 2018]. We find that this statistical procedure can provide robust constraints on some of the aforementioned parameters, in particular Ω_m.

This Thesis work demonstrates that the considered *deep learning* techniques based on state-of-the-art Artificial Neural Networks can be effectively employed in cosmological studies, in particular to constrain the main parameters of the cosmological framework by exploiting the statistics of the large-scale structure of the Universe.

Sommario

La struttura a grande scala dell'Universo è un fondamentale strumento di indagine per testare lo scenario cosmologico. Per descrivere la distribuzione spaziale e le proprietà dinamiche delle strutture cosmiche è necessario un approccio statistico, in grado di sfruttare le principali statistiche osservabili del campo di densità cosmico, come la funzione di correlazione a due punti, lo spettro di potenza e il bispettro, che dipendono direttamente dai parametri cosmologici. Attuali e futuri progetti, come DESI, Euclid, LSST, PSF, SKA e WFIRST, ricostruiranno le principali proprietà della cosiddetta *Ragnatela Cosmica*, su volumi sempre più grandi. Queste missioni, da terra e dallo spazio, produrranno un numero incredibilmente alto di dati che saranno utilizzati per discriminare i diversi scenari cosmologici. Il metodo standard per sfruttare questi dataset è attraverso un'analisi di inferenza statistica che richiede enormi risorse computazionali e necessita dell'assunzione di un modello di *likelihood* per la statistica considerata. Un alternativo e innovativo approccio al problema consiste nell'utilizzare algoritmi di *Machine Learning* e, in particolare, le cosiddette Reti Neurali Artificiali. Questo rappresenta un nuovo metodo di analisi dei dati in Cosmologia, che potrebbe esercitare una sempre maggiore attrattiva nel prossimo futuro per sfruttare al massimo l'enorme afflusso di dati, osservati e simulati, che diventeranno presto disponibili. Infatti, le Reti Neurali Artificiali possono approssimare ogni funzione continua e, dopo un'opportuna fase di allenamento, possono essere utilizzate per rappresentare complesse connessioni tra dati in input e parametri del modello.

Il principale scopo di questa Tesi è testare tecniche di *Machine Learning* per analisi cosmologiche. Abbiamo sviluppato e validato nuovi metodi e algoritmi numerici per vincolare i maggiori parametri del modello cosmologico standard, ovvero Ω_m, Ω_b, h, n_s e σ_8, sfruttando un'inferenza *likelihood-free*. Il dataset di allenamento considerato in questo lavoro consiste nell'enorme set di statistiche di secondo e terzo ordine del campo di densità della materia oscura, misurate dalle simulazioni *N-body* Quijote. Queste ultime sono uno dei più grandi set di simulazioni *N-body* di materia oscura attualmente disponibili, e considerano un intervallo significativo di parametri cosmologici del modello standard. Abbiamo implementato e allenato nuove Reti Neurali che prendono in input misure della funzione di correlazione a due punti, dello spettro di potenza e del bispettro e produ-

cono in output vincoli sui principali parametri cosmologici. Dopo le fasi di allenamento e validazione, abbiamo testato l'accuratezza dei nostri algoritmi di *Machine Learning* processando nuovi dataset, mai visti prima dalle Reti, generati con parametri cosmologici comparabili con quelli di Planck18 [Planck Collaboration et al., 2018]. Abbiamo verificato che questa procedura statistica è in grado di produrre robusti vincoli su alcuni dei suddetti parametri, in particolare Ω_m.

Questo lavoro di Tesi dimostra come le tecniche di *deep learning*, basate sullo stato dell'arte delle Reti Neurali Artificiali, siano utilizzabili per studi cosmologici, in particolare per vincolare i principali parametri dello scenario cosmologico, sfruttando le statistiche a grande scala dell'Universo.

Introduction

The concordance cosmological model, that up to now has exceeded all the main observational tests, is the so-called Λ-cold dark matter (ΛCDM) model. According to this framework, the Universe is dominated by two dark components, the Dark Energy, parameterised by the cosmological constant Λ, and the CDM. Within the ΛCDM model, we can describe the large-scale structures of the Universe following the evolution of the density perturbations. Indeed, the large-scale structure properties of the Cosmic Web provide a fundamental tool to constrain model parameters, and can be exploited with a statistical approach. Second-order and higher-order statistics of the matter density field contain crucial information about the dark component of the Universe (e.g. Peebles [2001]). In particular, the two-point correlation function and power spectrum of extra-galactic sources are key statistics to infer the primary clustering properties of the underlying density field (see e.g. Hawkins et al. [2003], Totsuji and Kihara [1969], Peebles [1980] and many others). Higher-order statistics, such as the three-point correlation function and the bispectrum (see Sefusatti et al. [2006]), are the lowest-order statistical tools to probe the non-Gaussian properties of the perturbations' field (e.g. Takada and Jain [2003]). These functions depend on the parameters of the cosmological model, and can be estimated from observations and simulations using proper statistical estimators (e.g. Peebles and Hauser [1974], Hamilton [1993], Landy and Szalay [1993] and many others). Exploiting the information content of these statistcs, it is possible to constrain model parameters with likelihood inference methods (e.g. Elsner et al. [2020] and Wilking et al. [2015]). Besides this approach, a novel data-driven technique that seems particularly convenient, especially to analyse large datasets, consists of performing likelihood-free inference making use of Machine Learning (ML) techniques. In brief, *standard* cosmological analyses require a customized likelihood for any statistics to be analysed, which involves assumptions from an underlying model. Then the inference method consists of finding the optimal parameters for that particular likelihood. Conversely, the ML approach starts from the most generic likelihood, that is the Artificial Neural Network, whose parameters are fixed to the best values without any assumptions, via the training and validation phases.

In this Thesis work we implement and validate deep Artificial Neural Networks for cosmological inference, that we train on second-order and third-order statistics of the dark matter density field, measured from the Quijote CDM N-body simulations [Villaescusa-Navarro et al., 2019]. ML techniques, especially the *deep learning* algorithms based on Neural Networks, shall become the mainstream toolkit to model the complex relationship between large-scale structure datasets and the underlying cosmology framework. Recently, ML techniques have been applied to many fields of Cosmology, including weak gravitational lensing (i.e. Schmelzle et al. [2017], Merten et al. [2019], Peel et al. [2019]), the cosmic microwave background (i.e. Caldeira et al. [2019], Münchmeyer and Smith [2019], Krachmalnicoff and Tomasi [2019]), the large-scale structure (i.e. Pan et al. [2019], Lucie-Smith et al. [2019], Kodi Ramanah et al. [2020]), gravitational waves (i.e. Dreissigacker et al. [2019], Gebhard et al. [2019]), cosmic reionization (i.e. La Plante and Ntampaka [2019], Gillet et al. [2019]) and supernovae (i.e. Muthukrishna et al. [2019], Moss [2018]).

This Thesis work is organized as follows:

- in Chapter 1, we give a general introduction to Cosmology, describing the theoretical basics of the standard ΛCDM cosmological model.
- in Chapter 2, we analyze the growth of density perturbations, from the primordial Universe to the non-linear regime.
- in Chapter 3, we describe the clustering properties of collapsed perturbations, introducing the fundamental statistical functions to probe the large-scale structure of the Universe.
- in Chapter 4, we give some basic concepts of ML, introducing the Artificial Neural Network properties.
- in Chapter 5, we present the analyzed Quijote N-body simulations and, in particular, their byproducts that will be used as training datasets for our Neural Networks.
- in Chapter 6, we exploit the Neural Networks training and validation phases, describing the data pre-processing and the Neural Network architectures.
- in Chapter 7, we summarize the Neural Networks outcomes for both training and testing datasets.
- in Chapter 8, we compare the outcomes of the ML approach implemented in this Thesis work to literature results and, finally, we suggest future improvements.

Capitolo 1

Basics of Cosmology

Cosmology is the branch of astrophysics that aims at describing the origin and the evolution of our Universe. The starting points of the investigation of the Universe, as a whole, are three fundamental principles:

- **Cosmological Principle**: The spatial mass distribution of the Universe is homogeneous and isotropic on large scales ($\gg 100\ Mpc$).
- **Fair Sample Principle**: If we measure the variance of a quantity by averaging over a sufficiently large volume, the results approach the true ensemble variance. In other words the average of a quantity over sufficiently large and independent volumes is equal to the average over different realizations of the Universe.
- **Copernican Principle**: Human observers do not rest in a privileged or special reference frame in the Universe.

The *Fair Sample principle* will be used for making statistical considerations over the Universe even if it is a unique and not reproducible object. The *Cosmological principle* can be furthermore simplified noting that, assuming the *Copernican principle*, isotropy implies homogeneity. In fact, since we observe an isotropic distribution of matter it means that every physical quantity, like, for example, the velocity field, can not have a preferred axis. For instance, the strain velocity tensor $\frac{\partial v_i}{\partial x_j}$ can be decomposed into symmetric and anti-symmetric parts. The anti-symmetric part corresponds to a rotation, so the velocity field, $\vec{v}$, can be written to the first order in $\vec{r}$ as follows:

$$\vec{v} = \Sigma \cdot \vec{r} + \Omega \vee \vec{r}, \tag{1.1}$$

where Σ and Ω are two tensors that account for radial and rotational contributions respectively. For isotropy, Ω must vanish because rotation would violate the hypoThesis, and every component of Σ must be equal. So there will not be a preferred axis. Thus the previous equation can be rewritten as:

$$\vec{v} = H\vec{r}, \tag{1.2}$$

where H is the value of each non-null components in Σ. This fundamental equation, that will be discussed in Section 1.4, states that the only motion that can preserve isotropy on a local scale is an expansion (or contraction).
The expansion of the Universe, mathematically derived by Friedmann and Lemâitre in the 1920s, was experimentally verified by Hubble in 1929.

1.1 Principles of General Relativity

If we consider the Universe as a whole, and especially at the largest scales, we can indeed assert that it is dominated by gravity. The best description of gravitational field in modern Physics is the general theory of relativity (GR), published by Albert Einstein in 1915 [Einstein, 1915]. This theory was introduced to conciliate Special Relativity (SR) and Newtonian Gravity and to solve empirical problems, such as anomalies of Mercury orbit, deflection of light and gravitational redshift.

Up to now, GR satisfied every experimental test and is able to predict and explain phenomena like gravitational waves, gravitational lensing and black holes that scientists discovered many decades after the Einstein's publication. GR is based on a set of physical and geometrical principles:

- **Spacetime structure**: The spacetime is a four-dimensional differentiable manifold of Lorentzian signature.
- **Principle of General Covariance**: It is possible to extend the laws of Special Relativity to a general reference frame using quantities expressed by a tensor of the spacetime manifold, instead of a Lorentz group tensor.
- **Equivalence principle**: The outcome of any local experiment in a free falling laboratory is independent of the location and velocity of the laboratory.

Starting from these three principles it is possible to build the equation of motion of GR. In what follows, Greek indices indicate spacetime and run from 0 to 3. The summation convention on repeated indices is also assumed.
If we consider a free falling observer, and we assume the *Equivalence principle*, the local metric in its reference frame ξ^μ is the canonical Minkowski metric [Minkowski, 1908], valid for the defined spacetime, at least in a small neighbourhood. It follows that a test particle subject only to gravity must move on a straight line given by the following equation (see [Weinberg, 1972]):

$$\frac{d^2\xi^\mu}{d\tau^2} = 0 \qquad \xi^\mu = (ct, x, y, z), \tag{1.3}$$

where τ is the invariant proper time. Therefore the test particle does not have any acceleration. This results can be also read saying that test particles will follow geodetics of a given spacetime metric.
In the Minkovski framework, the spacetime interval ds reads:

$$ds^2 = c^2 d\tau^2 = \eta_{\alpha\beta}\, d\xi^\alpha d\xi^\beta, \tag{1.4}$$

where $\eta_{\alpha\beta}$ is a diagonal matrix with the signature $(1, -1, -1, -1)$. It is possible to perform a transformation to some other coordinates x^μ:

$$d\xi^\mu = \frac{\delta\xi^\mu}{\delta x^\nu} dx^\nu. \tag{1.5}$$

We can insert this transformation in Eqs. (1.3) and (1.4) to derive the equations of dynamics in GR.

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0 \tag{1.6}$$

$$c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{1.7}$$

where the new quantities are:

$$\Gamma^\mu_{\alpha\beta} = \frac{\delta x^\mu}{\delta\xi^\nu}\frac{\delta^2\xi^\nu}{\delta x^\alpha \delta x^\beta}, \tag{1.8}$$

$$g_{\mu\nu} = \frac{\delta\xi^\alpha}{\delta x^\mu}\frac{\delta\xi^\beta}{\delta x^\nu}\eta_{\alpha\beta}. \tag{1.9}$$

The first one is the Christoffel symbol. The second one is the metric tensor which is symmetric and contains at most 10 independent components. The Christoffel symbols can be defined only in terms of the metric tensor, which is a crucial object that can define the structure of the spacetime and the motion of particles.
In order to find a covariant field equation, according to the *principle of General Covariance*, we have to construct a tensor starting from $g_{\mu\nu}$ up to its second partial derivatives. The simplest choice is the *Einstein tensor* $G_{\mu\nu}$, that contains the contractions of the *Riemann tensor*, the Ricci tensor $R_{\mu\nu}$, and the Ricci scalar R:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}. \tag{1.10}$$

This tensor is symmetric and covariantly conserved, thus it contains only 6 independent components. If the Einstein tensor determines the metric and how the particle moves, we have to find a tensor with the same mathematical properties that links the amount of energy and mass to the gravitational field. One such tensor is the *Energy-momentum tensor*, which for a perfect fluid with four-velocity u^μ, density ρ and pressure p (measured by an observer comoving with the fluid) is given by

$$T^{\mu\nu} = (p + \rho)u^\mu u^\nu + pg^{\mu\nu}. \tag{1.11}$$

Now, equalizing Eqs. (1.10) and (1.11), we can finally write the *Einstein field equation.*

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu}. \tag{1.12}$$

The above fundamental equation provides an interpretation of gravity as the geometry of the spacetime. Matter and energy are accounted into the right hand side of the equation and determine the curvature of the spacetime via the solutions of the metric tensor included into the Riemann tensor contractions on the left side.

1.2 Friedmann–Lemâitre-Robertson–Walker metric

If one tries to model the whole Universe within the GR framework, then will immediately face the problem of spacetime topology and boundary conditions.

In 1917 Albert Einstein [Einstein, 1917] opted for a Universe with a $\mathbb{R} \times S^3$ closed spatial topology with constant positive curvature. This topology was introduced to model a spatially finite Universe without a border, and thus to avoid the problem of unknown boundary conditions.

However to solve the *Einstein field equation* one has to define the metric of the Universe, the simplest way to do that is to start from the Minkovski metric and impose the *Cosmological Principle.*

Starting from Eq. (1.4) with the Minkovski signature, we can define the global time, t, as the time measured by different observers at rest with respect to their surrounding matter distribution. The metric matrix can be diagonalized if we take in consideration the isotropy. Furthermore we can decompose the spatial components into radial and transverse parts using the new coordinates (r, θ, ϕ) with the convention $d\Psi^2 = d\theta^2 + \sin^2\theta d\phi^2$. Finally, we can write distances as the product of a time-dependent function $a(t)$ and a time-independent comoving coordinate r.

$$ds^2 = c^2 d\tau^2 = c^2 dt^2 - a(t)^2[\, f(r)^2 dr^2 + g(r)^2 d\Psi^2], \tag{1.13}$$

where $f(r)$ and $g(r)$ are arbitrary functions. If we consider the metric on the surface with constant curvature, like Einstein suggested, we can get a very elegant functional form of the previous equation. Because the curvature must be constant, only three different geometries are possible: *Euclidean* if there is no curvature, *elliptical* for positive curvature, *hyperbolic* otherwise.

We can generalize this argumentation on a higher-dimensions space, for instance a 3-sphere. Furthermore we can write a functional form which embeds the three different geometries using a parameter k. This metric is known as the *Friedmann Lemâitre Robertson Walker metric* (FLRW).

$$ds^2 = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 d\Psi^2 \right), \tag{1.14}$$

where $k = +1$ is for positive curvature, $k = -1$ for negative one, $k = 0$ for flat geometry and $a(t)$ is the *scale factor* that accounts for the expansion of the Universe.

1.3 Distances

In Cosmology there are many ways to define the distance between two points. This is due to the expansion of the Universe, because of which the distance between any two objects is constantly changing. Every kind of cosmological distance measures the separation between events on trajectories of photons which terminate at the observer. The most intuitive distance definition, called the *proper distance*, can be derived starting from Eq. (1.14), setting a global proper time so that $dt = 0$ and a reference frame with $d\Psi = 0$:

$$d_{PR} = \int_0^r ds = a(t) \int_0^r \frac{dr}{\sqrt{1-kr^2}} = a(t) f(r), \tag{1.15}$$

where the function $f(r)$ depends on the parameter k. More specifically:

$$f(r) = \begin{cases} \sin^{-1} r, & \text{if } k = +1 \\ r, & \text{if } k = 0 \\ \sinh^{-1} r, & \text{if } k = -1 \end{cases} . \tag{1.16}$$

This kind of distance refers to the global proper time, thus it does not account for the expansion of the Universe and changes over time following the evolution of the scale parameter. If we carry off the scale factor, we can define another kind of distance, the *comoving distance*. This measure does not change, if we neglect the object peculiar velocities, i.e. it takes into account the Universe expansion. The comoving distance can be derived by calculating the proper distance at the present time t_0 (or the present scale factor $a(t_0) = a_0$) [1].

$$d_C = a(t_0) f(r) = a_0 f(r) = \frac{a_0}{a(t)} d_{PR}. \tag{1.17}$$

The comoving distance is time-dependent. It is possible to compute its variation in time as follows:

$$\frac{d(d_C)}{dt} = v_R = \dot{a}(t) f(r) \frac{a(t)}{a(t)} = \frac{\dot{a}(t)}{a(t)} d_{PR}. \tag{1.18}$$

---

[1] In what follows, the subscript 0 refers to present time

This important result states that the expansion of the Universe is such that the proper distance between any two points changes with a rate that depends on the scale factor. Every object moves away from all the others on the so-called *Hubble flow*.

1.4 Redshift and the Hubble-Lemâitre law

Let us consider an electromagnetic waves emitted with a certain wavelength λ. Its source will radially move with respect to the observer, following the Hubble flow. Thus an observer measures a wavelength λ_0 that is different than the original one due to the distortion of spacetime.
The relative difference between the original and the observed wavelengths is called redshift (or blueshift if the receiving wavelength is shorter than the emitting one):

$$z \equiv \frac{\lambda_0 - \lambda}{\lambda}. \tag{1.19}$$

If the source emitted the photon at t_{em}, and it reaches the observer at the time t_0 moving along a geodetic ($ds^2 = 0$), then we can use the FLRW metric to derive the connection between redshift and the scale factor. Without loss of generality, we can again set the reference frame in order to remove angular parameters:

$$ds^2 = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - kr^2} \right) = 0. \tag{1.20}$$

Integrating along the photon path, r, and over the time, t, and using Eq. (1.15) we get:

$$\int_{t_{em}}^{t_0} \frac{c}{a(t)} dt = \int_0^r \frac{dr'}{\sqrt{1 - kr'^2}} = f(r). \tag{1.21}$$

If the source sends a second photon right after the first one, at $t_{em} + \delta t_{em}$, that reaches the observer at $t_0 + \delta t_0$, we can rewrite the previous equation as follows:

$$\int_{t_{em}+\delta t_{em}}^{t_0+\delta t_0} \frac{c}{a(t)} dt = f(r). \tag{1.22}$$

We can also assume that $a(t_0 + \delta t_0) = a(t_0)$ and $a(t_{em} + \delta t_{em}) = a(t_{em})$, which holds for small time intervals.
Now, we can equal the left side of the last two equations. Then using the relations between time, frequency and wavelength, we obtain the redshift definition that depends only on the scale factor:

$$\frac{\delta t_0}{a_0} = \frac{\delta t_{em}}{a(t_{em})}, \tag{1.23}$$

$$z = \frac{a_0}{a(t_{em})} - 1. \tag{1.24}$$

In the first decade of '900 many astronomers measured the redshift of close galaxies. They found that almost all the extragalactic objects are moving away from Earth. Thus all photons that reach our telescopes have a larger wavelength with respect to the time at which they were emitted. This was the first hint of an expanding Universe. In fact the scale factor in Eq. (1.24) grows with t, thus the Universe is expanding. It was also possible to constrain the velocity of this expansion, starting from the comoving distance definition given by Eq. (1.18). The expansion of the Universe is described by the well-known *Hubble-Lemâitre Law* [Hubble, 1929].

$$v = H(t)r, \tag{1.25}$$

where

$$H(t) \equiv \frac{\dot{a}(t)}{a(t)} \tag{1.26}$$

is known as the *Hubble parameter* or, for $t = t_0$, the *Hubble constant*, H_0. This law tells us that the expansion of the Universe is such that every object is moving away from a generic observer with a velocity proportional to its distance.

The value of the Hubble constant is not yet precisely known, and there are discrepancies from the results coming from different independent methods as it can be seen in Fig. 1.1. One of the last empirical results, coming from the ESA Planck mission 2018 [Planck Collaboration et al., 2018], measured the following value of the Hubble constant from the cosmic microwave background radiation:

$$H_0 = 67.4 \pm 0.5 \ km \ s^{-1} \ Mpc^{-1}. \tag{1.27}$$

The value of the Hubble constant can also be expressed using the dimensionless parameter h:

$$H_0 = 100 \ h \ km \ s^{-1} \ Mpc^{-1}. \tag{1.28}$$

H_0 has units of s^{-1}. Assuming that the Universe had always the same expansion rate we can use the inverse of the Hubble constant as an estimator of the age of the Universe.

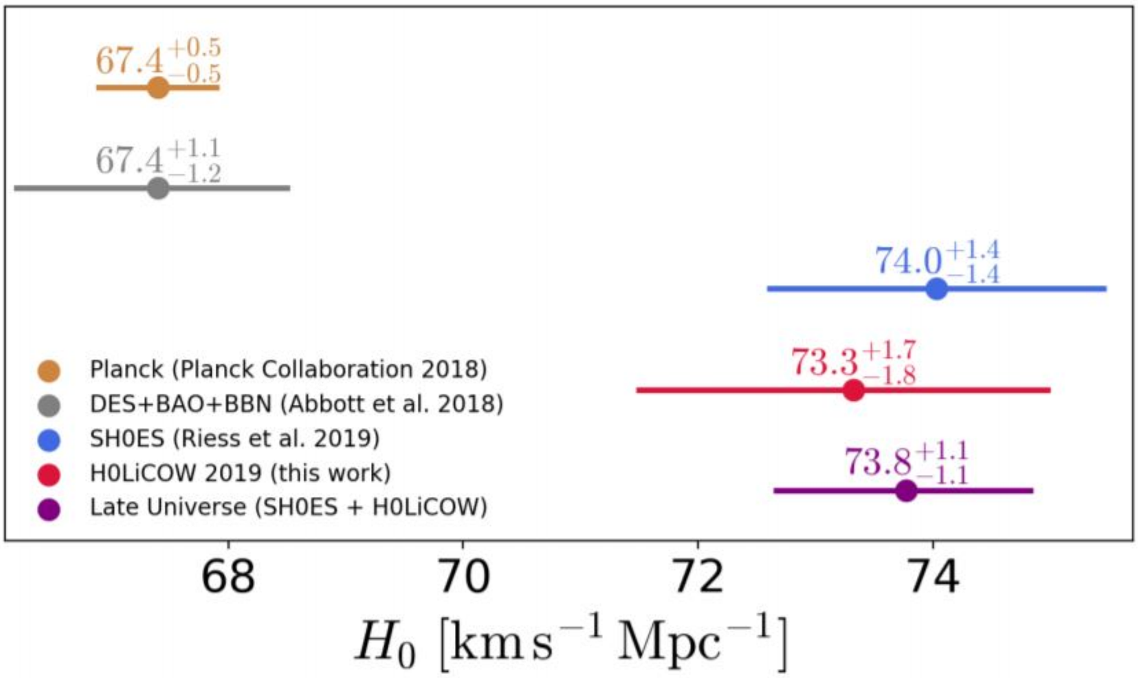


Figura 1.1: Comparison of H_0 constraints for early-Universe and late-Universe probes in a flat ΛCDM cosmology [Wong et al., 2019]. The early-Universe probes shown are from CMB [Planck Collaboration et al., 2018] and a combination of clustering, weak lensing data, BAO, and big bang nucleosynThesis [Abbott et al., 2018]. The late-Universe probes shown are the latest results from SH0ES [Riess et al., 2019], obtained from Large Magellanic Cloud Cepheid standards, and H0LiCOW [Wong et al., 2019] from lensed quasar. When combining the late-Universe probes (purple), we find a 5.3σ tension with Planck [Wong et al., 2019].

1.5 Friedmann equations

Under the assumption that the Universe is a perfect fluid of matter and energy, it is possible to use the isotropic and homogeneous metric (1.14) into the equation of GR (1.12). The results are two non-null equations called *Friedmann equations* [Friedmann, 1922].

$$\ddot{a}(t) = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right), \tag{1.29}$$

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

These two differential equations describe the time evolution of the scale factor as a function of the density and pressure of the cosmic fluid and the curvature parameter

k. The Universe is indeed a closed system, thus it is expanding without energy loss. Therefore, it is possible to impose the adiabatic condition, that can also be used to derive one Friedmann equation starting from the other:

$$dU = -pdV \longrightarrow d(\rho c^2 a(t)^3) = -pd^3a. \tag{1.31}$$

From the Friedmann equations, two useful quantities can be defined:

$$q \equiv -\frac{a(t)\ddot{a}(t)}{\dot{a}(t)^2} \qquad \textit{deceleration parameter}, \tag{1.32}$$

$$\Omega \equiv \frac{8\pi G}{3H(t)^2}\rho \equiv \frac{\rho}{\rho_c} \qquad \textit{density parameter}, \tag{1.33}$$

where ρ_c is called the *critical density.*

$$\rho_c = \frac{3H(t)^2}{8\pi G}. \tag{1.34}$$

Thus the second Friedmann equation (1.30) can be rewritten as:

$$\Omega - 1 = \frac{k}{H(t)^2 a(t)^2} \tag{1.35}$$

where the fraction sign depends on the ratio between the density of the Universe and the critical density (1.34). Using the last two equations and recalling the only three geometries of our Universe, the following cases are possible:

$$\begin{cases} \rho > \rho_c, & \Omega > 1, \quad k = +1 \quad \text{Spherical/Closed Universe} \\ \rho = \rho_c, & \Omega = 1, \quad k = 0 \qquad \text{Euclidean/Flat Universe} \\ \rho < \rho_c, & \Omega < 1, \quad k = -1 \quad \text{Hyperbolic/Open Universe} \end{cases} \tag{1.36}$$

1.6 Cosmic fluids

The Friedmann equations have been obtained under the assumption that the matter/energy content of our Universe can be approximated as a perfect fluid. Let us assume that this fluid is composed of different components, each of them satisfies the equation of state

$$p = w\rho c^2, \tag{1.37}$$

where w is a constant that is specific for every different components. This fluid must obviously satisfies the adiabatic condition (1.31). Hence, using the equation of state, we get:

$$\frac{\dot{\rho}}{\rho} = -3(1+\omega)\frac{\dot{a}(t)}{a(t)} \longrightarrow \frac{\rho}{\rho_0} = \left(\frac{a(t)}{a(t_0)}\right)^{-3(\omega+1)}. \tag{1.38}$$

Thus if there are different components in the cosmic fluid, their densities will evolve with the scale factor in different ways, namely:

$$\begin{cases} w = 0, & \text{matter} \quad \rho(t) \propto a(t)^{-3} \\ w = \frac{1}{3}, & \text{radiation} \quad \rho(t) \propto a(t)^{-4} \end{cases} \tag{1.39}$$

The evolution of matter and radiation densities can be easily understood if we notice that for the matter component the density scales with the volume element, while for the radiation component the density scales with volume and with redshift.

1.7 Cosmological constant

We have seen that the *Cosmological principle* is the starting point of all cosmological studies. However, to a few scientists it seemed that this principle was incomplete. In fact it allowed the Universe to be invariant under observer position changes, but it does not say anything about the temporal dimension. Thus Bondi, Gold and Hoyle [Bondi and Gold, 1948, Hoyle, 1948] proposed a Universe homogeneous and isotropic both in space and time as the basic principle of their *steady-state theory*. This principle, called *Perfect Cosmological principle*, has been proved to be wrong. However it had a lot of attraction on scientists who did not want to abandon the idea of a static Universe. According to this principle the large scale structure of the Universe must essentially be the same for all observers, regardless of their spacetime location. In mathematical terms, the first application of the Perfect cosmological principle was that the scale factor $a(t)$ must be constant with time, so $\ddot{a}(t) = \dot{a}(t) = 0$.

If these equivalences hold, then the Friedmann equations can be used to derive the physical quantities (pressure and density) of the fluid that composes this static Universe. Using Eq. (1.29) it can be easily seen that the static hypoThesis leads to a non physical condition:

$$\rho = -\frac{3p}{c^2}. \tag{1.40}$$

In order to solve this problem, and save the steady-state theory Albert Einstein changed his General Relativity field equation (1.12) in the simplest possible way. He added a constant term, Λ, called the *Cosmological Constant* [Einstein, 1931]:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T^{\mu\nu}. \tag{1.41}$$

This energy term is a source of repulsive gravity and can be seen as the curvature of empty space. If we move the Cosmological Constant on the right, inside the energy-momentum tensor, is it possible to give another interpretation of this component as representation of the vacuum matter/energy content. Therefore it can be derived the contribution of this constant to pressure and density of the cosmic fluid:

$$\tilde{p} = p - \frac{\Lambda c^4}{8\pi G}, \tag{1.42}$$

$$\tilde{\rho} = \rho + \frac{\Lambda c^2}{8\pi G}. \tag{1.43}$$

From these equations the value of ω in the equation of state can be carried out:

$$\tilde{p} = -c^2\tilde{\rho} \longrightarrow \omega = -1. \tag{1.44}$$

Hence the Cosmological Constant can be seen as a new component of the Universe, called *Dark Energy*. Recalling Eq. (1.39), for Dark Energy holds:

$$w = -1 \rightarrow \rho(t) \propto const \tag{1.45}$$

1.7.1 The Einstein' Universe

With the new description of the gravitational field, Einstein proposed a model for Universe filled only with matter, i.e. $w = 0$ and $\tilde{p} = -\frac{\Lambda c^4}{8\pi G}$. Let us insert the new definitions of pressure and density into Eq. (1.40) that describes a static Universe:

$$\tilde{\rho} = -\frac{3\tilde{p}}{c^2} = \frac{3\Lambda c^2}{8\pi G}. \tag{1.46}$$

Now we can use the second Friedmann Eq. (1.30) setting $\dot{a}(t) = 0$:

$$\tilde{\rho} = \rho + \frac{\Lambda c^2}{8\pi G} = \frac{kc^2}{8\pi G a(t)^2}. \tag{1.47}$$

The density of ordinary matter must be positive for definition, thus $k = +1$ in this model, the so-called *Einstein Universe* with the topology described in Section (1.2). Finally, applying the equation of state, we get:

$$\tilde{p} = -\frac{\Lambda c^4}{8\pi G} = -\frac{kc^4}{8\pi G a(t)^2} = -c^2\tilde{\rho}, \tag{1.48}$$

$$\Lambda = \frac{k}{a(t)^2} = a(t)^{-2}. \tag{1.49}$$

This elegant and simple model has been abandoned after the Hubble's discovery of the expansion of the Universe. After that, Einstein changes his point of view, declaring that the Cosmological Constant was no longer necessary and theoretically unsatisfactory. Nevertheless the Cosmological Constant did not disappear and was reintroduced in modern cosmological theories after the discover of the accelerated expansion of the Universe made with observation of supernovae 1a as standard candles (see Riess et al. [1998] and Perlmutter et al. [1999]). In fact, its one of fundamental building blocks of the ΛCDM model that is named for its main component, Dark Energy, Λ, and Cold Dark Matter, CDM.

1.7.2 De Sitter Universe

To better understand the contribute of the Cosmological Constant to the expansion of the Universe, it can be useful to derive the scale factor behaviour in a flat and empty Universe [de Sitter, 1917], i.e. $k = 0$, $\rho = 0 \rightarrow \tilde{\rho} = \frac{\Lambda c^2}{8\pi G}$. So, again, let us start from the second Friedmann equation, Eq. (1.30), with the above hypotheses:

$$\dot{a(t)}^2 + \cancel{kc^2} = \frac{8\pi G}{3}\tilde{\rho}a(t)^2 = \frac{\Lambda}{3}a(t)^2c^2. \tag{1.50}$$

We can solve this differential equation, obtaining:

$$a(t) = A\exp\left(\sqrt{\frac{\Lambda}{3}}ct\right). \tag{1.51}$$

Therefore, if the Universe was empty and flat, it will still experience an exponential expansion driven by the Cosmological Constant and its repulsive action.

1.8 Friedmann model

The term *Friedmann model* is used to indicate a cosmological model for a matter-only Universe, through its equations can be easily extended for multi-components Universes. In order to solve the Friedmann equations, let us combine the Hubble law and Eq. (1.30). Then we can generalize the density parameter as the sum of different components contributions:

$$H(t)^2 = H_0^2\left(\frac{a(t_0)}{a(t)}\right)^2\left[1 - \sum_i \Omega_i + \sum_i \Omega_{0,w_i}\left(\frac{a(t_0)}{a(t)}\right)^{1+3w_i}\right], \tag{1.52}$$

where the sum is performed over the different components, each one with a state equation parameter w_i. The term $1 - \sum_i \Omega_i \equiv \Omega_k$ is known as the *curvature density parameter*. From Eq. (1.35) it is possible to see that the sign of the curvature density parameter is concordant with the one of k. Eq. (1.52) is a crucial equation that links redshift,

comoving distances and densities of the different components of the Universe.
Now, using Eq. (1.35) we can derive the redshift dependence of the density parameters. Let us assume a single component for simplicity:

$$\Omega(z)_w^{-1} - 1 = \frac{\Omega_{0,w}^{-1} - 1}{(1+z)^{1+3w}}. \tag{1.53}$$

This equation tells us that, since z grows from now back to the Big Bang, the density parameter will tend to unity regardless of its actual value, thus the geometry of early the Universe is fixed.
This statement is not valid for density of dark energy, that does not change during the Universe evolution.

1.8.1 Einstein-de Sitter Universe

By the early 1930s, after the Hubble's discovery, it has been established that the static models of cosmos were not able to reproduce the observations. At the same time de Sitter investigated a model of an empty, hence not physical, Universe.
The starting point for building a model of the Universe consistent with Hubble law and with a physical meaning is through Eq. (1.52), that can be solved analytically in a few cases. The most simple case is the so-called *Einstein-de Sitter Universe* (EdS) [Willem de Sitter, 1932] that was proposed in 1932. This model is valid for a Euclidean Universe ($k = 0$, thus for $\Omega = 1$) with a mono-component cosmic fluid, a flat geometry ($\Omega = 1$) and without the Cosmological Constant. Under these assumptions, we can use Eqs. (1.52) and (1.30) to get the following equation:

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+\omega)}}. \tag{1.54}$$

Starting from this calculation, it is possible to derive the other cosmological quantities that depend only on the w factor:

	Radiation $w = 1/3$	Matter $w = 0$
$a(t)$	$a_0 \left(\frac{t}{t_0}\right)^{1/2}$	$a_0 \left(\frac{t}{t_0}\right)^{2/3}$
$\rho(t)$	$\frac{3}{32\pi G t^2}$	$\frac{1}{6\pi G t^2}$
$H(t)$	$H_0(1+z)^2$	$H_0(1+z)^{3/2}$
t	$t_0(1+z)^{-2}$	$t_0(1+z)^{-3/2}$
t_0	$\frac{1}{2H_0}$	$\frac{2}{3H_0}$
q	1	$\frac{1}{2}$
R_h	$2ct$	$3ct$

Tabella 1.1: Physical quantity trends for EdS Universe

The Einstein-de Sitter model is a fundamental framework as it can accurately describe the Universe in its first phases. Using Eq. (1.53) we can deduce that the early Universe was almost flat. At the same time, the density trend with scale factor, described in Eq. (1.38), and the currently known values of the density ρ_0 for each components of the Universe can be linked together to obtain the plot shown in Fig 1.2.

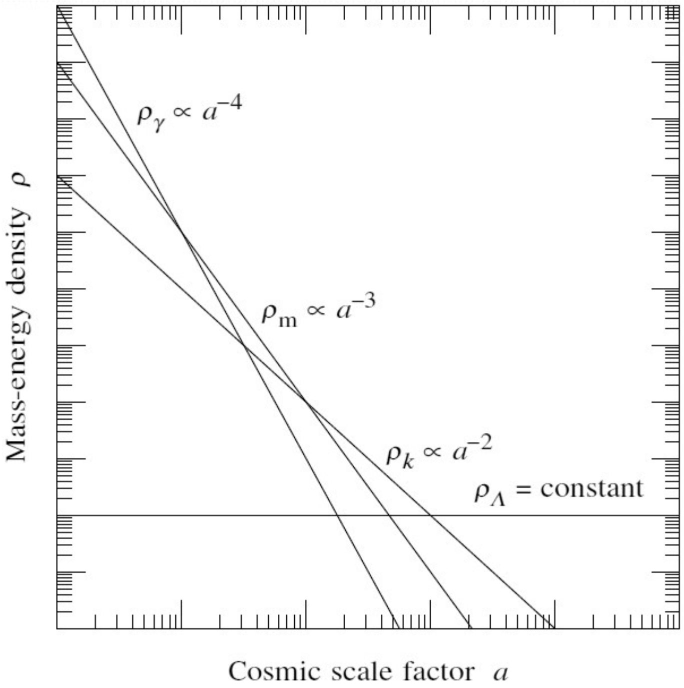


Figura 1.2: The density trend of all the cosmic fluid components, as a function of the scale factor (credits to https://jila.colorado.edu/~ajsh/courses/astr3740_19/evol.html).

Thus the primordial flat Universe was dominated by radiation and can be approximated as a mono-component Universe like the one described by the Einstein-de Sitter model.

1.8.2 Open and Closed Universes

The Friedmann model can be solved also for Open or Closed mono-component Universes. Let start from Eq. (1.52), that can be written in a most compact form as follows:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2\left[\Omega_k + \Omega_0\left(\frac{a_0}{a}\right)^{(1+3w)}\right]. \tag{1.55}$$

The evolution of the scale parameter depends on the competition between the two terms inside the square parenThesis. Let us compare these two terms for a generic scale factor a^*, or redshift z^*:

$$\Omega_k = \Omega_0\left(\frac{a_0}{a^*}\right)^{(1+3w)} \rightarrow \frac{a_0}{a^*} = 1 + z^* = \left(\frac{1-\Omega_0}{\Omega_0}\right)^{(1+3w)}. \tag{1.56}$$

We have already seen that for $z \gg z^*$ the Universe is well described by an EdS model. On the other hand, for $z \ll z^*$, the curvature parameter Ω_k determines the evolution of the scale factor in Eq. (1.55) and the other term inside the parenThesis can be neglected. Thus two different cases are possible, one for Open Universes and one for Closed Universes. For Open Universes, $\Omega_0 < 1$ and so $\Omega_k = 1 - \Omega_0 > 0$, asymptotically we obtain:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2\Omega_k \rightarrow \dot{a} = a_0 H_0 \sqrt{1-\Omega_0} = const. \tag{1.57}$$

Thus the Universe will experience a constant expansion with:

$$a(t) \propto t, \quad H \propto t^{-1}, \quad q \rightarrow 0. \tag{1.58}$$

For Closed Universes, $\Omega_0 > 1$ and $\Omega_k < 0$, hence the previous differential equation admits an undulatory solution. Therefore for Closed Universes the scale factor will reach a maximum at t_{max} and then will return to initial condition after $2t_{max}$. The evolution of the scale parameter in Flat, Open and Closed models are summarized in Fig. 1.3.

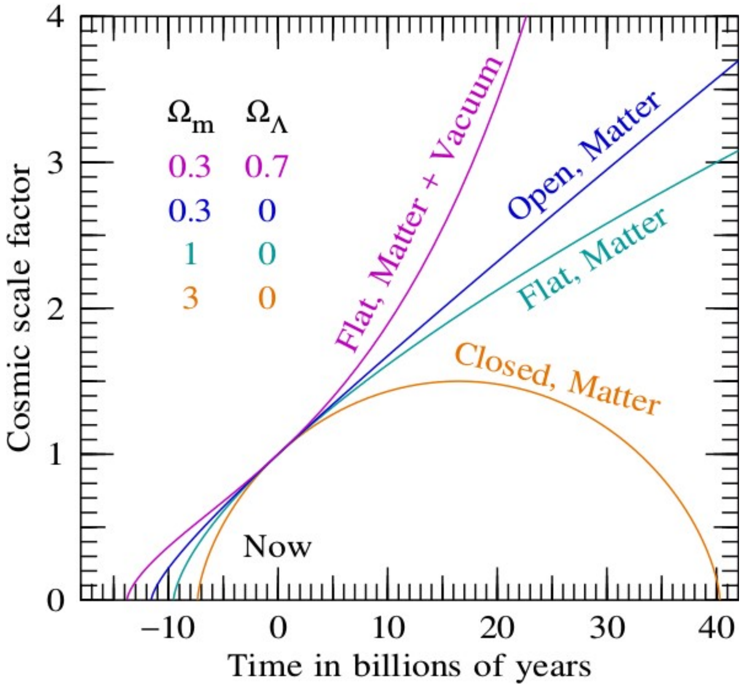


Figura 1.3: Scale factor evolution as a function of time from Big Bang for different cosmological models (credits to https://jila.colorado.edu/~ajsh/courses/astr3740_19/evol.html).

1.9 The concordance cosmological model

The standard model in modern cosmology is the ΛCDM model. Up to now, this model has fulfilled every observational tests, with some tensions (see Sec. 1.4 and Fig. 1.1) and it was also able to predict and describe a wide variety of cosmic phenomena, such as the *Cosmic microwave background* (CMB). The CMB is a relict of low-energy photons generated after the recombination epoch that permeate the observable Universe with a temperature today of $2.726 \pm 0.005\ K$. The starting point of ΛCDM model is the well accepted *Big Bang* theory which predicts that the Universe starts from a singularity on the Planck scale, that rapidly inflated (see [Peebles, 2014]). During the first $\sim 10^{-37}s$, the matter/energy content of the Universe was so high that all the fundamental forces were fused together. As soon as the Universe expanded, it cools down so that the forces separated from each other during the so-called *phase transitions*. Phase transitions are crucial moments of the Universe evolution, as they allowed temporary violations of conservations. For instance, during the first of these phases, at $\sim 10^{-37}s$, the Universe could form baryons breaking the baryon number conservation law. The seeds of baryons are generated at the end of the *Inflationary epoch* [Martin, 2018] from the decay of a hypothetical particle called *Inflaton* with a stochastic process. The early Universe was indeed radiation-dominated, hence the radiation drove the thermalization, and the cosmic fluid was relativistic. Particles with a certain mass m_x could become non-relativistic as soon as the temperature of the Universe reached a value given by the following relation:

$$k_B T = m_x c^2. \tag{1.59}$$

Thus, since the temperature decreased, the lightest particles could detach from the relativistic fluid. One fundamental cosmic time, called the equivalence time t_{eq}, is reached when the radiation density equals the matter density. However, at t_{eq}, the Universe thermalization is still driven by radiation. Hence another fundamental time could be defined, called the decoupling time t_{dec}, at which matter and radiation temperature evolution became unbounded. Within this framework, the ΛCDM model specifies the qualities and quantities of the components of the Universe. According to this model, the Universe is made up of three main components, radiation, matter and Dark Energy. Radiation, which consists of photons and the other relativistic particles such as neutrinos, was dominant during the first phases but nowadays is negligible. Its actual density could be calculated using CMB observation (see [Planck Collaboration et al., 2019]). Using the black-body laws and the density trend for an EdS model we get a value of $\Omega_{0,r} \sim 10^{-5}$. Another fundamental component of the Universe is matter, which can be divided into two different classes, ordinary matter and Dark Matter. The density parameter of ordinary matter, made up of baryons, can be estimated in various ways. One estimation comes from galaxies luminosity function which can be used to constrain Ω_b assuming a certain M/L and ρ_b/ρ_{gal} ratios. Ordinary matter is not the dominant matter component. In fact, starting from the Zwicky seminal work [Zwicky, 1933], astrophysics found out evidence for a new collisionless kind of matter, called *Dark Matter* (DM). DM is a hypothetical typology of matter that interacts only through gravity and weak nuclear force. Thus it can only be detected with dynamical probes, such as galaxies light curves or gravitational lensing [Cacciato et al., 2013, Herrera-Martín et al., 2019, de Salas et al., 2019]. Up to now there are several theories that try to describe the constituents and the nature of DM. Overall there are two theoretical classes for the hypothetical DM particles:

- **Cold Dark Matter, CDM** DM particles that decoupled from radiation when they were no longer relativistic. An example of this class are the so called weakly interactive massive particles (WIMPs) [Roszkowski et al., 2018].
- **Hot Dark Matter, HDM** DM particles that became non relativistic after t_{dec}. Examples of this class are neutrinos [Boyarsky et al., 2019].

The total matter density parameter can be estimated from CMB data, that suggest a value of $\Omega_m \sim 0.3$. Finally, the main component of the ΛCDM model is the Dark Energy that can be parametrized by the Cosmological Constant. Observational evidences for Dark Energy could be found by standard candle, like SN1a [March et al., 2011], magnitude-redshift relation. Furthermore, other evidences can be found, for instance, by CMB acoustic peaks, [Percival et al., 2010], and X-ray gas fraction in clusters, [Sasaki, 1996]. The effects of this kind of energy are embedded into Eq. (1.29) that, using the equation of state for a mono-component fluid, reads:

$$\ddot{a} = -\frac{4}{3}\pi G(1+3w)a \tag{1.60}$$

Therefore for $w < -\frac{1}{3}$ the Universe will experience a positive acceleration. According to the most recent CMB studies (combining Planck temperature and polarization data with Planck lensing and baryon acoustic oscillation measurements, Planck Collaboration et al. [2018]), our Universe has a flat geometry:

$$\Omega_k = 0.0007 \pm 0.0019. \tag{1.61}$$

Thus, $\Omega_0 = \Omega_{0,r} + \Omega_{0,m} + \Omega_{0,\Lambda} = 1$, hence we have $\Omega_\Lambda \sim 0.7$.

Capitolo 2

History of cosmic structure formation

2.1 Introduction to the Jeans theory

The global properties of our Universe are well described by the ΛCDM model with a flat FLRW metric. Within this framework, the presence of large-scale structures can be explained as the end product of the evolution of small gravitational perturbations.

The first perturbations could start their growth after the decoupling between matter and radiation. Due to its non-collisional nature, CDM first detached from radiation and thus formed seeds inside which baryons fell after decoupling.

The growth of perturbations from matter clouds could be described as the struggle between gravitational collapse and internal pressure (or velocity dispersion) repulsive action. This kind of studies has its roots in the so-called *Jeans theory*, proposed by Sir James Jeans in 1902. The aim of this model is to predict when a cloud of gas collapses, and how the density contrast between the perturbation and the background distribution of matter and energy evolves in time.

The density contrast, or *density perturbation field*, can be defined as follows:

$$\delta(x) \equiv \frac{\rho(x) - <\rho>}{<\rho>}. \tag{2.1}$$

Thus $\delta(x)$ runs from -1 to $+\infty$. The linear regime is for $\delta \ll 1$.

The Jeans theory can be applied to non-relativistic matter, on scales smaller than the horizon one (described in the following Section) and in the linear regime. Moreover this theory can be generalized to an expanding Universe and, under some assumptions, in the quasi-linear regime. Before starting with the proper perturbations' theory, we have to introduce two fundamental scales that will be used in the following discussion:

- *Cosmic horizon*, R_h
- *Jeans scale length*, λ_J.

2.1.1 Cosmic horizon

The cosmic horizon, $R_h(t)$, is defined as the comoving scale length of the region in causal connection with an observer at the given time or, in other words, the path length of the photon emitted immediately after the Big Bang that reaches the observer at a certain time t. Its mathematically formulation can be derived from comoving geodetics of FLRW metric:

$$R_h(t) = a(t)\int_{tBB}^{t}\frac{cdt'}{a(t')} = \begin{cases} R_h \propto a^2 & \text{for } a < a_{eq} \\ R_h \propto a^{3/2} & \text{for } a > a_{eq} \quad . \end{cases} \tag{2.2}$$

This equation tells us that above a certain scale R_h it is possible to neglect every micro-physics effect since there were no time for two different regions of the Universe to be in causal connection. Therefore, over the cosmic horizon, the only relevant force is gravity which leads every perturbation of the dominant component to collapse.
It is possible to derive the previous statement handling a given perturbation like a closed Universe inside a flat one. The two Universes must satisfy the second Friedmann equation:

$$H_p^2 + \frac{c^2}{a^2} - \frac{8}{3}\pi G\rho_p = 0 \quad \text{perturbation, k=+1,} \tag{2.3}$$

$$H_b^2 - \frac{8}{3}\pi G\rho_b = 0 \quad \text{background, k=0.} \tag{2.4}$$

The Hubble parameter in the two equations must be the same. Hence, from the density evolution for the background one, Eq. (1.38), it is possible to derive the following relation:

$$\delta \propto a^{1+3w} \rightarrow \begin{cases} w = 0, & t > t_{EQ} \quad \delta \propto a^{2/3} \\ w = \frac{1}{3}, & t < t_{EQ} \quad \delta \propto a^2 \quad . \end{cases} \tag{2.5}$$

Beyond the cosmic horizon, micro-physics effects are negligible, thus all the components will follow the dominant one. So it can be concluded that every perturbation will grow.

2.1.2 Jeans scale length

The second fundamental scale in Jeans theory is the so-called Jeans length. It is defined as the scale length beyond which a generic perturbation can collapse. Let us consider a perturbation with a scale λ, pressure P and mean density ρ. The internal pressure will act against the gravity of the perturbation, with an acceleration given by:

$$a_P = \frac{P\lambda^2}{M} \sim \frac{v_s^2}{\lambda}, \tag{2.6}$$

where v_s is the sound velocity, calculated under the assumption of constant entropy:

$$v_s^2 \sim \frac{P}{\rho} \tag{2.7}$$

At the same time, the gravitational force will compress the cloud and trigger the collapse, with the following acceleration:

$$a_g = \frac{GM}{\lambda^2} = G\rho\lambda. \tag{2.8}$$

If we equalized the two accelerations, we get an estimator of the Jeans length:

$$\lambda_J = \frac{v_s}{\sqrt{G\rho}}. \tag{2.9}$$

When $\lambda > \lambda_J$, gravity wins the competition against internal pressure and the perturbation collapses. On the other hand, for $\lambda < \lambda_J$ the internal pressure causes the evaporation of the perturbation. For DM, the process that lead to evaporation is, instead, the velocity dispersion.

2.1.3 Growth of perturbations in an expanding Universe

The evolution of cosmological perturbations is described starting from the fundamental equations governing non relativistic and adiabatic fluid motion:

$$\frac{D\rho}{Dt} = -\rho\nabla \cdot v \quad \text{Continuity equation,} \tag{2.10}$$

$$\frac{Dv}{Dt} = -\frac{\nabla p}{\rho} - \nabla\Phi \quad \text{Euler equation,} \tag{2.11}$$

$$\nabla^2\Phi = 4\pi G\rho \quad \text{Poisson equation,} \tag{2.12}$$

where the first two equations describe the density and momentum conservation, while the Poisson equation links gravitational potential, Φ, and density. D/Dt is the convective derivate. These equations will be treated within the small perturbation linear theory. Let us assume a known steady state solution, $s_b = (\rho_b, p_b, \Phi_b, v_b = 0)$, to which we apply a small perturbation, $\delta s = (\delta\rho, \delta p, \delta\Phi, \delta v)$. We can put both the solution and the perturbation in Eqs. (2.10), (2.11), (2.12). Under the assumption of small perturbations, $|\delta s| \ll 1$, we can neglect the second-order, and higher, terms. The results are conventionally expressed in wave form in Fourier space, and using Eq. (2.1) instead of density:

$$\delta(x,t) = \sum_k \delta(k,t) \exp{(ikx)}. \tag{2.13}$$

To take into account the expansion of the Universe, we have to set the unperturbed velocity, to the Hubble flow, $\delta v = Hr$, and change the coordinate space, from proper coordinates to comoving ones. After some calculations, a single equation can be obtained:

$$\ddot{\delta}(k) + \dot{\delta}(k)\frac{2\dot{a}}{a} + \delta(k)\left(\frac{k^2 v_s^2}{a^2} - 4\pi G\rho_b\right) = 0. \tag{2.14}$$

This second-order differential equation holds for every kind of perturbation with scale length below the cosmic horizon, and takes into consideration the Hubble flow. The first-order time derivative terms represent the tidal cosmological effect, while the terms inside parentheses are linked to pressure, via v_s^2, and to the background density gravitational field. Eq. (2.14) is a fundamental relation to derive to evolution of every kind of perturbations in different cosmological epochs.

2.1.4 Dissipation processes

Beyond the cosmic horizon the perturbations will always grow. Below the horizon scale their evolution is given by Eq. (2.14), which does not take into account the micro-physics processes that tend to spread the over-densities. For baryons, the dissipation process is called the *diffusion damping*, and it is originated when photons decoupled from matter. Photons diffuse from overdense to underdense regions, carrying baryons with them and thus leading to a dissipation process. If the mean free path of photons is higher than the perturbation scale length then the photons can carry away baryons and then the overdensity will not grow. In terms of mass, the so called *Silk mass* [Silk, 1969] can be defined which reaches a maximum at the decoupling time where $M_{Silk}(t_{dec}) \sim 10^{12} M_\odot$, while the Jeans mass at the same time is $M_J(t_{dec}) \sim 10^{16} M_\odot$. Thus every perturbation with a mass between these values will oscillate and can not grow until decoupling. While, on scale smaller than $M_{Silk}(t_{dec})$ the perturbations are completely cancelled.

For DM particles there is a similar process, which however is not a dissipation, due to the non-collisionally nature of DM, called *free streaming*. DM overdensities are damped up to a length scale λ_{fs}:

$$\lambda_{fs}(t) = a(t)\int_{t_{BB}}^{t} \frac{v(t')dt'}{a(t')}. \tag{2.15}$$

The functional form of λ_{fs} is very similar to the cosmic horizon definition, where $v(t)$ is the velocity of the perturbation oscillations. In term of mass, the free streaming mass reaches a maximum at the equivalence time, that is comparable with the Jeans mass for DM at the same time, thus every perturbation greater than these values will grow freely, while perturbations on smaller scales are cancelled.

2.1.5 Jeans theory solutions

In order to describe the growth of perturbations, we have to modify and solve Eq. (2.14) for the different components of the Universe.

Radiation

First of all we consider radiation, which is the dominant components for $t < t_{eq}$, thus every other element follows its evolution. If we assume an EdS model, that is a perfect assumption for early epochs (see Eq. (1.53)), we can compare the cosmic horizon with the Jeans scale length using Eqs. (2.2) and 1.1.

$$R_h = 2ct \quad , \qquad \lambda_J \sim \frac{v_s}{\sqrt{G\rho}} \sim \sqrt{\frac{32\pi}{9}}ct. \tag{2.16}$$

Hence,

$$\lambda_J > R_h. \tag{2.17}$$

Therefore only perturbations with scales larger than the cosmic horizon can collapse and grow. Hence every radiation perturbation inside the cosmic horizon can only oscillate and due to the high velocity, on the average we have $\delta \sim 0$.

Dark matter

Dark matter perturbations start growing before t_{eq}, in a radiation dominated Universe. Solving Eq. (2.14) before the equivalence time inside the horizon, we get the following results:

$$\delta_{DM}(k) = 1 + \frac{3}{2}\frac{a}{a(t_{eq})}. \tag{2.18}$$

This result can be better understood if we compare the density constant parameter of the DM perturbations at the equivalence time with the same parameter calculated when the perturbation crosses the cosmic horizon:

$$\frac{\delta(t_{eq})}{\delta(t_h)} = \frac{1+\frac{3}{2}}{1+\frac{3}{2}\frac{a(t_h)}{a(t_{eq})}}. \tag{2.19}$$

Therefore DM perturbations can grow by a factor $\frac{5}{2}$ at most since they entered into the cosmic horizon. This effect is called *stagnation*, or *Meszaros* effect [Meszaros, 1974].

Baryonic matter

Baryonic matter perturbations can grow only after t_{dec}. Thus they found gravitational holes already generated by DM particles. The solution of the Jeans equation for this phase takes into account these phenomena, thus baryonic perturbations will follow DM perturbations. This process is known as *baryon catch-up*. Solving Eq. (2.14) in a EdS Universe, we get the following:

$$\delta_b(k) = \delta_{DM}(k)\left(1 - \frac{a(t_{dec})}{a}\right). \tag{2.20}$$

Total matter

When baryons and DM perturbations are bounded together in the same potential holes, we can describe the evolution of total matter perturbation assuming an EdS model, at least for high redshift, and solving Eq. (2.14) looking for functional forms as $\delta(k) = t^{\alpha}$. There will be different set of solutions for perturbations scale length larger or smaller than the Jeans one. For smaller scales the overdensities cannot grow, and thus oscillate (α is a complex number). For scale $\lambda \gg \lambda_J$ there are two solutions, one increasing and one decreasing in time:

$$\delta(k) = A\delta_+ + B\delta_- \rightarrow \begin{cases} \delta_- \propto t^{-1} \propto a^{-3/2} & \text{decreasing} \\ \delta_+ \propto t^{2/3} \propto a & \text{increasing} \end{cases}, \tag{2.21}$$

where A and B are integration constants. Since the decaying solution does not lead to growth of gravitational instabilities, the only relevant solution is the increasing one. For a generic Universe, the functional form of the increasing solution is given by the following integral:

$$\delta_+ = H(z)\int_z^{\infty} \frac{(1+z')dz'}{H^3(z')} \tag{2.22}$$

2.2 Non-linear models

The previous Section focused on the evolution of perturbations in linear regime, $\delta \ll 1$. At the present time, the Universe is filled with perturbations with $\delta \sim 10^2$ and higher: thus to properly describe the evolution of the cosmic structure at late times we have to develop a non-linear evolution theory of density perturbations. A significant step in this direction was done by Yakov Zeldovich, who proposed a kinematic approach to extend linear theory in a quasi-linear regime. In his work, Zeldovich assumed that particles move along the Hubble flow, with some perturbations consisting of a time-independent displacement field scaled with a time-dependent function. Thus, using linear theory, he linked this displacement field with the so-called *Strain tensor*:

$$\rho(r,t) = \frac{<\rho>}{a^3}(1-\delta_+\lambda_1)^{-1}(1-\delta_+\lambda_2)^{-1}(1-\delta_+\lambda_3)^{-1}, \tag{2.23}$$

where λ_i are the eigenvalues of the strain tensor, and δ_+ is the time-dependent factor that scales the displacement field found within the linear theory model. The Zeldovich model is exact in one dimension, and it is a very important tool to generate initial conditions of N-body simulations. Another fundamental theory to describe the non-linear evolution is the so-called *Spherical model*, that describes an overdense sphere, with initial comoving velocity $v_i = 0$, as a closed Universe inside a flat one. Within this framework, the perturbation size reaches a maximum at t_{max} with a contrast density of $\delta(t_{max}) \sim 4.6$, that is obviously different from the value that would be obtained in the linear regime $\delta_{lin}(t_{max}) \sim 1.07$. The time-dependent value of ρ_p is given by (see [Gunn and Gott, 1972]):

$$\rho_p(t_{max}) = \frac{3\pi}{32 G t_{max}^2} \tag{2.24}$$

Then, after this phase called *turnaround*, the perturbation continues its evolution. In linear theory, this sphere will collapse in $t_e = 2t_{max}$, neglecting the effect of pressure that acts against gravity. Using the virial theorem, it is possible to derive that the collapsing sphere virialized at $t_{vir} = 3t_{max}$ after an oscillating phase with a radius $R_{vir} = \frac{1}{2} R_{max}$, thus $\rho_p(t_e) = 8\rho_p(t_{max})$. Then, from the linear theory we would get the following contrast density value, calculated for convention at t_e:

$$\delta_{lin}(t_e) = \delta_{lin}(t_{max}) \left(\frac{t_e}{t_{max}}\right)^{\frac{2}{3}} = 1.686, \tag{2.25}$$

while from the spherical collapse theory, we get:

$$\delta(t_e) = \frac{\rho_p(t_e)}{\rho(t_e)} - 1 = 8\frac{\rho_p(t_e)}{\rho(t_e)} \left(\frac{t_e}{t_{max}}\right)^2 - 1 \approx 179. \tag{2.26}$$

2.2.1 N-body simulations

The exact solution of the equations of non-linear perturbations can not be obtained analytically. Thus scientists implemented numerical algorithms to follow the evolution of simulated perturbations. Usually the density field is decomposed into a set of discrete particles, so that these simulations are known as *N-body simulations.* During the past decades the calculators evolved dramatically, and many different implementations of N-body simulations have been proposed and evaluate. In what follows, some different N-body simulation algorithms will be presented, that takes into account only gravitational interactions.

The simplest approach is called *direct gravitational* N-body simulations. Within this method at every time-step the gravitational field of all the particles is solved, exactly. Then, using the equation of motion, the positions and velocities of all particles, at the

following time-step, are calculated. This approach has a computational complexity of $\mathcal{O}(N^2)$ because for each particles must be evaluated the value of the gravitational acceleration produced by the other $(N-1)$ particles.
The problem of solving the Poisson's equation can be alternatively addressed using a different, faster method, called *particle mesh*(PM) [Hockney and Eastwood, 1988]. This method decomposes the density field on a regular grid in Fourier space using Fast Fourier transform techniques and then calculates the k-space gravitational potential on the grid. Finally, the inverse Fourier transform is performed on the gravitational potential mode $\delta\Phi_k$ to obtain the real space components of the gravitational force. The main advantage of the Fourier space analysis is that it is possible to solve the Poisson's equation locally, without integrating on the whole real space. In fact the integration is hidden in the Fourier space transform $\delta(k) \propto \int \delta(x) \exp(ikx) d^3x$. The computational complexity of the particle mesh algorithm is $\mathcal{O}(N \log(N))$. This method can be further improved using an adaptive grid instead of a regular one, enhancing the resolution where the density is higher (see [Bryan and Norman, 1998]). The resolution of the PM code is indeed related to the number of particles and the size of the mesh. Increasing the latters makes the algorithm more computationally expensive. In order to alleviate this issue, variations of PM algorithms have been proposed. For example, the so-called *particle-particle-particle mesh* (P3M) [Hockney and Eastwood, 1988] algorithm provides an intermediate approach between PM and direct gravitational methods. In the P3M algorithm the gravitational forces are calculated exactly for near particles, while for distant ones the forces are estimated on a grid using the PM scheme. Another similar method is the so-called *tree code* (TreePM) [Barnes and Hut, 1986], which implements a grid where neighbouring cells are considered individually while distant ones are treated like a single particle. The computational complexity of the two last methods is still $\mathcal{O}(N \log(N))$. The starting point of every N-body simulations is usually provided assuming the Zeldovich approximation in quasi-linear regime, starting from Gaussian initial conditions at high redshift. Another important feature of all the N-body codes is the *softening length* ε, introduced to prevent infinity forces when the distance between two particles approach zero, defined as $F \propto (r+\varepsilon)^{-2}$ (see [Aarseth, 1963]).

Capitolo 3

Clustering

3.1 Statistical analysis of the density field

As described in the previous Sections, the gravitational instabilities generated during the inflation are the building blocks of the large-scale structure that we observe nowadays. Thus a fundamental parameter to model the gravitational perturbations is the dimensionless density contrast defined in Eq. (2.1). In order to describe the current matter distribution a statistical approach is necessary. The first issue in this kind of analysis is the uniqueness and the non-reproducibile nature of the Universe. To solve this problem, let us assume the *Fair Sample principle* which states that it is possible to consider independent and sufficiently large volumes as different realizations of the Universe.

Furthermore, it is convenient to study the statistical properties of the Universe in Fourier space, as it was done for the Jeans model.

3.1.1 Properties of the initial perturbation field

The seeds of the first gravitational perturbations arise at the end of the inflationary epoch. Due to the stochastic nature of quantum fluctuations, the phases of the Fourier modes of the gravitational potential, $d\Phi_k$, associated via Poisson equation to the perturbations of the density field, δ_k, are independent one of the other. Hence, they had a random and uniform distribution in the interval $[0, 2\pi]$. The density perturbations, $\delta(x)$, are the inverse Fourier transform of $\delta(k)$, so they are a superposition of Fourier modes with independent random phases. Therefore, from the central limit theorem of statistics, it follows that the density probability distribution $P(\delta)$ at any point in space is Gaussian (truncated below -1) with a zero mean value:

$$P(\delta) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\delta^2}{2\sigma^2}\right). \tag{3.1}$$

Hence, the fundamental statistical quantity of perturbations' field is the variance $\sigma^2 =< \delta(x)^2 >$, which is the average of the mean values computed in large and independent volumes.

3.1.2 Mass variance

As said, the statistical properties of the initial Gaussian density contrast field $\delta(x)$ are fully described by the variance σ^2. However, it can be very difficult to define the punctual variance for two main reasons. First of all $\delta(x)$ can not be calculated in each point in the Universe because there is a lack of observational probes to define the density in every point. Furthermore, any extragalactic sources, such as galaxies, are biased tracers of the underlying density field. Due to these issues it is convenient to introduce a different quantity, σ_M, called *mass variance*, given by:

$$\sigma_M^2 =< \delta_M^2 >, \tag{3.2}$$

where δ_M is the density contrast filtered on a given scale R, in real space, that corresponds to the mode k in Fourier space. Starting from the mass variance, it is also possible to define the variance of the perturbations' field related to observable probes, like galaxies, δ_{gal}, or DM haloes:

$$\delta_{gal}(x) \equiv \frac{\rho_{gal}(x) - < \rho_{gal} >}{< \rho_{gal} >}. \tag{3.3}$$

We consider galaxies for clarity, but the following discussion can be equally applies to other probes. We can assume that the ratio between the galaxy contrast field and the total mass perturbations' field can be described with a linear function of $\delta(x)$ that is independent of the length scale (as a first approximation). This ratio is known as the *bias factor*:

$$b_{gal} \equiv \frac{\delta}{\delta_{gal}}. \tag{3.4}$$

Another important quantity, linked to the variance, is the two point auto-correlation function of the density field, $\xi(r)$.

$$\xi(r) =< \delta(x)\delta(x+r) >, \tag{3.5}$$

where the angle brackets indicate two averaging operations, one over the normalization volume V, thus at every point x, and the other over each point with distance r from x. Assuming the *Cosmological principle*, this function depends only on the module of r. Let us translate this definition into the Fourier space:

$$\xi(r) = (2\pi)^{-6} \int\int < \delta(k)\delta(k') > \exp\left(ik(x+r) + ik'x\right) dk dk'. \tag{3.6}$$

Therefore, using the properties of the three-dimensional Dirac delta function, we get the following result:

$$\xi(r) = (2\pi)^{-3} \int P(k) \exp{(ikr)} d^3k, \tag{3.7}$$

where $P(k)$ is known as *Power Spectrum*. The last result, that comes from the *Wiener–Khinchin theorem*, states that the power spectrum is the Fourier transform of the correlation function. Hence, these two functions share the same information, one in real space and the other in Fourier space. From the last two equations it is possible to relate the power spectrum with the decomposition of the density perturbations' field:

$$P(k) \propto < |\delta(k)|^2 > . \tag{3.8}$$

Using the Parseval theorem, we can connect the previous definition to the variance of the density perturbations' field in real space:

$$\sigma^2 = \frac{1}{2\pi^2} \int k^2 P(k) dk. \tag{3.9}$$

With these results we can implement the aforementioned filtering operation on $\delta(x)$ to get the mass variance. This operation does not change the statistical properties of $\delta(x)$, and it is usually performed convolving δ with a filter $W(R)(\hat{W}(k,R)$ in Fourier space):

$$\sigma_M^2 = (2\pi)^{-3} \int P(k) \hat{W}^2(k,R) d^3k, \tag{3.10}$$

where the original variance definition, Eq. (3.9), can be recovered if the filtering scale tends to zero. In conclusion, if we assume an initial Gaussian density probability distribution with zero mean for the perturbation field, the only relevant statical quantity is the variance of the field, which is linked via Fourier transform to the power spectrum and its anti-transform, the correlation function. These two fundamental functions provide powerful probes for discriminating among different cosmological models and will be analysed in details in the next Sections.

3.1.3 Primordial power spectrum

In the previous Section we showed that the power spectrum is one of the central quantities in cosmology. For decades scientists tried to model its functional form with increasing accuracy. According to standard inflationary scenarios, the initial power spectrum must not contain any preferred length scale, thus we can assume a featureless power law as a functional form:

$$P(k) = Ak^n, \tag{3.11}$$

where n is known as the *spectral index* and gives information about the balance between large and small scale power, and A is a normalization constant. Let us insert this functional form into Eq. (3.9):

$$\sigma^2 = \frac{A}{2\pi^2} \int k^{2+n} dk \propto k^{n+3}. \tag{3.12}$$

One hint about the initial value of the spectral index comes from the inflationary epoch. At the end of this phase, there were stochastic perturbations of the metric, so $d\Phi_k$ must be a white noise, equals on every length scale, $d\Phi_k \propto k^0$. Hence, using Eq. (3.12), we get:

$$d\Phi_k \propto \frac{GdM}{R} \propto d\rho R^2 \propto \sigma R^2 \propto k^{(n-1)/2}. \tag{3.13}$$

Then, the initial spectral index must be close to 1. This functional form of the primordial power spectrum is known as the *Zeldovich spectrum*, and it is the best choice to represent the initial power spectrum of DM, or the power spectrum of every perturbation that is outside the cosmic horizon. Furthermore, when a density perturbation crosses the horizon it has always the same amplitude because the mass variance at the cosmic horizon is not time-dependent for the Zeldovich spectrum. At the same time, if an overdensity of DM crosses the horizon before the equivalence time, it will experience the stagnation process, described is Sec. 2.1.5. Thus, if we want to model the DM power spectrum at the equivalence time t_{eq}, we have to consider that for scale lengths larger (or smaller wavenumber if we are in Fourier space, recalling that $k \propto R^{-1}$) than the cosmic horizon scale, the initial power spectrum is preserved. For scales smaller than $R_h(t_{eq})$, the DM particles had experienced the stagnation until t_{eq}, while the free streaming process erases perturbations up to a time-dependent scale λ_{FS}, as described in Sec. 2.1.4. This effect can be parametrized by the so-called *Transfer function*, $T(k)$, that takes into account the previous assertions. The amplitude at the equivalence time of the perturbations that suffer from stagnation depends on the ratio between the time of entry in the horizon t_h and initial time t_{in}. At t_h, recalling Eq. (2.2), the following relation holds:

$$a(t_h) \propto R_h^{1/2} \rightarrow a(t_h)^2 \propto k^{-2}. \tag{3.14}$$

Then,

$$\delta(k, t_{eq}) = \delta(k, t_{in}) \frac{a(t_h)}{a(t_{in})}^2 \propto k^{-2}, \tag{3.15}$$

so the power spectrum $P(k) \propto \delta^2$, at the equivalence time, reads:

$$P(k, t_{eq}) \propto P(k, t_{in}) T(k)^2, \tag{3.16}$$

where

$$T(k)^2 = \begin{cases} k^{-4} & \text{for} \quad k > k_h(t_{eq}) \\ k^0 & \text{for} \quad k < k_h(t_{eq}) \ . \end{cases} \tag{3.17}$$

3.1.4 Correlation function

If the power spectrum has a fundamental role in cosmology, its Fourier anti-transform, the correlation function, shares the same importance. In Sec. 3.1.2 we defined the auto-correlation function of the perturbations' density field, see Eq. (3.5). An alternative definition of the *two-point correlation function* for a discrete sample of objects was given by Davis and Peebles [1983]. This function gives the excess probability, with respect to a random distribution, of finding two objects, such as galaxies, voids or DM haloes, in two comoving volumes dV_1 dV_2, at distance r:

$$dP = n_v^2 dV_1 dV_2 (1 + \xi(r)). \tag{3.18}$$

The two-point correlation function can be easily extended to a larger number of points:

$$dP^N = n_v^N (1 + \xi^N(r)) dV_1 ... dV_N \tag{3.19}$$

where ξ^N is the so-called *N-point correlation function*. From the definition of the correlation function in terms of probability, Eq. (3.5), it is clear that ξ must be greater than -1. The N-point correlation function contains more information than the two-point one. Let us consider for instance the three-point correlation function, that can be written in terms of the two-point correlation function in the following way:

$$\xi^3 = \xi(r_{1,2}) + \xi(r_{1,3}) + \xi(r_{2,3}) + \varsigma(r_1, r_2, r_3), \tag{3.20}$$

where $\varsigma(r_1, r_2, r_3)$ is the reduced three-point correlation function, that contains only the excess of probability with respect to the two-point function. The above description of the correlation function can also be extended for continuous density fields of equal mass particles $\rho(x) = mn(x)$. Following this assumption we can rewrite Eq. (3.18) in terms of density field. Then it is possible to calculate the expected number of objects within a certain radius r from a given one:

$$N(< r) = \int_V dP(2|1) = \int_V \frac{\rho(x)}{m} dV_2 + \int_V \xi(r) dV_2 = \frac{4\pi\rho(x) r^3}{3m} + \frac{4\pi\rho(x)}{m} \int_0^r r'^2 \xi(r') dr'. \tag{3.21}$$

The correlation function can also be declined for specific observed objects, such as galaxies, clusters or DM haloes, recalling the bias definition in Eq. (3.4). Let us consider e.g. galaxies, then the following relation holds:

$$\xi_{gal}(r) =< \delta_{gal}(x)\delta_{gal}(x + r) >= b_{gal}^2 < \delta(x)\delta(x + r) >= b_{gal}^2 \xi(r). \tag{3.22}$$

At scales $10 \leq r[\frac{Mpc}{h}] \leq 50$, the observed functional form of the galaxy correlation function is the following:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{\gamma}, \tag{3.23}$$

where the values of the correlation slope is set to $\gamma \sim 1.8$, and of the correlation length, r_0, depend on the selected cosmic tracers (i.e. for galaxies $r_0 \sim 5$Mpc/h). This functional form is not valid in all the scale range $[0, +\infty]$ because, if we consider the total volume of the Universe, then the second term in Eq. (3.21) must tend to zero, so the correlation function must be negative somewhere.

3.2 Observational Cosmology

The theoretical framework built in the previous Section is a powerful tool to constrain model parameters with observations. In particular, power spectrum and correlation functions can be measured from extragalactic surveys and compared with theoretical expectations to test the assumed cosmological model. To solve this task it is essential to take into account empirical effects that modify the observed features, such as *redshift-space effects* or *geometrical distortions*.

3.2.1 Geometrical distortions

Let us consider the comoving separation, r, between two objects at cosmological distances. It can be decomposed into two orthogonal components with respect to the line of sight, $r_\perp$ and $r_\parallel$. The perpendicular distance $r_\perp$ can be computed using the *angular diameter distance* definition [Weinberg, 1972]:

$$r_\perp = d_A = \frac{d_{tC}}{1+z}, \tag{3.24}$$

where d_{tC} is the *transverse comoving distance* that is related to the comoving distance d_C (see Eq. (1.17)) by:

$$d_{tC} = \begin{cases} \frac{c}{H_0\sqrt{1-\Omega_0}} \sinh\left(c^{-1} d_C H_0 \sqrt{1-\Omega_0}\right) & \text{for } \Omega_0 > 1 \\ d_C & \text{for } \Omega_0 = 1 \\ \frac{c}{H_0\sqrt{|1-\Omega_0|}} \sin\left(c^{-1} d_C H_0 \sqrt{|1-\Omega_0|}\right) & \text{for } \Omega_0 < 1 \end{cases} \tag{3.25}$$

We can notice that, for flat Universes, there is no difference between the angular diameter distance and the proper distance, Eq. (1.15). However, from the previous equations we see that $r_\perp$ is linked to the assumed cosmological model. Therefore, if we assume two different cosmological models, referred to the subscripts 1 and 2, the following relation holds:

$$r_{\perp,1} = \frac{d_{A,2}}{d_{A,1}} r_{\perp,2}. \tag{3.26}$$

Analogous calculation can be done for the parallel distance $r_{\parallel}$, which depends on the cosmological model through the Hubble parameter, namely:

$$r_{\parallel,1} = \frac{H_2(z)}{H_1(z)} r_{\parallel,2}. \tag{3.27}$$

Hence, the assumption of a wrong cosmology modifies the measured distance between two objects, or the shape of a resolved celestial body. This effect, called *geometrical distorsion*, can also be used as an empirical test to constrain the geometry of the Universe. In fact, if the real shape is known, it is possible to fine-tuning the cosmological parameters in order to obtain the expected form. This procedure is known as *Alcock-Paczynski test* [Alcock and Paczynski, 1979].

3.2.2 Redshift space distortions

If we want to map the three-dimensional distribution of cosmic objects we have to take into consideration also the so-called *redshift space distortions* (RSDs). This kind of distortions arise from the difference between cosmological/real and observed redshift. Specifically, the observed redshift is related to the real one by the radial velocity $v_{\parallel}$ of the object with respect to the observer, and its empirical random error σ_z:

$$z_{obs} = z_{cosm} + \frac{v_{\parallel}}{c}(1 + z_{cosm}) + \frac{\sigma_z}{c}. \tag{3.28}$$

RSDs arise when we measure distances with the observed redshift instead of the cosmological one, that is often unknown. The peculiar velocity $v_{\parallel}$ that caused the difference between z_{obs} and z_{cosm} is due to different phenomena. On small spatial scales, < 1 Mpc, peculiar velocities are due to the virialized motion of galaxies within DM halos. Therefore, groups and clusters appear stretched along the line of sight, forming the so-called *Fingers of God*. On larger scales, peculiar velocities are due to the streaming of galaxies toward the higher density regions. Hence, the density field becomes squashed along the line of sight. This phenomenon is known as *Kaiser effect*. The RSDs can be theoretically modelled in both linear and non-linear regimes. In linear regime, RSD effects on perturbations' density field are given by the Kaiser formula [Kaiser, 1987].

$$\delta(k)_{RSD} = (1 + f\mu^2)\delta(k), \tag{3.29}$$

where f is the logarithmic derivative of the growth factor with respect to the scale factor $(f = \frac{d\ln(\delta_+)}{d\ln(a)})$ and μ is the cosine of the angle to the line of sight.

Using Eq. (3.8), the RSD effect on the power spectrum reads:

$$P(k)_{RSD} = (1 + f\mu^2)^2 P(k). \tag{3.30}$$

3.2.3 Observed power spectrum and bispectrum

The power spectrum provides a fundamental tool to describe the evolution of perturbations' field, and can be used to constrain cosmological parameter values. The primordial power spectrum, which refers to DM since it is the first component that can collapse, has been described in Sec. 3.1.3 and can be expressed with the following functional form:

$$P(k) = AT(k)^2 k^n. \tag{3.31}$$

The normalization constant A can be expressed using the mass variance σ_8 on a fixed scale of 8 $\frac{\text{Mpc}}{h}$, which is another essential parameter of the cosmological model:

$$\sigma_8^2 = \int P(k)\hat{W}^2(kR = 8\ \text{Mpc}/h)d^3k. \tag{3.32}$$

To compare the initial, theoretical, power spectrum with the observed one, we have to describe how it evolves with cosmic time. The power spectrum, calculated at t_{eq}, will evolve differently at different scale lengths. For every scale k, lower than the scale of the cosmic horizon at the equivalence time $k_h(t_{eq})$, the perturbations can grow freely and their evolution is described by the linear theory. Thus, we expect a self-similar growth governed by the parameter δ_+ presented in Eq. (2.21). On the other hand, Fourier scales larger than $k_h(t_{eq})$ can grow inside the cosmic horizon, then they will reach the non-linear regime and the evolution of the power spectrum can be computed with the methods introduced in Sec. 2.2. Finally, the turning point of the spectrum, which is linked to the cosmic horizon scale, evolves according to Eq. (2.2), hence it will rigidly moves toward lower Fourier scales.

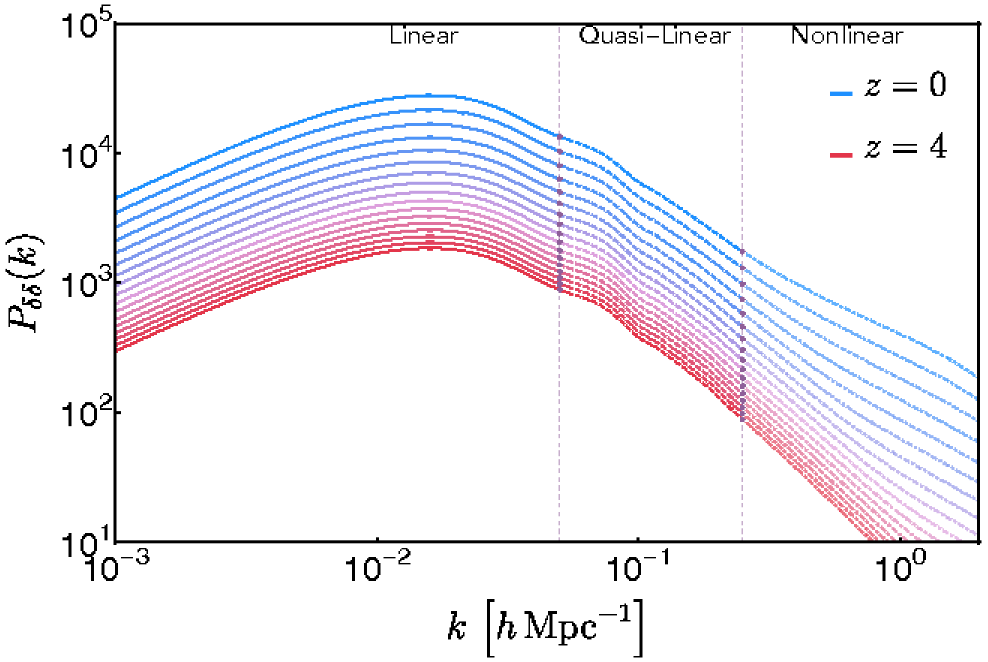


Figura 3.1: Redshift evolution of the matter power spectrum between $z = 0$ (top most line in lower plot) and $z = 4$ (bottom most line in lower plot) [Pratten et al., 2016].

The observed power spectrum will also suffer from RSDs and for selection bias. The RSDs modify the shape of the spectrum according to Eq. (3.30), in linear regime. To study the effects of RSDs, it can be useful to decompose the power spectrum in Legendre polynomial bases. Using the Legendre polynomials, L_l, and starting from Eq. (3.30) we get:

$$P^m(k) = \frac{2l+1}{2}\int_{-1}^{1} P(k,\mu)L_l(\mu)d\mu. \tag{3.33}$$

The shape of the observed power spectrum depends on many factors. For instance, the amplitude A of the spectrum is linked to σ_8, via Eq. (3.32), and the selection bias b, via Eq. (3.4). The evolution of the amplitude is instead linked to the growing factor f, within the linear regime. Hence, some cosmological parameters cannot be constrained from the observed power spectrum, as their effects are degenerate. One way to overcome this issue is to consider also the so-called *bispectrum* $B(k_1, k_2)$, defined as follows:

$$B(k1, k2) \propto < \delta(k_1)\delta(k_2)\delta(k_3) > . \tag{3.34}$$

The bispectrum is the Fourier anti-transformation of the three-point correlation function and contains information about both the primordial and gravitationally induced non-Gaussianity features. As it was done for the power spectrum, also the bispectrum can be decomposed using spherical harmonics for the multipoles' expansion [Scoccimarro et al., 1999]:

$$B_s(k_1, k_2) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} B_s^{l,m}(k_1, k_2) Y_{l,m}(\omega, \varphi), \tag{3.35}$$

where $Y_{l,m}(\omega, \varphi)$ are the spherical harmonics that depends on the polar, ω, and azimuthal, φ, angles with respect to the line of sight. Furthermore, from the bispectrum two terms can be derived, the *reduced bispectrum* and the *shot noise correction*. The reduced bispectrum is the analogue of the reduced three-point correlation function, see Eq. (3.20). Thus it is a term that contains only the information in excess with respect to the power spectrum. Assuming a Poisson sampling, then the shot noise term modifies both the power spectrum and the bispectrum, in two different ways. It adds a white noise component proportional to $\frac{1}{\bar{n}}$ and, moreover, it increases statistical uncertainties.

Capitolo 4

Introduction to machine learning

In this Chapter a brief introduction to *machine learning* (ML) will be given, for further details we recommend Alpaydin [2014], Shalev-Shwartz and Ben-David [2014] and https://developers.google.com/machine-learning/crash-course/ml-intro. For an introduction of ML techniques for cosmological applications see Ntampaka et al. [2019], He et al. [2019] and https://github.com/georgestein/ml-in-cosmology.

The term ML refers to the automated detection of meaningful patterns in data, commonly implemented using computing systems called *Artificial Neural Networks* (or, more simply *Neural Networks*) and inspired by the animals brain functioning. In the past couple of decades, ML has become a common tool in almost any task that requires information extraction from large data sets. ML algorithms have been introduced to solve various tasks, when human expertise does not exist or when we are unable to find a model behind some empirical results. For these features ML algorithms are widely used in astrophysical studies. The main advantage of these techniques is the capability of dealing with huge dataset finding non-linear relations between input features. Aside from this, ML can also reduce the need for computational power. For instance it can substitutes the standard likelihood inference for constraining cosmological parameters. In what follows we will present the building blocks of the *multi-layer perceptron Neural Network* (MLP), which is one of the simplest ML architecture. MLP can be used for regression tasks, such as to constrain cosmological parameters starting from N-body simulations, which is the primary focus of this Thesis work. Beside MLP, there are many ML architectures that can be used for analysing N-body simulations results. For instance, we may cite Convolutional Neural Network (CNN). CNN are networks, typically exploited for image recognition/classification, that use filter kernels to analyse patterns at different scales on 2D or 3D images, for further details see Zhang et al. [2019] and Pan et al. [2019].

The main topics of a ML algorithm that will be analyzed in this Chapter are:

- The dataset
- The Neural Network architecture
- The training phase
- The validation phase
- The test phase

4.1 The dataset

The starting point of every ML algorithm is the dataset. The learning technique depends on the input dataset. More specifically the learning can be supervised or unsupervised. Unsupervised Neural Networks search for patterns in the input data without a known output value. In the supervised case, like the MLP one, the input data is split into the *training dataset* and the *validation dataset*. The training dataset contains input features and labels. The Neural Network can improve its performances comparing its outputs with the labels, i.e. the expected results. The output data of MLP can be numerical values, for *regression* tasks, or classes, for *classification* problems. The training dataset will be used during the so-called training phases, when it will be processed by the Neural Network to compare the outputs with the expected labels. The parameters of the network will change according to the difference between outputs and labels within a certain metric. The test dataset is, instead, used after the training process to evaluate the precision of the Neural Network on data it has never processed. Conventionally the ratio between training and validation datasets is 4 : 1. Starting from the training data another set may be created, called the *validation dataset*, which is used among training phases to estimate how the network generalizes the results during training epochs. In fact a common problem that can arise in ML analyses is the so-called *overfitting*, when the network excellently connects training data to training labels, but it is not able to achieve the same precision when it is applied on the validation set. Using test sets, one can stop the training phase when the network begins to overfit, or when it reaches the desired accuracy.

4.2 The Neural Network architecture

From the operational point of view, ML is commonly implemented with Neural Networks, one of whose is the aforementioned MLP. Like other Neural Networks, the MLP is made up of different layers (at least three) of neurons. The neuron, the atom of a Neural Network, is a component which performs the following linear operation.

$$\vec{y} = f(\vec{W} \cdot \vec{x} + \vec{B}), \tag{4.1}$$

where $\vec{x}$ is the input vector and $\vec{y}$ the output one. $\vec{W}$ and $\vec{B}$, which are the *weight* and *bias* vectors, respectively, are two learnable parameters of the neuron, i.e. they will be changed during the training phases to improve the network precision. The function $f(x)$ is called *activation function*. This function was introduced to mirror the human neuron behaviour, and for MLP is usually non-linear. The choice of a non-linear activation function, except for the last layer one, makes the network able to distinguish data non-linearly separable. This fundamental feature could be explained in terms of the *universal approximation theorem* [Hornik, 1991]. This theorem states that a multilayer feed-forward network can potentially approximate every continuous functions. Up to now, the most frequently used activation function for MPL is the Leaky Rectified Linear activation (leaky-relu), first introduced in acoustic models [Maas et al., 2013]. Its functional form is the following:

$$f(x) = \begin{cases} x & \text{for } x > 0 \\ \frac{x}{c} & \text{for } x \leq 0 \end{cases} , \tag{4.2}$$

where c is a fixed arbitrary parameter greater than $+1$, that was introduced to solve the null gradient problem for inputs close to zero. Stacking up a certain number of neurons, with the same activation function, forms one *layer*. There are three types of layers: *input*, *hidden* and *output* ones. The input layer is connected to the input data, thus it must have the number of neurons equal to the number of input features. The data are then processed by the hidden layers until the output layer, that must have a number of neurons equal to the number of output labels. In the MLP architecture, the layers are fully connected, thus every neuron of hidden layers is linked to all neurons of the previous and following layers. The MLP is a feed-forward artificial Neural Network, hence connections between the nodes do not form a cycle. Summing up, in a MPL input data are linked to input layer, then are passed to hidden layers and finally to the output one. Each neuron implements a linear operation, with two free parameter vectors, to which an activation function, like the one described in Eq. (4.2), is applied. Therefore every layer contains two learnable matrices of weight and bias values.

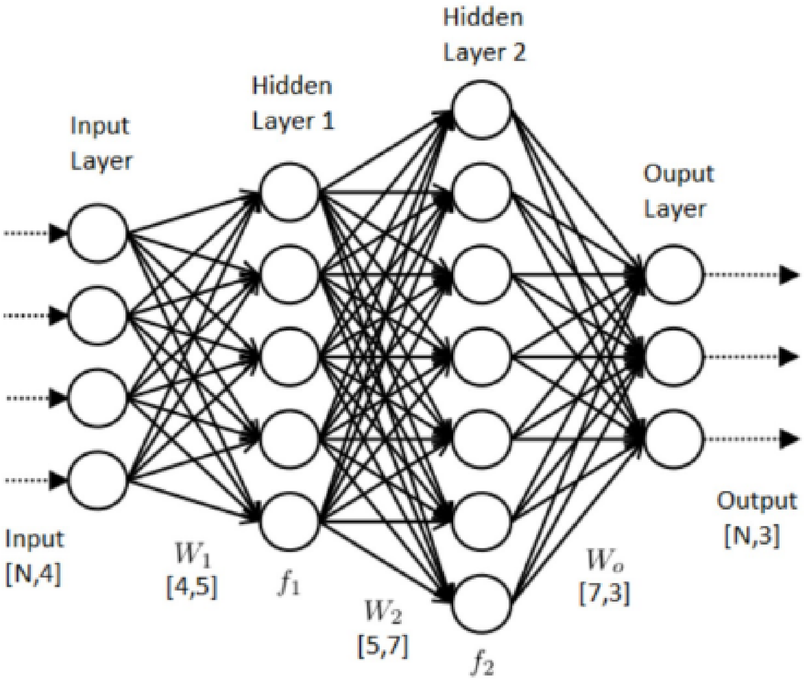


Figura 4.1: Tipical architecture for a MLP. The N input data, consistent of four elements, pass through the net until the output layer that returns 3 values for each input element. The weight matrix for each layer, except the input one, are indicates with the letter W (credits to https://www.datasciencecentral.com/profiles/blogs/how-to-configure-the-number-of-layers-and-nodes-in-a-neural).

4.3 The training phase

During the so-called training phase the network processes the input, and then compares the output with the expected labels. The results are then used to change the network parameters leading to better predictions with a technique called *backpropagation* [Rumelhart et al., 1986]. The training phase is determined by two fundamental functions, the *loss function* and the *optimizer*. The loss function is used as an estimator of the network precision and depends on the specific problem. In what follow we will consider only regression tasks. The standard loss function is the mean squared error (MSE), also known as squared L2 norm, which, for output o and label l over N input data, reads:

$$MSE(o,l) = \frac{1}{N}\sum_{i=1}^{N}(l_i - o_i)^2. \tag{4.3}$$

Other suitable choices are the L1 norm (mean absolute error) or the Smooth L1 loss, which computes the L1 loss if the mean absolute error is below 1 and the L2 loss otherwise.

As states above, the loss function is an estimator of the network precision. Then we need one more component that modifies the free parameters of the MLP according to the loss function results. This component is known as the optimizer, or optimization algorithm. Its purpose is to refresh the learnable parameters in order to minimize the loss function. The standard optimization algorithm for MLP is the *stochastic gradient descent* (SGD) method. Within this algorithm the gradient calculated from the data is replaced with a stochastic approximation (derivated from a random subsample of the training data) in order to converge faster to the desired solution. The iterative method, starting from a loss function of the i-th input expressed in term of the free parameters $L_i(\omega)$, can be described with the following equation:

$$\omega_{new} = \omega_{old} - \eta \nabla L_i(\omega_{old}), \tag{4.4}$$

where the algorithm iterates over a random subsample of the input data. The parameter η in known as the *learning rate*, and is a fundamental hyperparameter of Neural Networks that determines the step size toward the direction of the loss function minimum. In order to optimally train a Neural Network this parameter has to be carefully fine-tuned, because the value of the learning rate is a trade-off between the convergence speed and the achievement of the absolute minimum. Another optimizer, that will be used in the following, is the *adaptive moment estimation* algorithm (ADAM) [Kingma and Ba, 2014] which only involves first-order gradients with little memory requirement.

4.4 The validation phase

At the end of each training epoch, the network provides the loss function evaluated on the training and validation dataset. Comparing validation and training datasets is useful to finetuning the network hyperparameters, such as the learning rate or the number of layers/neurons. Three different cases, summarized in Fig 4.2, are possible:

- **High train and validation loss:** This case is called *underfitting*. The network is not able to properly adjust its learnable parameters and the predictions are not improving.
- **High validation loss and low train loss:** This case is called *overfitting*. The network is good at predicting labels for the training data, but it is not able to generalize.
- **Low training and validation loss, of comparable values:** This is the optimal result and tells us that the network has adequately learned.

By analysing the validation and training loss evolution during epochs it is possible to find fluctuating trend due to the stochastic nature of the optimization algorithm or caused by wrong tuned parameters. For instance a large learning rate could lead to oscillation around the absolute minimum of the parameter space.

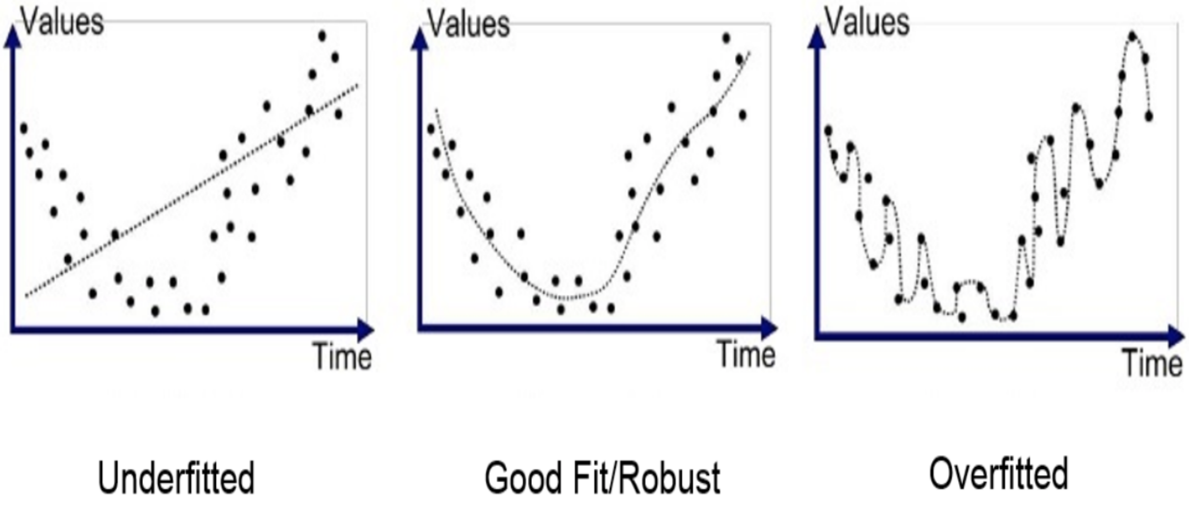


Figura 4.2: Possible training outcome for a regression Neural Network (credits to https://towardsdatascience.com/how-to-increase-the-accuracy-of-a-neural-network-9f5d1c6f407d).

4.5 The test phase

However the test bench for the network is the validation set. Thus we expect that the test loss is a little bit higher than the validation one since the validation set could be a random subsample of the training data while the test set has been never seen by the network

Capitolo 5

The Quijote N-body simulations

In order to fully describe the large-scale structure of the Universe, from linear to non-linear scales, we need to analyze numerical simulations. The test bench for the statical analysis of the density field described in Sec. 3 are the 3-dimensional surveys of cosmological probes. Unfortunately, surveys are limited to quasi-linear or non-linear scales by observational bias and represent amazing but difficult scientific and engineering challenges. One way to tackle this problem is to generate and study numerical simulations that can provide a large dataset to constrain cosmological parameters via statistical and/or numerical approaches. This motivation lead Francisco Villaescusa-Navarro and his team to develop one of the largest set of N-body simulations, the *Quijote simulations* [Villaescusa-Navarro et al., 2019].

5.1 Simulation features

The full set of Quijote N-body simulations of DM particles contains 43100 realizations, spanning on 7000 different cosmological models. The full dataset, consisting of simulation snapshots and cosmological statistics occupies more than 1 Petabyte and was assembled in more than $3.5 \cdot 10^7$ CPU hours. The Quijote simulations have been generated using GADGET III, an improved version of GADGET-II [Springel, 2005] TreePM code (see Sec. 2.2.1). All the simulations have a cosmological volume of 1 $(\mathrm{Gpc/h})^3$. Most of them are filled with 512^3 CDM particles, while there are also low- (256^3) and high- (1024^3) resolution simulations. The particles started on a regular three-dimensional grid. The softening length for the gravitational force is 100, 50 and 25 kpc/h for low-, fiducial- and high- resolution simulations, respectively . For every simulation snapshots at redshift $z = 0$, 0.5, 1, 2 and 3 are available. The initial conditions have been generated at redshift 127 in different ways: in most cases with second-order perturbation theory (see Sec. 2.1.3) with the code provided at https://cosmo.nyu.edu/roman/2LPT/, while in few cases with the Zeldovich approximation (see Sec. 2.2).

Name	Ω_m	Ω_b	h	n_s	σ_8	M_ν(eV)	w	realizations	simulations	ICs	$N_c^{1/3}$	$N_\nu^{1/3}$
Fid	$\underline{0.3175}$	$\underline{0.049}$	$\underline{0.6711}$	$\underline{0.9624}$	$\underline{0.834}$	$\underline{0}$	$\underline{-1}$	15000 500 500 1000 100	standard standard paired fixed standard standard	2LPT Zeldovich 2LPT 2LPT 2LPT	512 512 512 256 1024	0 0 0 0 0
Ω_m^+	$\underline{0.3275}$	0.049	0.6711	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
Ω_m^-	$\underline{0.3075}$	0.049	0.6711	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
Ω_b^{++}	0.3175	$\underline{0.051}$	0.6711	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
Ω_b^+	0.3175	$\underline{0.050}$	0.6711	0.9624	0.834	0	-1	500	paired fixed	2LPT	512	0
Ω_b^-	0.3175	$\underline{0.048}$	0.6711	0.9624	0.834	0	-1	500	paired fixed	2LPT	512	0
Ω_b^{--}	0.3175	$\underline{0.047}$	0.6711	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
h^+	0.3175	0.049	$\underline{0.6911}$	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
h^-	0.3175	0.049	$\underline{0.6511}$	0.9624	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
n_s^+	0.3175	0.049	0.6711	$\underline{0.9824}$	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
n_s^+	0.3175	0.049	0.6711	$\underline{0.9424}$	0.834	0	-1	500 500	standard paired fixed	2LPT	512	0
σ_8^+	0.3175	0.049	0.6711	0.9624	$\underline{0.849}$	0	-1	500 500	standard paired fixed	2LPT	512	0
σ_8^-	0.3175	0.049	0.6711	0.9624	$\underline{0.819}$	0	-1	500 500	standard paired fixed	2LPT	512	0
LH	[0.1 , 0.5]	[0.03 , 0.07]	[0.5 , 0.9]	[0.8 , 1.2]	[0.6 , 1.0]	0	-1	2000 2000	standard fixed	2LPT	512 512	0

Tabella 5.1: Cosmological model parameters of the Quijote set of N-body simulations. The underlined value refers to the fiducial cosmology. The parameters n_s is the spectral index of the power spectrum (see Eq. (3.11)) [Villaescusa-Navarro et al., 2019]

Let us consider the initial perturbation density field in Fourier space $\delta(k) = Ae^{i\phi}$. Recalling the discussion on the Gaussian properties of the density field, provided in Sec. 3.1.1, it can be demonstrated that ϕ has a random uniform distribution in the $[0, 2\pi]$ interval while the amplitude A is extracted from a Rayleigh distribution related to the value of $P(k)$. This is the standard way to generate initial conditions. We will define *fixed* simulations the sub-set of simulations in which the amplitude A is fixed to the square root of the variance of the Rayleigh distribution, while two fixed simulations with density fields $\delta(k)_1 = Ae^{i\phi_1}$ and $\delta(k)_2 = Ae^{i\phi_2}$ with $|\phi_1 - \phi_2| = \pi$ are called *pair-fixed* simulations (see [Angulo and Pontzen, 2016] and [Villaescusa-Navarro et al., 2018]).

5.1.1 Cosmological models

The Quijote simulations set contains simulations with different values of the following parameters: Ω_m, Ω_b, h, n_s, σ_8, w_{DE} and M_ν. In this Thesis work we will not consider simulations with massive neutrinous or with $w_{DE} \neq -1$ (see Eq. (1.37)) because they were not enough in number, and so the Neural Network cannot reach an acceptable accuracy. The parameters of the cosmological models that will be analysed in the following Sections are summed up in the Table 5.1: The values of the cosmological parameters of the fiducial model are consistent to Planck 2018 results [Planck Collaboration et al., 2018].

The fundamental sub-set of Quijote simulations that will be used as training data for the following Neural Network analysis is the one called *latin hypercube* . The name of this set refers to *latin hypercube sampling* (LHS) a sampling method parameter space that attempts to distribute samples evenly [McKay et al., 1979]. Hence, for the number of simulations and for the finely sampling of the parameter space, the latin hypercube is an optimal training dataset for the Neural Network algorithms discussed in this Thesis work.

To study the impact of cosmological parameters on the main statistical functions, $\vec{S}$, such as power spectrum and the two-point correlation function, it may be also useful to calculate the partial derivatives of these functions with respect to each parameter, θ, namely:

$$\frac{\partial \vec{S}}{\partial \theta} \sim \frac{S(\vec{\theta} + d\vec{\theta}) - S(\vec{\theta} - d\vec{\theta})}{2d\theta}, \tag{5.1}$$

In order to calculate the partial derivatives, many simulations have been generated by Villaescusa-Navarro and his team. More specifically, the simulations with superscripts "+" and "-" in Fig. 5.1, share all the cosmological parameters of the fiducial simulations except one, which is slightly larger or smaller. The difference between the fiducial parameter and the modified one are lower than 5%, to guarantee enough accuracy in approximating the derivative with Eq. (5.1), and higher than 1.8% to keep numerical noise contribution low.

5.2 Data products

The full dataset of the Quijote simulations is stored in the Gordon cluster of the San Diego Supercomputer Center, and it can be accessed through the web service globus (https://www.globus.org/). The simulation snapshots contain positions and velocities of DM particles at redshifts 0, 0.5, 1, 2, 3. Initial conditions at $z = 127$ are also provided. To read the snapshots and calculate many cosmological functions the *Pylians* library (https://github.com/franciscovillaescusa/Pylians) can be used. However, due to the huge size of the simulation snapshots, we focus only on subproducts generated by Villaescusa-Navarro and his team and stored in the San Diego Supercomputer. So, in what follows, we will consider three products, namely, the two-point correlation function, the power spectrum and the bispectrum, both in real and redshift space. There are other important byproducts, such as density fields (see in Fig. 5.1), halo and void catalogues, marked power spectrum and probability density functions, that have not been considered in this Thesis work, and will be exploited in a future work.

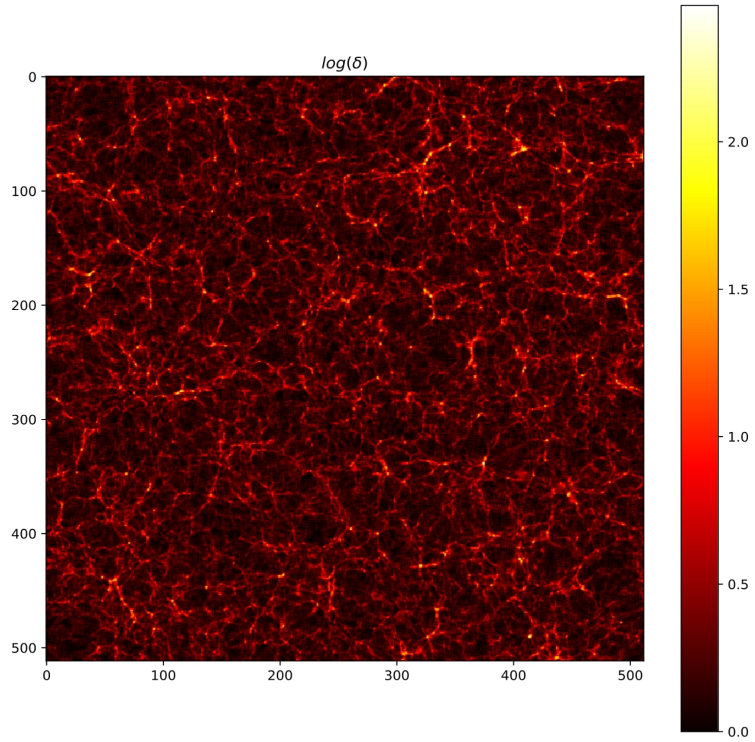


Figura 5.1: Contrast density field for the fiducial simulations at $z = 0$ generated with the Pylians library [Villaescusa-Navarro et al., 2019].

The two-point correlation function

To measure the two-point correlation function of the Quijote simulation snapshots, the Pylians library used by Villaescusa-Navarro and his team performs the following procedure:

- First, a regular grid was created with N^3 cells (where N^3 is the total number of particles). Then, every particle has been assigned to the closest grid points using the Cloud-in-Cell (CIC) algorithm [Laux, 1996].

- The density field has been calculated, as defined in Eq. (2.1).

- The Fourier transform of the density field has been performed, then the modules of each Fourier mode have been computed. The resulting $|\delta(k)^2|$ provide an estimate of the power spectrum.

- Finally, the power spectrum is anti-transformed, obtaining the two-point correlation function.

This procedure holds for real and redshift space. In redshift space, RSDs have been introduced along one Cartesian axis, and the correlation function monopole ($\nu = 0$), quadrupole ($\nu = 2$) and hexadecapole ($\nu = 4$) have been calculated as follows (see [Grasshorn Gebhardt and Jeong, 2018]):

$$\varepsilon^{\nu}(r) = \frac{1}{2\pi^2} \int_0^{\infty} k^2 P(k) \frac{j_l(kr)}{(kr)^{\nu}} dk, \tag{5.2}$$

where $j_l(kr)$ are the Bessels functions. An example of two-point correlation function monopole extracted from the same simulation at different redshifts is shown in Fig. 5.2.

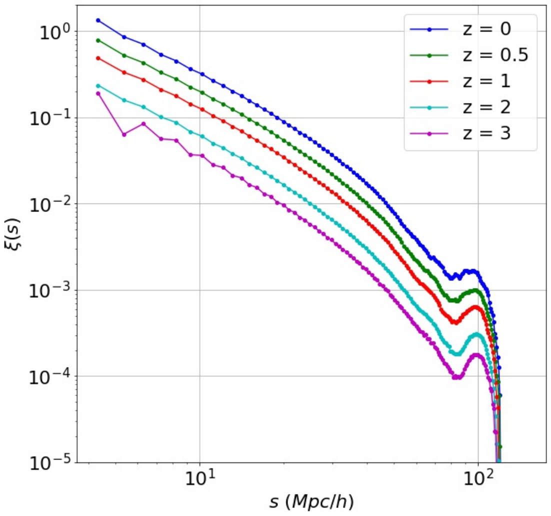


Figura 5.2: Two-point correlation function monopole in real space for the fiducial cosmological model at high resolution (1024^3 particles) and different redshifts. [Villaescusa-Navarro et al., 2019].

Power spectrum and bispectrum

The power spectrum is calculated following the same procedure as for the correlation function, negletcting the last step and normalizing $|\delta(k)^2|$. As for the two-point correlation function, the power spectrum has been measured in both real and redshift space. In redshift space, RSDs have been introduced along the three Cartesian axis, and power spectra monopole, quadrupole and hexadecapole have been derived. An example of power spectrum monopole extracted from the same simulation at different redshifts is shown in Fig. 5.3.

The bispectrum monopole is measured as follows:

$$B_0(k_1, k_2, k_3) = \frac{1}{V_B} \int_{k_1} d^3q_1 \int_{k_2} d^3q_2 \int_{k_3} \delta_D(q_1, q_2, q_3)\delta(q_1)\delta(q_2)\delta(q_3) d^3q_3 - B^{SN}, \quad (5.3)$$

where δ_D is the Dirac delta function, V_B is a normalization factor and B^{SN} is the correction for Poisson shot noise. Specifically, the bispectra are calculated using the code publicly available at https://github.com/changhoonhahn/pySpectrum. The perturbation field $\delta(x)$ is decomposed on a grid with $N_{grid} = 360$ and triangle configurations with bins of width $\Delta k = 0.01885$ h Mpc^{-1}, with $k_{max} = 0.5$ h Mpc^{-1}. As for the previous two statistical functions, bispectrum is computed both in real and redshift space.

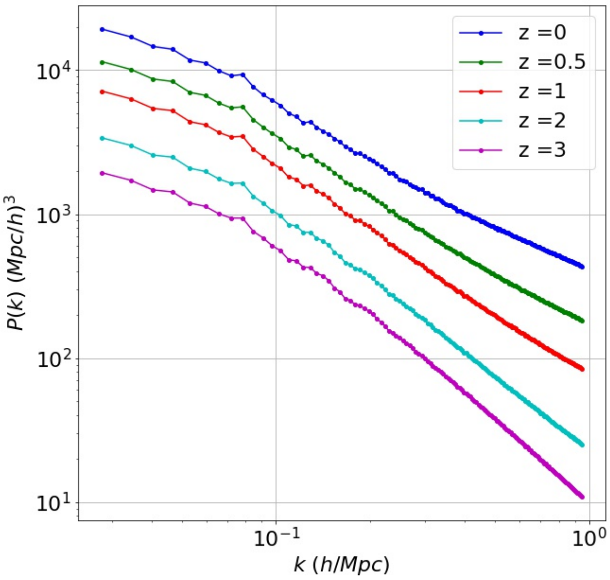


Figura 5.3: Matter power spectrum in real space for the fiducial cosmological model at the fiducial resolution. Color code indicates different redshifts. [Villaescusa-Navarro et al., 2019].

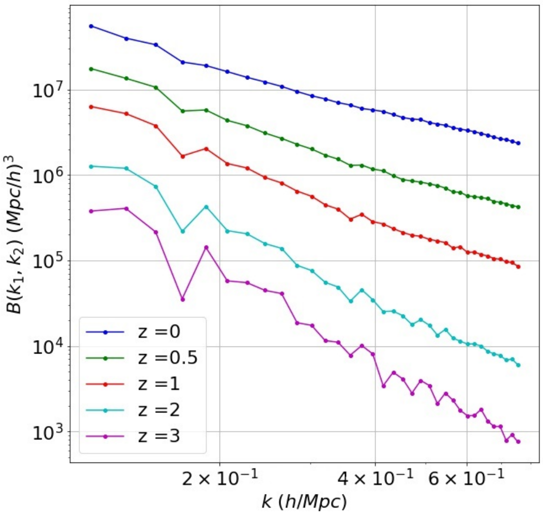


Figura 5.4: Unreduced bispectrum in real space for the fiducial cosmological model at the fiducial resolution. Color code indicates different redshifts. [Villaescusa-Navarro et al., 2019].

5.3 Statistical results

The Quijote N-body simulations provide an optimal dataset for statistical analyses to constrain cosmological parameters emulating the main summary statistics of the density field. These kind of analyses have been implemented by Villaescusa-Navarro and his team. Their results will be used to pre-process the input data of our ML algorithms and compared with the output of our Neural Networks. One way to quantify the information content of a set of cosmological parameters, $\vec{\theta}$, given a statistics $\vec{S}$, is by using the *Fisher matrix* F:

$$F_{ij} = \sum_{\alpha\beta} \frac{\partial S_\alpha}{\partial \theta_i} \frac{\partial S_\beta}{\partial \theta_j} C^{-1}_{\alpha\beta}, \tag{5.4}$$

where S_i is the i-th element of the statistics $\vec{S}$ and C is the covariance matrix:

$$C_{\alpha\beta} =< (S_\alpha - \bar{S_\alpha})(S_\beta - \bar{S_\beta}) > . \tag{5.5}$$

The Quijote simulations provide all elements to calculate Eqs. (5.4) and (5.5). In particular, one set of simulated data has been used to compute the covariance matrix, while the other simulations have been used to estimate the derivative of the statistics with respect to cosmological parameters as described in Sec. 5.1.1. The creators of the

Quijote simulations used this approach to forecast the uncertainties of the main ΛCDM cosmological parameters from the matter power spectrum in real space at $z = 0$, assuming the fiducial cosmological model, for different values of the maximum Fourier mode, k_{max}. The results of this analysis are presented in Fig. 5.5, that shows the constraints of the ΛCDM cosmological parameters coming from the Fisher's analysis of the matter power spectrum as a function of the maximum wavenumber considered, k_{max}. It is evident how, moving towards higher scales in Fourier space, the constraints improve.

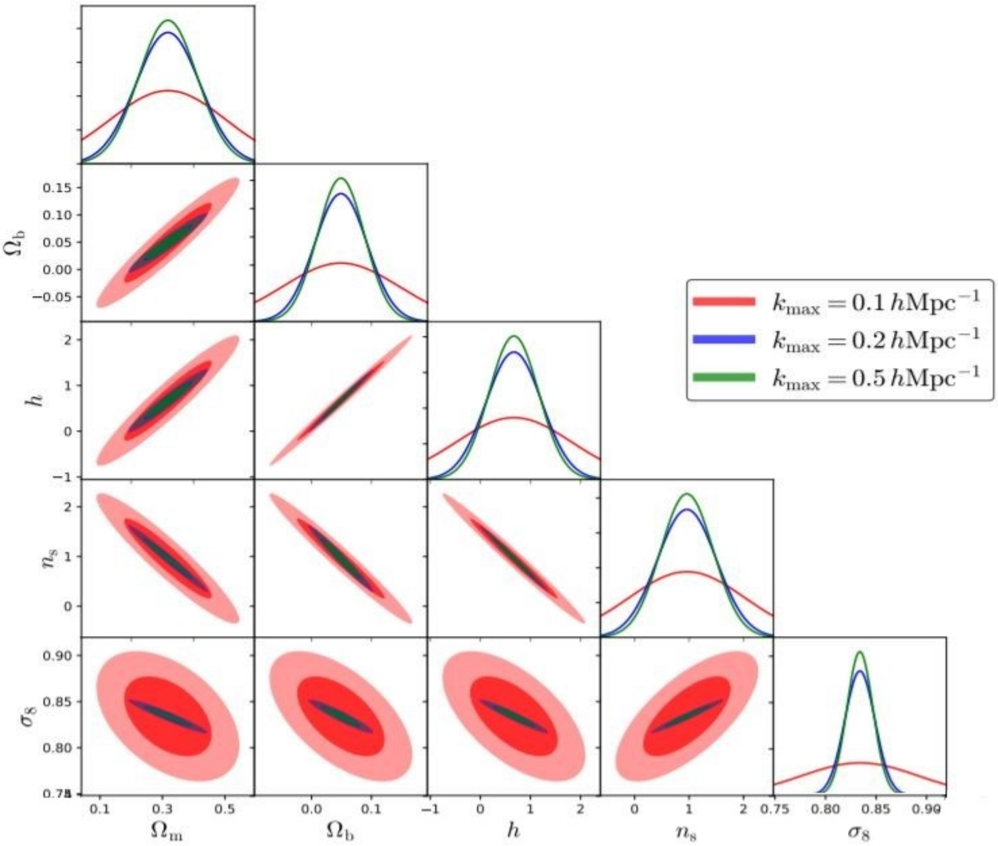


Figura 5.5: Constraints on the values of ΛCDM cosmological parameters from the matter power spectrum in real space, at $z = 0$, for different k_{max}. The ellipses indicate 1σ and 2σ constraints. The last panels in each row show the marginalized posterior probability function for each cosmological parameter [Villaescusa-Navarro et al., 2019].

A powerful way to search for new statistics, starting from the Fisher matrix, is via Information Maximising Neural Networks (IMNN) [Charnock et al., 2018]. The goal of IMNN is to maximize the Fisher matrix using latin-hypercube simulated data. Besides this approach, the Quijote simulations and, in particular, its latin-hypercube dataset can be used to train machine learning algorithms. Thus Villaescusa-Navarro and his team develop a *random decision forest* algorithm [Ho, 1995]. The inputs for this machine learning algorithm are the probability density functions, while the labels are five cosmological parameters, Ω_m, Ω_b, h, n_s, σ_8. The input dataset has been split in train and test datasets (see Sec. 4.1), with 1600 simulations for the first dataset and 400 for the second one. For this purpose the one-dimensional probability density functions have been smoothed on a scale of 5 Mpc/h using a top hat filter at $z = 0$. The result of the random forest is a function f that maps inputs, $\vec{pdf}$, into labels $\vec{\theta}$:

$$\vec{\theta} = f(\vec{pdf}(1 + \delta)) \tag{5.6}$$

The machine learning approach does not need any template likelihood. As it can be seen in Fig. 5.6, it is only able to accurately predict the value of σ_8 and, poorly, Ω_m. For the other parameters the code always predicts the mean value of the considered parameter on the training set.

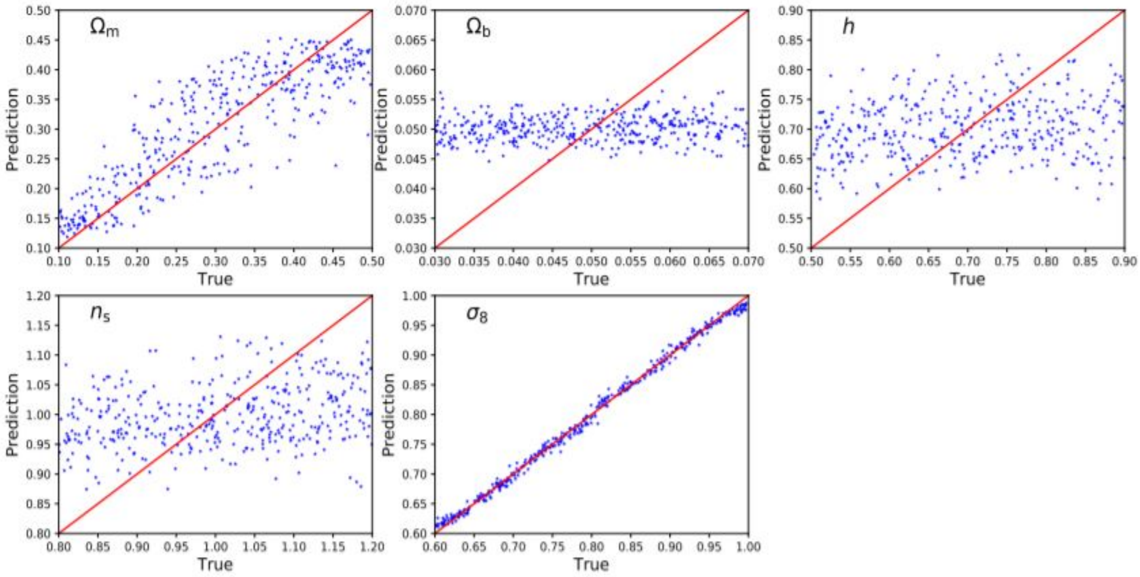


Figura 5.6: Results of the random decision forest algorithm on the test set of 400 probability density functions calculated starting from latin-hypercube simulations for five cosmological parameters, as labeled is each panel [Villaescusa-Navarro et al., 2019].

Another possible approach to the problem, which can improve the previous results, is the one that uses CNN, introduced in Sec. 4, to identify feature directly from the three-dimensional density field. However this method needs a more accurate hyperparameter fine-tuning, more training data and a higher computation power during the training phase, thus it is slow for this kind of datasets.

Capitolo 6

Machine learning analysis of Quijote simulations

The Quijote simulations presented in the previous Section, and more specifically the latin-hypercube dataset, are an exceptional test bench for ML algorithms that aim to constrain the cosmological parameters starting from N-body simulation results, such as two-point correlation functions, power spectra and bispectra. In this Chapter we will describe the implementation of new ML algorithms, trained on these data, that is able to derive satisfactory constraints on five fundamental ΛCDM cosmological parameters.

6.1 Machine learning framework

The ML algorithms developed in this Thesis work has been built within the *Pytorch* framework [Paszke et al., 2019]. PyTorch is an open source machine learning library based on the Torch library [Collobert et al., 2002] that provides an imperative and Pythonic programming style, focused on usability and speed. PyTorch is written in Python and supports hardware accelerators such as graphics processing unit (GPU). The PyTorch package provides two high-level features:

- Tensor computation with strong GPU acceleration
- Deep Neural Networks based on automatic differentiation, i.e. an automatic system for gradients computation

The GPU acceleration is implemented with the CUDA parallel computing platform (https://developer.nvidia.com/cuda-zone) that allows to speed up the training phase up to a factor 100 with respect to standard CPUs. The fundamental modules of the Pytorch library that have been used in this Thesis work are:

- **The Autograd module :** Autograd is a reverse automatic differentiation system that records on a graph all the operations applied on data. In this graph, leaves are input tensors and roots are output tensors. By tracing the graph from roots to leaves, Autograd automatically computes the gradients using the chain rule. Hence, this module is a fundamental tool to save time during the training phase.

- **The Optim module :** The Optim module implements optimization algorithms. Within this module, the optimization codes described in Sec. 4.3 and tools to adjust net's hyperparameters between training epochs, such as the *learning rate schedulers* (see [Wu et al., 2019]), can be found.

- **The nn module :** The nn module contains every building blocks that we need in order to implement a Neural Network architecture, such as layers and loss functions. This module depends on the Autograd one to define models and differentiate them, and on Optim module to adjust the parameters during the training phase.

Finally, all the algorithms have been implemented on the Google Colaboratory (Colab) notebook (https://colab.research.google.com/) , a Jupyter-like notebook that allows the user to write and execute Python programs on the cloud. Colab also gives free access to GPUs and TPUs with no configuration required.

6.2 Analysed dataset

The training dataset of the Neural Network analysis presented in this Section is the set of two-point correlation functions, power spectra and bispectra calculated from the latin-hypercube sub-set of the Quijote N-body simulations. All the input data have been normalized in the $[0, 1]$ range. The normalizing operation on the data generally speeds up the learning, and leading to faster convergence (see [Sola and Sevilla, 1997]). The latin-hypercube dataset has been split into training and validation datasets, according to the rule described in section 4.3. The test dataset, instead, comes from 500 fiducial simulations, and, again, all the data have been normalized in the $[0, 1]$ range. For the analyses of this Thesis work, we have used data at redshift 0, 0.5 and 1 calculated in redshift space. For each of the aforementioned statistics, we have considered all the three different realizations, where the RSDs have been placed on the three different spatial axes. In summary, for the training and validation phases we use, in total, 2000 different simulations coming from the latin hypercube dataset, each one at $z = 0$, $z = 0.5$ and $z = 1$ with 3 different realizations of RSDs. So, the total number of input data is 36000, of which 27000 have been used for training phase, while a random subsample (changed after every training epochs) of 9000 data have been used for the validation phase. The five cosmological parameters Ω_m, Ω_b, h, n_s and σ_8 have been used as labels.

As reported in the Fig 5.1, the labels lie in the following intervals:

- $0.1 \leq \Omega_m \leq 0.5$,
- $0.03 \leq \Omega_b \leq 0.07$,
- $0.5 \leq h \leq 0.9$,
- $0.8 \leq n_s \leq 1.2$,
- $0.6 \leq \sigma_8 \leq 1$.

6.2.1 Power spectrum

The redshift-space power spectrum monopole, quadrupole and hexadecapole moments have been measured by Villaescusa-Navarro et al. [2019] from the latin-hypercube simulations, using the Pylians library. An example of power spectra dataset member used as Neural Networks training data in this Thesis work are shown in Fig. 6.1.

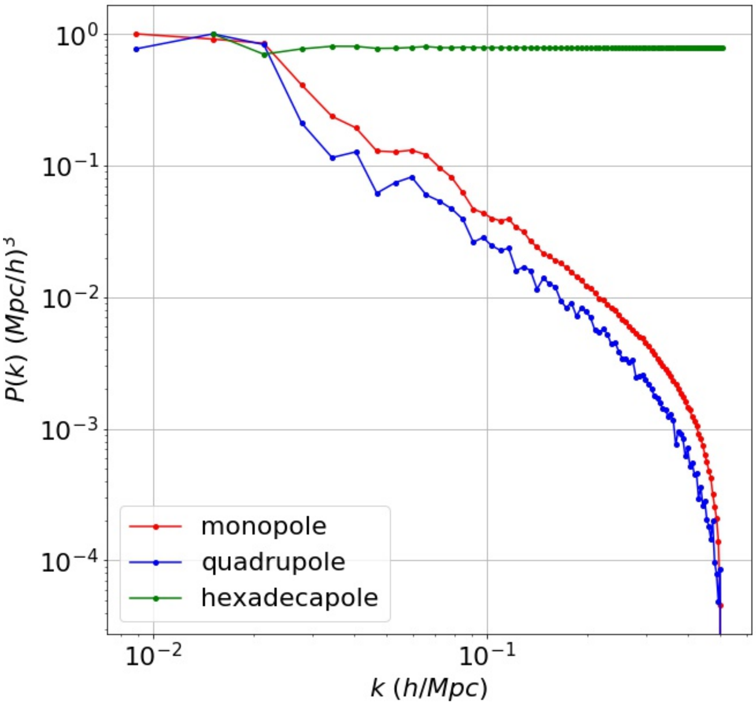


Figura 6.1: Power spectrum monopole, quadrupole and hexadecapole at $z = 0$ with RSDs placed on the x axis. normalized in the $[0, 1]$ range. The power spectrum has been calculated from the first realization of the latin-hypercube simulations with cosmological parameters $\Omega_m = 0.1755$, $\Omega_b = 0.06681$, $h = 0.7737$, $ns = 0.8849$, $\sigma_8 = 0.6641$.

The maximum value of k considered in the training phase, k_{max}, has been choosen by comparing the relative difference between the real-space power spectrum of the Quijote simulations at the fiducial cosmology and the power spectrum calculated with the CAMB non-linear code [Lewis and Challinor, 2011]. As illustrative case, Fig. 6.2, shows the result at $z = 0$. Another possible way to chose the value of k_{max} comes from the original analysis of Villaescusa-Navarro et al. [2019], described in Sec. 5.3, who provided k-dependent constrains on the cosmological parameters as a function of k_{max}, as shown in Fig. 6.3.

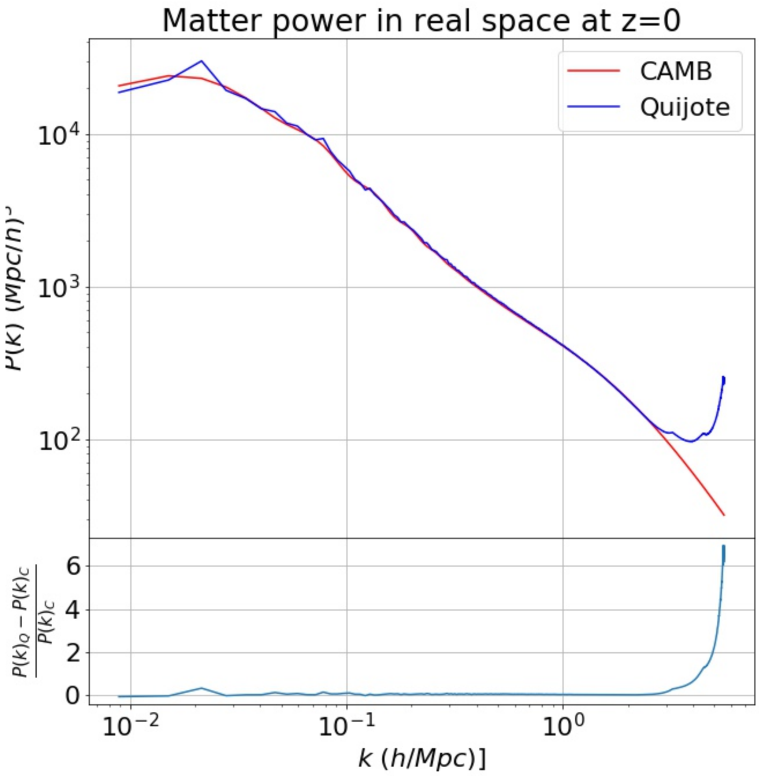


Figura 6.2: Comparison between the real-space power spectrum monopole of the Quijote simulations, at $z = 0$, at the fiducial cosmology and the corresponding power spectrum estimated with CAMB.

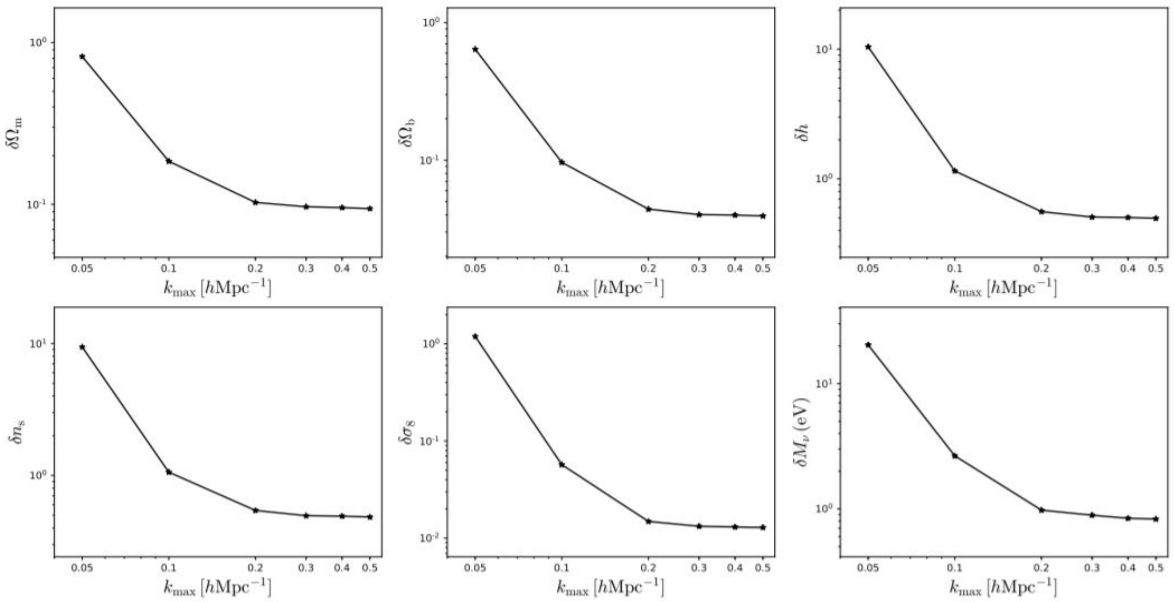


Figura 6.3: Marginalized 1σ constraints on the value of the cosmological parameters from the analysis of the matter power spectrum in real space at $z = 0$ presented in Sec. 5.3 as a function of k_{max} [Villaescusa-Navarro et al., 2019].

Considering the two methods, we decided to cut the power spectrum multipoles at $k_{max} = 0.5$h/Mpc to limit the analysis at large enough scales where the simulations' results are reliable. The resulting power spectrum consists of 80 linear bin from $k = 0$ to $k = k_{max}$ for each power spectrum multipole. Finally, the different multipoles have been concatenated into a single vector, with 240 elements. This vector, without the k mode that are the same for every spectrum, is one of the input of the Neural Network algorithm.

6.2.2 Bispectrum

The bispectra of the latin-hypercube dataset of the Quijote simulations has been calculated by Villaescusa-Navarro et al. [2019] using the Pylians library. The measured bispectra are store in files with 10 columns. The first three columns refer to the length of the triangle sides, while the following three are the power spectrum calculated at these values of k. The 7-th, 8-th and 9-th columns are, respectively, the bispectrum monopole $B_0(k_1, k_2)$ (see. (3.35)), already shot-noise corrected, the reduced bispectrum $Q(k_1, k_2)$, the analogue of the reduced three-point correlation function, and the shot noise correction $B_{SN}(k_1, k_2)$. The last column refers to the number of triangles in the bin. For the Neural Network analysis of this Thesis work, we considered only the columns with the bispectrum monopole and the reduced bispectrum. The values of k, the shot-noise correction and the number of triangles in the bin have been neglected. An example of the bispectrum monopole and of the reduced bispectrum monopole is shown in Fig. 6.4. The considered k range goes from $\frac{6\pi}{1000}$ to $\frac{240\pi}{1000}$ h/Mpc with a bin interval of $\frac{6\pi}{1000}$ h/Mpc. Thus there are 6350 triangle configurations. Hence, the input vector for the Neural Network generated from the bispectrum is made of 12700 elements.

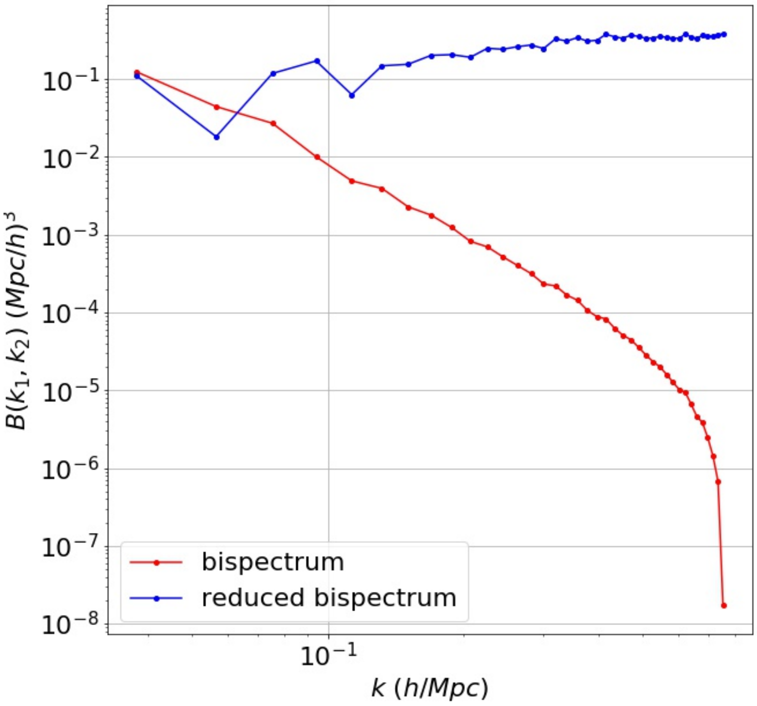


Figura 6.4: Equilateral ($k_1 = k_2 = k_3$) and normalized bispectrum and reduced bispectrum at $z = 0$ with RSDs placed on the x axis. The bispectrum has been calculated from the first realization of the latin-hypercube simulations with cosmological parameters $\Omega_m = 0.1755$, $\Omega_b = 0.06681$, $h = 0.7737$, $ns = 0.8849$, $\sigma_8 = 0.6641$.

6.2.3 Two-point correlation function

As for the power spectrum and the bispectrum, the two-point correlation function monopole, quadrupole and hexadecapole have been computed with the Pylians library in redshift space. All multipoles have been cut at r_{max} corresponding to the k_{max} of the power spectrum. An example of the two-point correlation function multipoles is shown in Fig. 6.5. The resulting input vector used for the training of the Neural Network is made of 240 elements.

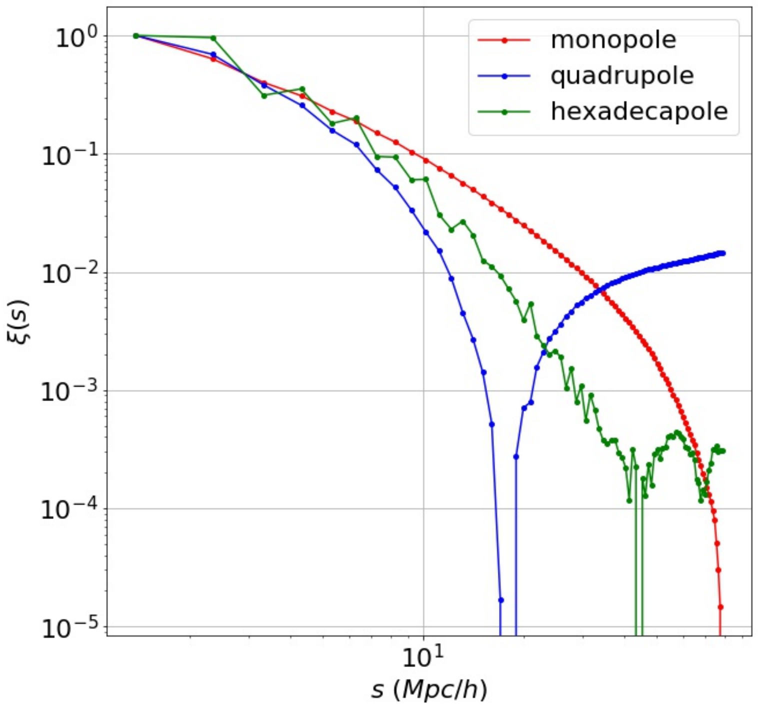


Figura 6.5: Normalized two-point correlation function monopole, quadrupole and hexadecapole at $z = 0$ with RSDs placed on the x axis. The two-point correlation function has been calculated from the first realization of the latin-hypercube simulations with cosmological parameters $\Omega_m = 0.1755$, $\Omega_b = 0.06681$, $h = 0.7737$, $ns = 0.8849$, $\sigma_8 = 0.6641$.

6.2.4 Test

A subset of the simulations at the fiducial cosmology have been used for the test phase. Specifically, for every statistical function in input, the test is performed with 500 data coming from the fiducial simulation for every redshifts, thus 1500 data in total. Once the network is trained, it predicts a value for the five parameters that are hopefully close to the true values. Hence, after a Gaussian fit of the prediction occurrences, it is possible to use the mean value as a guess for the true value of each parameter, and the variance as an estimator of the error. Recalling that the network has never seen the test dataset before the end of the training phase, we expect an higher value for the loss function on this dataset.

6.3 Neural Network architecture

The Neural Networks that we implement in this Thesis work take in input one cosmological function, and predict the 5 output labels, i.e. the 5 cosmological parameters. Specifically, we have chosen MLP Neural Networks that perform regression task. Since there are three different input functions, i.e. the power spectrum, bispectrum and the two-point correlation function, we created three different Neural Networks that share most of the features. In order to build a general model for all the three input types, an accurate hyperparameters search was performed, to be able to predict accurate values for the output labels with the same net architecture from the three different inputs. To achieve this, only the first layer has been changed according to the number of input features. The main features of this net architecture are the following:

- **Input layer :** The input layer contains a number of neurons equal to the number of input features, thus 240 for the correlation function and power spectrum, and 12700 for the bispectrum.

- **Hidden layers :** There are three hidden layers with 1024, 2048 and 1024 units. These values have been accurately fine-tuned. A deeper network may achieve a better accuracy, because it has a larger number of learnable parameters, but this increases the risk of overfitting.

- **Dropout layer :** This kind of layer was applied only on the last hidden layer. The dropout layer is a particular type of layer, active only during training phase that neglects the values coming from a given percentage (50% for our networks) of the previous layer's neurons, in order to reduce the overfitting (see [Srivastava et al., 2014]).

- **Output layer :** The output layer consists of 5 units, one for each cosmological parameter that we want to constrain.

- **Batch normalization layers :** The batch normalizations (bn) layer have been applied onto all the layers, except the output one. This special kind of layer normalizes the input data of each neuron with an operation that introduces two more learnable parameters. The aim of the bn layers is to speed up the training process and to reduce the overfitting (see [Ioffe and Szegedy, 2015]).

- **Activation functions :** For every layers, except the output one, we choose the leaky ReLu activation function described in Eq. (4.2).

- **Batch-size :** The batch size, i.e. the number of input data that will be propagated together through the network within each epochs, is set to 256. This value is a good trade-off between convergence speed and accuracy (see [Shirish Keskar et al., 2016]).
- **Optimizer :** After testing with all the optimizers implemented in the pytorch library, we choose the ADAM optimizer, described in Sec. 4.3, that gives the best results with low computational time requirement.
- **Learning rate :** The learning rate choose for this Thesis analysis is set to 0.005. This value, accurately fine-tuned after many tests, is a good choice for the first training epochs. Then, after few epochs, it has to be changed to find the absolute minimum in the gradient space rather than a relative one. To perform this operation a learning rate scheduler has been used, implemented in the Optim module, that multiplies the learning rate for a constant γ at the given epochs. The optimal values found for γ is 0.6, while the scheduler has been triggered after 50, 200 and 400 epochs, regardless of the input data.
- **Loss function :** As we perform a typical regression task, we choose the standard MSE loss function, described in Eq. (4.3). However, we also tried with other loss functions, such as the mean absolute error, but the final results did non improved.
- **Training epochs :** The Neural Network has been trained for 500 epochs, since we found that this value is large enought to reach stable results.

The resulting architecture is shown in Fig 6.6. The full codes used for this Thesis work and the training, validation and test dataset are available at https://drive.google.com/open?id=1dGu2Wn6zBpkP73m4_DmaE0InS3JiUWEX The results of the training, validation and test phases on this kind of Neural Network architecture will be presented and described in the following Chapter.

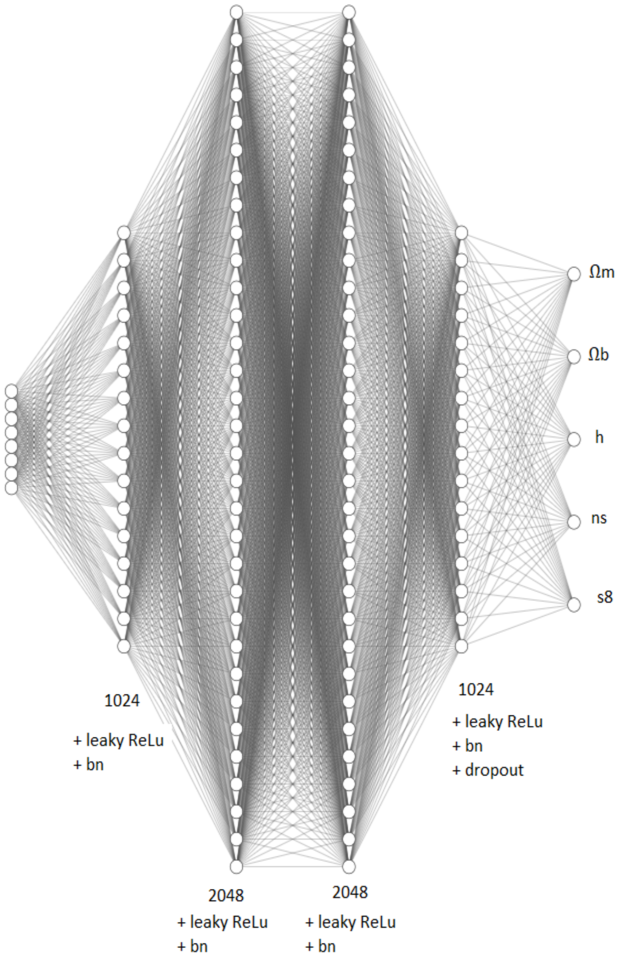


Figura 6.6: Schematic rappresentation of the Neural Network architecture used for this Thesis work. This figure has been generated with http://alexlenail.me/NN-SVG/index.html

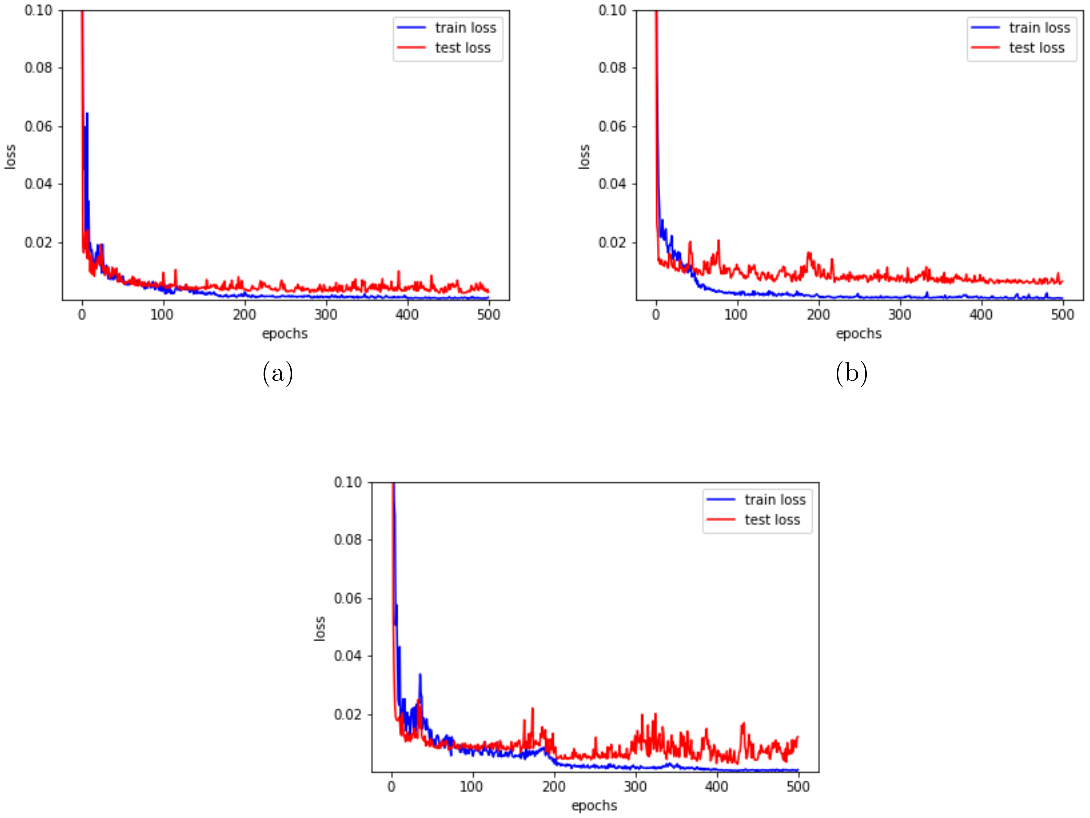
Capitolo 7

Results and conclusions

In the previous Chapter, we have described the training and validation datasets that have been given as input for presented Neural Networks presented in this Thesis work. The aim of this analysis is to investigate the capability of our ML infrastructure in constraining the ΛCDM cosmological parameters from the two-point correlation functions, the power spectra and the bispectra of CDM particles catalogues in redshift space. In this Chapter, we will analyze the outcomes of the training and validation phases, such as loss trends and predicted labels. Then, we will apply the trained net on the test dataset, made up of a subset of the fiducial simulations, to assess the performances in constraining the values of the five cosmological parameters considered. Finally, we will examine the results, comparing with the ones of recent literature works, and we will discuss future improvements, applications and perspectives for Neural Network analyses of the large-scale structures of the Universe.

7.1 Training and validation outcomes

To analyze the training and validation phases, we have to look at their MSE loss trends as a function of the training epochs, as shown in Fig. 7.1. Hence, recalling the discussion in Sec. 4.3, we may expect a validation MSE loss slightly higher than the training one. Fig. 7.1 represents the losses for the considered input functions over all the 500 training epochs. However, for the following analysis we only keep the weights relative to the best epoch, i.e. the epoch where the minimum values of training and validation loss have been achieved. In Fig. 7.1 we can also notice the effects of the learning rate scheduler, described in Sec. 6.3. The scheduler is triggered after 50, 200 and 400 epochs, in fact we can observe a small change in loss trends. After the aforementioned epochs the learning rate, from its initial value of 0.005, changes to 0.003, 0.0018 and 0.00108 respectively. Furthermore, we can see that the validation loss oscillates more than the training loss. This is due to the lower number of samples in the validation datasets, with respect to the training one. The losses plotted in Fig. 7.1 are the mean MSE losses calculated above each output, thus every the prediction error weights more on the validation loss than on the training one.



(a) (b)

(c)

Figura 7.1: The validation and train MSE loss trends as a function of training epochs for two-point correlation functions (a), power spectra (b) and bispectra (c) inputs. The epochs with the lowest values of train and validation loss are the 446-th (a), 446-th (b) and 489-th (c), respectively.

The saved weights for the three different input datasets refer to the following epochs:

- 446-th epoch with train loss$\sim$ 0.00040 and validation loss$\sim$ 0.00239 for two-point correlation functions.
- 446-th epoch with train loss$\sim$ 0.00049 and validation loss$\sim$ 0.00557 for power spectra.
- 489-th epoch with train loss$\sim$ 0.00034 and validation loss$\sim$ 0.00273 for bispectra.

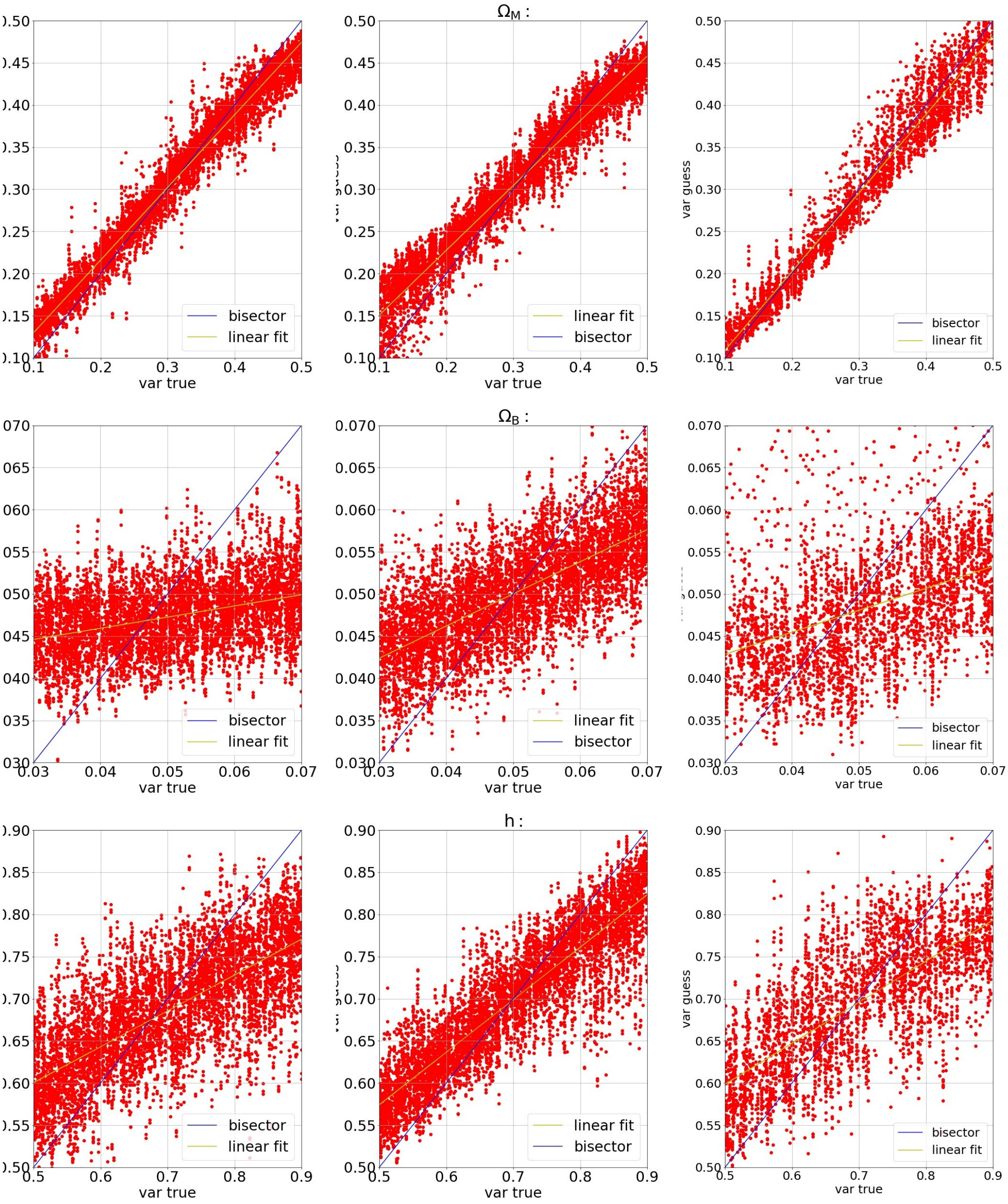
Theoretically, we have to save weight that refer to epochs at which the training and validation loss where comparable. We decide to continue training process until a minimum in both, training and validation loss, is reached. At that epoch the network suffers, a little, from overfitting. However, the resulting accuracy is higher than the case with comparable train and validation loss. In summary, looking at loss trends, we can conclude that the Neural Networks have been properly trained. In the next Section we will analyze the predictions made over the training datasets. These investigations are made as a sanity check before the test phase.

7.2 Network predictions over the training data

To visualize how the Neural Networks implemented in this Thesis work predict the labels for each training input, we plot the target labels with respect to the guess ones. The results are shown in Fig. 7.2. In this Figure we can notice the networks capabilities in linking each input with the corresponding label values. We also perform a linear fit on the predicted labels that, ideally, have to correspond to the bisector of the first quadrant. Fig. 7.2 provides an indicator of which labels the network predicts correctly, corresponding to the relations with small scatter and with a fit close to the bisector. From the results shown in Fig. 7.2, and the loss values described in the previous Section, we observe that the better predictions comes from the two-point correlation functions, while the worse results come from the power spectra. These two statistics have been calculated with the Pylians library as described in Sec. 5.2. Specifically, the two-point correlation functions are directly extracted from the power spectra, with the inverse Fourier transform. Hence we did not expect to find any difference in network predictions for these functions, in contrast to what is evident in the losses' plot and in Fig. 7.2; the network find clear links between two-point correlation functions and cosmological parameters, while this statement is not valid for power spectra. This unexpected founding deserves further investigations, that we postpone for a future dedicate work. As regard the bispectra, the prediction are, in general, noisier than the ones coming from the other two cosmological functions. To quantify the accuracy over training labels recognition, we can look at the linear fit coefficients m and q shown in Tab 7.1, from which we can conclude that the networks are good at predicting only certain label values for each training input. Specifically, Ω_m is accurately assessed by all the three different networks, that have linear fit coefficients comparable with the bisector ones. The outcomes for σ_8 and n_s labels is, instead, fairly reliable only in networks that takes as input two-point correlation functions and bispectra. As regards h and Ω_b, the three networks are not able to properly assess these labels. This result is due to the fact that the considered statistics do not strongly depend on these parameters. To summarize, we can conclude that the networks developed in this Thesis work are able to assess the values of Ω_m and, less accurately, n_s and σ_8 from the training datasets.

Label	Two-point correlation function	
	m	q
Ω_m	0.964 ± 0.008	0.002 ± 0.002
Ω_b	0.119 ± 0.013	0.0411 ± 0.0007
h	0.462 ± 0.016	0.377 ± 0.012
n_s	0.856 ± 0.013	0.144 ± 0.013
σ_8	0.942 ± 0.005	0.039 ± 0.004
	Power Spectrum	
	m	q
Ω_m	0.994 ± 0.009	0.014 ± 0.003
Ω_b	0.380 ± 0.016	0.0307 ± 0.0008
h	0.714 ± 0.014	0.200 ± 0.009
n_s	0.633 ± 0.013	0.372 ± 0.013
σ_8	0.661 ± 0.013	0.276 ± 0.011
	Bispectrum	
	m	q
Ω_m	0.937 ± 0.009	0.026 ± 0.003
Ω_b	0.220 ± 0.019	0.036 ± 0.001
h	0.48 ± 0.02	0.356 ± 0.016
n_s	0.805 ± 0.015	0.157 ± 0.015
σ_8	0.830 ± 0.012	0.117 ± 0.009

Tabella 7.1: The linear fit coefficients for the predictions over the training datasets.


$\Omega_M:$
$\Omega_B:$
$h:$
var true
var guess
bisector
linear fit
0.1
0.2
0.3
0.4
0.5
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.03
0.04
0.05
0.06
0.07
0.030
0.035
0.040
0.045
0.050
0.055
0.060
0.065
0.070
0.6
0.7
0.8
0.9
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90

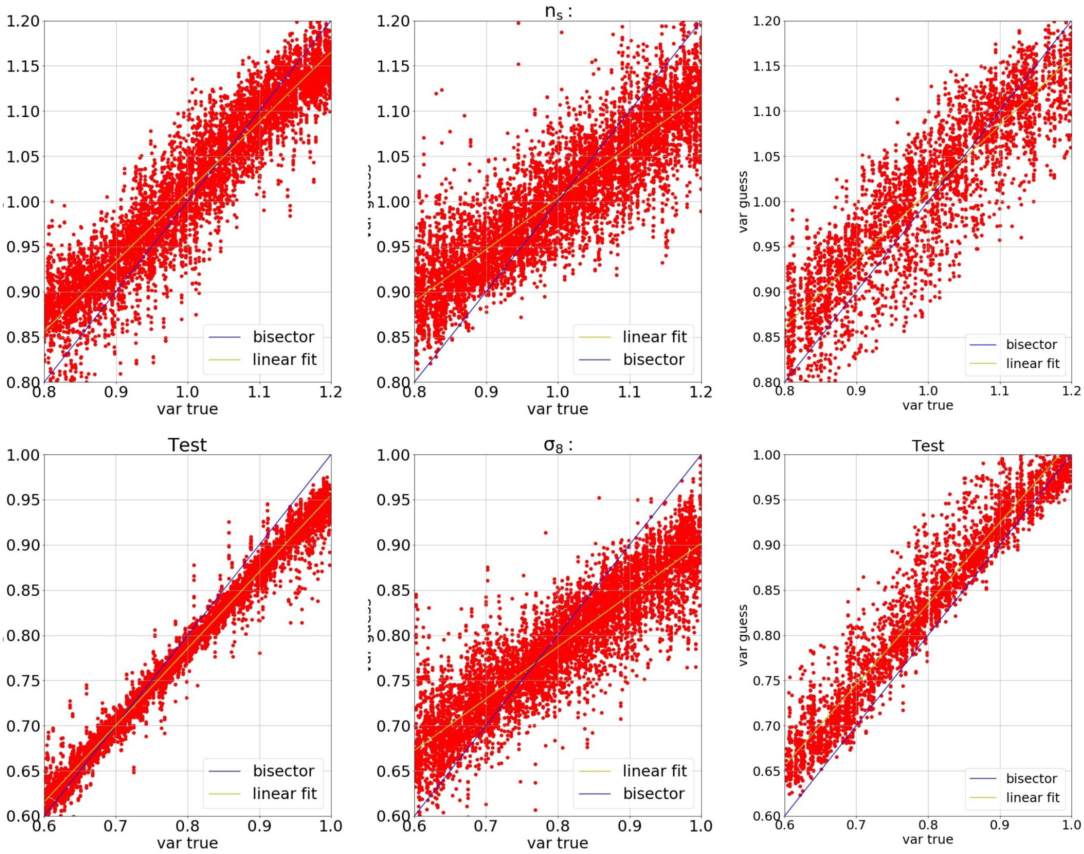


Figura 7.2: Cosmological parameters predicted by the Neural Network on the training set compared to the true parameters of the Quijote simulations (from left to right: two-point correlation functions, power spectra and bispectra).

7.3 The predictions on the test set

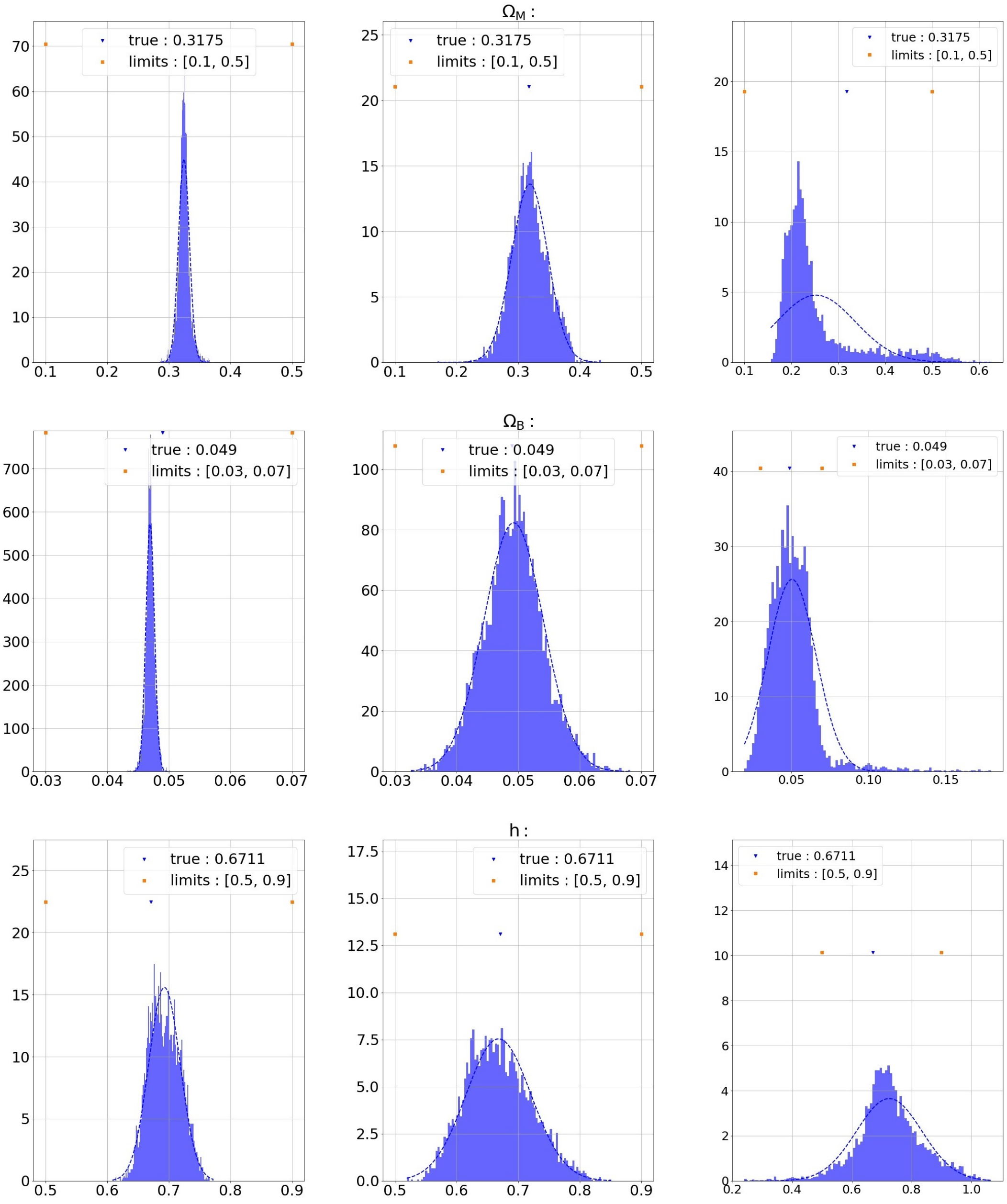
In the previous Section, we showed the results of the training and validation phases, where the input data come from the latin-hypercube subsample of the Quijote simulations. We may conclude that the three networks are well trained and able to predicts label values close to the real ones, for what concerns Ω_m, n_s and σ_8. However, the real proving ground for our ML algorithms is the test phases, that will be analyzed in this Section. In particular, we want to investigate how the networks predict labels for data that they have never seen. The test dataset is made of 1500 samples for each considered functions, extracted from 500 simulation runs at the fiducial cosmology, for the three different redshifts considered. For each label we built an histogram with the occurrences of the predicted values, in 100 linear bins. The resulting histograms, one for each label and for each network, are shown in Fig 7.3. In order to assess the five cosmological parameters of the fiducial simulations, we perform a Gaussian fit on these histograms. The two fitting parameters, the mean, μ, and the variance, σ, will be used as estimators for the predicted value and the absolute error of each cosmological parameters. The fit results are shown in Tab 7.2.

Label	Two-point correlation function	
	μ	σ
Ω_m	0.305	0.011
Ω_b	0.045	0.002
h	0.685	0.066
n_s	0.968	0.047
σ_8	0.826	0.007
	Power Spectrum	
	μ	σ
Ω_m	0.333	0.041
Ω_b	0.047	0.005
h	0.662	0.062
n_s	0.977	0.059
σ_8	0.828	0.051
	Bispectrum	
	μ	σ
Ω_m	0.252	0.084
Ω_b	0.051	0.015
h	0.725	0.109
n_s	1.051	0.109
σ_8	0.877	0.077

Tabella 7.2: The Guassian fit coefficients for the predictions over the test datasets.

Label	$\xi(r)$	$P(k)$	$B(k_1, k_2)$	Planck18
Ω_m	0.305 ± 0.011	0.33 ± 0.04	0.25 ± 0.08	0.315 ± 0.007
Ω_b	0.045 ± 0.002	0.047 ± 0.005	0.051 ± 0.016	0.04934 ± 0.0008
h	0.69 ± 0.07	0.66 ± 0.06	0.72 ± 0.11	0.674 ± 0.005
n_s	0.97 ± 0.05	0.98 ± 0.06	1.01 ± 0.11	0.965 ± 0.004
σ_8	0.826 ± 0.007	0.83 ± 0.05	0.88 ± 0.08	0.811 ± 0.006

Tabella 7.3: Comparison between our network results and Planck18 constrains.


Ω_M :
true : 0.3175
limits : [0.1, 0.5]
0 10 20 30 40 50 60 70
0.1 0.2 0.3 0.4 0.5
0 5 10 15 20 25
0.1 0.2 0.3 0.4 0.5
0 5 10 15 20
0.1 0.2 0.3 0.4 0.5 0.6
Ω_B :
true : 0.049
limits : [0.03, 0.07]
0 100 200 300 400 500 600 700
0.03 0.04 0.05 0.06 0.07
0 20 40 60 80 100
0.03 0.04 0.05 0.06 0.07
0 10 20 30 40
0.05 0.10 0.15
h :
true : 0.6711
limits : [0.5, 0.9]
0 5 10 15 20 25
0.5 0.6 0.7 0.8 0.9
0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5
0.5 0.6 0.7 0.8 0.9
0 2 4 6 8 10 12 14
0.2 0.4 0.6 0.8 1.0

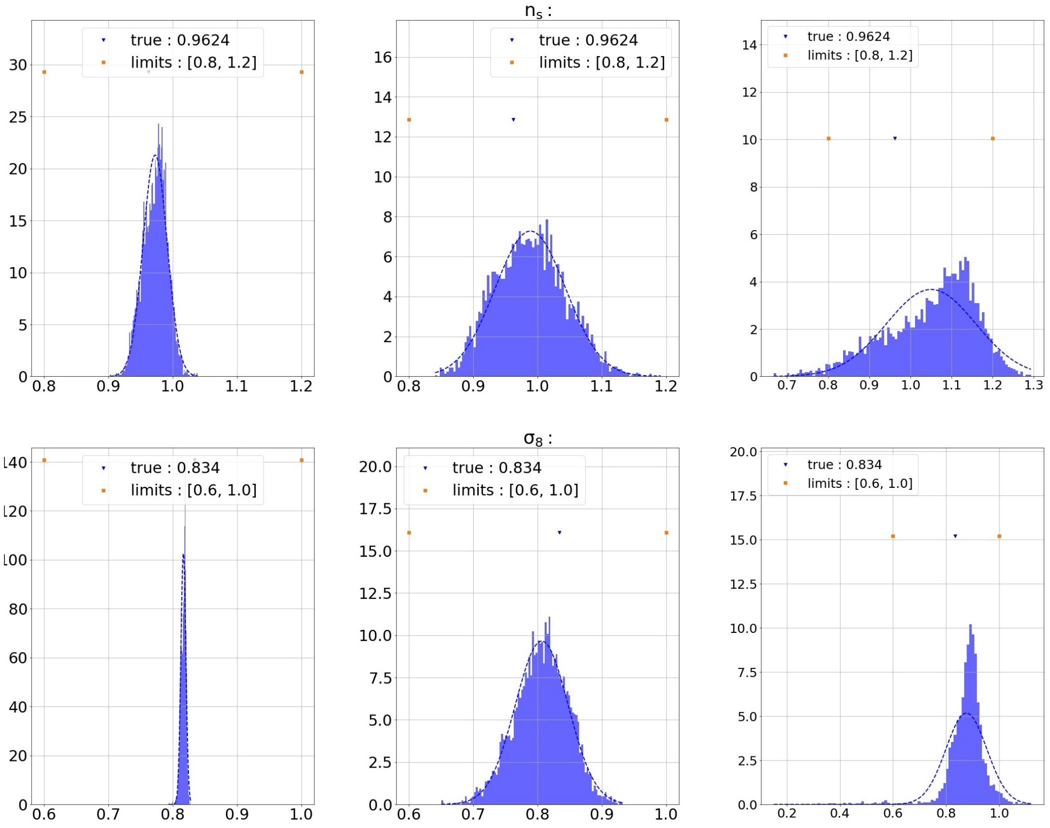


Figura 7.3: Histograms with the prediction for the test labels. The blue triangles indicate the expected values of the labels while the squares define the intervals of each cosmological parameters for the latin-hypercube simulations (from left to right: two-point correlation functions, power spectra and bispectra). The bi-modal trend of some histograms is an unexpected features that has to be investigated in future works.

Finally, we compare the 1σ constrains derived from the ML analysis of the fiducial subset of the Quijote simulations with the Planck18 results [Planck Collaboration et al., 2018]. The comparison is shown in Tab 7.3.

Capitolo 8

Conclusions

In this Thesis work we developed ML techniques to assess cosmological parameters from two-point correlation functions, power spectra and bispectra of the DM density field, trained on the snapshots of the Quijote N-body simulations. This kind of analysis represents a novel tool in the cosmological field, that aims to constrain cosmological parameters with a data-driven likelihood-free inference analysis. Here we summarize the main outcomes of this Thesis work. Then, we will give suggestions for possible future improvements that will pave the way for applications on real datasets.

The Neural Networks results, summarized in Table 7.3, are the final outcomes of our ML analysis of the Quijote N-body simulations. Almost all the presented results are fully consistent with expectations, that is with the cosmological parameters of the fiducial subset used as test dataset. However, the constraints depend on the statistics used to train and validate the Network. In particular, the Neural Network trained on the two-point correlation functions provides accurate constraints on most of the cosmological parameters, while the Neural Network trained with the power spectra provides constraints with larger confidence intervals, in particular for Ω_m and σ_8. This unexpected issue, also discussed in the previous Section, deserves further investigations. Finally, the Neural Network that has been trained with the bispectrum measurements, provides constraints on Ω_m and n_s that are much larger than in the other cases. This is due to the larger number of input features and the sparsity of the binning. We can assume that, with a larger number of training samples, and with a deeper Network, the results coming from this analysis could be tighter.

Summing up, the best results of the Neural Network analyses presented in this Thesis work are the outcomes of the Networks trained with the two-point correlation functions. For comparison, the resulting 1σ constraints on the Ω_m parameter of the ΛCDM model are a factor ~ 1.6 larger than the outcome of Planck18 collaboration. In Fig. 8.1, we show the comparison between the constraints on Ω_M derived in this Thesis work and some literature results. We recall that our result was obtained with a Neural Network trained and validated over 36000 two-point correlation function monopoles, quadrupoles, hexadecapoles in redshift space, at $z = 0$, 0.5 and 1, measured from mock DM catalogues in $V = 1(\mathrm{Gpc/h})^3$.

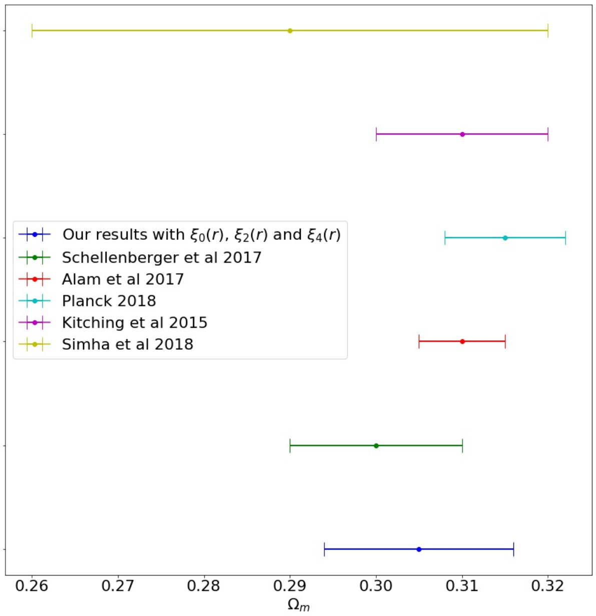


Figura 8.1: Constraints on Ω_m coming from our Neural Network compared to the latest literature outcomes. Specifically, Alam et al. [2017] and Simha and Cole [2013] derived constraints from the galaxy clustering of the SDSS survey, Kitching et al. [2015] from weak lensing shear maps and Schellenberger and Reiprich [2017] from galaxy clusters in X-ray band.

8.1 Future perspectives

The statistical approach consisting of exploiting ML algorithms to constrain the cosmological parameters of our Universe represents a newborn field in Cosmology. This branch became relevant only in the last few years and, up to now, it is in constant growth. The analysis presented in this Thesis work should be considered as a preliminary attempt to pave the way for a full ML exploitation of real cosmological datasets. In fact, our Neural Network analysis can be improved in several ways. First, if we train the Networks with larger datasets, spanning over more cosmological models, we will be able to extract tighter constaints on the cosmological parameters. Moreover, larger training datasets can make the Networks able to predict values for more labels, thus extending the number of cosmological parameters that the Network can constrain. Furthermore, with more computational power, it will be possible to exploit deeper Neural Networks. Recalling the overfitting problem, described in Sec. 4.4, if we want to work with larger Networks we have to accurately fine-tune the hyperparameters. This problem can be solved using algorithms that automate the search for optimal values in the hyperparameters space.

Other possible improvements can be obtained trying different Neural Networks architectures and training data. About the Network architecture, one of the spearheads of ML are the so-called Convolutional Neural Networks (see [Pan et al., 2019]). This kind of Networks, usually employed for image recognition, can be trained directly over simulation snapshots, bypassing the issue of finding statistical functions and estimators that contain enough information to train the Network. However, with this approach the number of input features, i.e. the computational power for Neural Networks training and exploitation, is extremely high. Thus, the alternative approach considered in this Thesis, which exploits statistical functions that can embed information on the large-scale structure of the Universe with a limited number of features, seems more convenient.

Finally, the training dataset should be as close as possible to real datasets. Indeed, we plan to extend the current analysis by training our Neural Networks with realistic mock galaxy catalogues. To investigate if the outcomes of the ML analysis are independent of the algorithms used for populating DM haloes with galaxies, we aim to train the Networks with datasets coming from different mock galaxy catalogues, constructued with different galaxy formation and evolution algorithms, with distinct initial conditions, different resolutions and considering different observables.

We would like to comment that the primary goal of ML techniques applied to cosmological studies should be to provide tools in support of the standard statistical analyses for a deeper understanding of our Universe and, in particular, of its large-scale structures. The ML approach has not to be used as a heuristic process that always gives the best answer without an understandable, logical procedure. Instead, this tool should be exploited within a theoretical framework, to support the discovery of general laws that can properly describe our Universe.

Bibliografia

S. J. Aarseth. Dynamical evolution of clusters of galaxies, I. *Monthly Notices of the Royal Astronomical Society*, 126:223, Jan 1963. doi: 10.1093/mnras/126.3.223.

T. M. C. Abbott, F. B. Abdalla, J. Annis, K. Bechtol, J. Blazek, B. A. Benson, R. A. Bernstein, G. M. Bernstein, E. Bertin, D. Brooks, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, C. L. Chang, T. M. Crawford, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, C. Davis, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, A. Drlica-Wagner, A. E. Evrard, E. Fernand ez, B. Flaugher, P. Fosalba, J. Frieman, J. García-Bellido, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, W. G. Hartley, J. W. Henning, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, E. Krause, K. Kuehn, S. Kuhlmann, N. Kuropatkin, O. Lahav, A. R. Liddle, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, A. Manzotti, M. March, J. L. Marshall, R. Miquel, J. J. Mohr, T. Natoli, P. Nugent, R. L. C. Ogando, Y. Park, A. A. Plazas, C. L. Reichardt, K. Reil, A. Roodman, A. J. Ross, E. Rozo, E. S. Rykoff, E. Sanchez, V. Scarpine, M. Schubnell, D. Scolnic, I. Sevilla-Noarbe, E. Sheldon, M. Smith, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, A. R. Walker, R. H. Wechsler, J. Weller, W. Wester, W. L. K. Wu, J. Zuntz, Dark Energy Survey Collaboration, and South Pole Telescope Collaboration. Dark Energy Survey Year 1 Results: A Precise H_0 Estimate from DES Y1, BAO, and D/H Data. *Monthly Notices of the Royal Astronomical Society*, 480(3):3879–3888, Nov 2018. doi: 10.1093/mnras/sty1939.

S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, J. Comparat, A. J. Cuesta, K. S. Dawson, D. J. Eisenstein, S. Escoffier, H. Gil-Marín, J. N. Grieb, N. Hand, S. Ho, K. Kinemuchi, D. Kirkby, F. Kitaura, E. Malanushenko, V. Malanushenko, C. Maraston, C. K. McBride, R. C. Nichol, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, M. Pellejero-Ibanez, W. J. Percival, P. Petitjean, F. Prada, A. M. Price-Whelan, B. A. Reid, S. A. Rodríguez-Torres, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martín, S. Saito, S. Salazar-Albornoz, L. Samushia, A. G. Sánchez, S. Satpathy, D. J. Schlegel, D. P. Schneider, C. G. Scóccola, H.-J. Seo, E. S. Sheldon, A. Simmons, A. Slosar, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. V. Magaña, J. A. Vazquez, L. Verde, D. A. Wake, Y. Wang, D. H. Weinberg, M. White, W. M. Wood-Vasey, C. Yèche, I. Zehavi, Z. Zhai, and G.-B. Zhao. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 470(3):2617–2652, Sept. 2017. doi: 10.1093/mnras/stx721.

C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, Oct 1979. doi: 10.1038/281358a0.

E. Alpaydin. *Introduction to Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 3 edition, 2014. ISBN 978-0-262-02818-9.

R. E. Angulo and A. Pontzen. Cosmological N-body simulations with suppressed variance. *Monthly Notices of the Royal Astronomical Society*, 462(1):L1–L5, Oct 2016. doi: 10.1093/mnrasl/slw098.

J. Barnes and P. Hut. A hierarchical O(N log N) force-calculation algorithm. *Nature*, 324(6096):446–449, Dec 1986. doi: 10.1038/324446a0.

H. Bondi and T. Gold. A new model for the expanding universe. *Monthly Notices of the Royal Astronomical Society, Vol. 108, No. 5, pp. 372–382*, 1948.

A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy. Sterile neutrino Dark Matter. *Progress in Particle and Nuclear Physics*, 104:1–45, Jan 2019. doi: 10.1016/j.ppnp.2018.07.004.

G. L. Bryan and M. L. Norman. Statistical properties of x-ray clusters: Analytic and numerical comparisons. *The Astrophysical Journal*, 495(1):80–99, mar 1998. doi: 10.1086/305262. URL https://doi.org/10.1086%2F305262.

M. Cacciato, F. C. van den Bosch, S. More, H. Mo, and X. Yang. Cosmological constraints from a combination of galaxy clustering and lensing - III. Application to SDSS data. *Monthly Notices of the Royal Astronomical Society*, 430(2):767–786, Apr 2013. doi: 10.1093/mnras/sts525.

J. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story. DeepCMB: Lensing reconstruction of the cosmic microwave background with deep neural networks. *Astronomy and Computing*, 28:100307, July 2019. doi: 10.1016/j.ascom.2019.100307.

T. Charnock, G. Lavaux, and B. D. Wandelt. Automatic physical inference with information maximizing neural networks. *Physical Review D*, 97(8):083004, Apr 2018. doi: 10.1103/PhysRevD.97.083004.

R. Collobert, S. Bengio, and J. Marithoz. Torch: A modular machine learning software library. 11 2002.

M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V. The two-point position and velocity correlations. *The Astrophysical Journal*, 267:465–482, Apr 1983. doi: 10.1086/160884.

P. F. de Salas, K. Malhan, K. Freese, K. Hattori, and M. Valluri. On the estimation of the local dark matter density using the rotation curve of the Milky Way. *Journal of Cosmology and Astroparticle Physics*, 2019(10):037, Oct 2019. doi: 10.1088/1475-7516/2019/10/037.

W. de Sitter. On einstein's theory of gravitation and its astronomical consequences. *Monthly Notices of the Royal Astronomical Society, 78, pp. 3–28*, 1917.

C. Dreissigacker, R. Sharma, C. Messenger, R. Zhao, and R. Prix. Deep-learning continuous gravitational waves. *Physical Review D*, 100, 08 2019. doi: 10.1103/PhysRevD.100.044009.

A. Einstein. Zur allgemeinen relativitätstheorie. *Preussische Akademie der Wissenschaften Sitzungsberichte, pp. 778–801*, 1915.

A. Einstein. Kosmologische betrachtungen zur allgemeinen relativitätstheorie. *Preussische Akademie der Wissenschaften, Sitzungsberichte, pp. 142–152*, 1917.

A. Einstein. Zum kosmologischen problem der allgemeinen relativitätstheorie. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse, pp. 235–237*, 1931.

F. Elsner, F. Schmidt, J. Jasche, G. Lavaux, and N.-M. Nguyen. Cosmology inference from a biased density field using the eft-based likelihood. *Journal of Cosmology and Astroparticle Physics*, 2020:029–029, 01 2020. doi: 10.1088/1475-7516/2020/01/029.

A. Friedmann. Uber die krummung des raumes. *Zeitschrift für Physik A, pp. 377–386,*, 1922.

T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf. Convolutional neural networks: a magic bullet for gravitational-wave detection? *Phys. Rev.*, D100(6):063015, 2019. doi: 10.1103/PhysRevD.100.063015.

N. Gillet, A. Mesinger, B. Greig, A. Liu, and G. Ucci. Deep learning from 21-cm tomography of the cosmic dawn and reionization. *Monthly Notices of the Royal Astronomical Society*, 484(1):282–293, Mar. 2019. doi: 10.1093/mnras/stz010.

H. S. Grasshorn Gebhardt and D. Jeong. Fast and accurate computation of projected two-point functions. *Physical Review D*, 97(2):023504, Jan 2018. doi: 10.1103/PhysRevD.97.023504.

J. E. Gunn and I. Gott, J. Richard. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *The Astrophysical Journal*, 176:1, Aug. 1972. doi: 10.1086/151605.

A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *The Astrophysical Journal*, 417:19, Nov. 1993. doi: 10.1086/173288.

E. Hawkins, S. Maddox, S. Cole, O. Lahav, D. S. Madgwick, P. Norberg, J. A. Peacock, I. K. Baldry, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, B. Jones, I. Lewis, S. Lumsden, W. Percival, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *Monthly Notices of the Royal Astronomical Society*, 346(1):78–96, Nov. 2003. doi: 10.1046/j.1365-2966.2003.07063.x.

S. He, Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, W. Chen, and B. Póczos. Learning to predict the cosmological structure formation. *Proceedings of the National Academy of Science*, 116(28):13825–13832, Jul 2019. doi: 10.1073/pnas.1821458116.

A. Herrera-Martín, M. Hendry, A. X. Gonzalez-Morales, and L. A. Ureña-López. Strong Gravitational Lensing by Wave Dark Matter Halos. *The Astrophysical Journal*, 872 (1):11, Feb 2019. doi: 10.3847/1538-4357/aafaf0.

T. K. Ho. Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*, ICDAR '95, page 278, USA, 1995. IEEE Computer Society. ISBN 0818671289.

R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. Taylor Francis, Inc., USA, 1988. ISBN 0852743920.

K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251 – 257, 1991. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(91)90009-T. URL http://www.sciencedirect.com/science/article/pii/089360809190009T.

F. Hoyle. A New Model for the Expanding Universe. *Monthly Notices of the Royal Astronomical Society*, 108:372, Jan 1948. doi: 10.1093/mnras/108.5.372.

E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *PNAS, pp. 168-173*, 1929.

S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv e-prints*, art. arXiv:1502.03167, Feb 2015.

N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227:1–21, Jul 1987. doi: 10.1093/mnras/227.1.1.

D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv e-prints*, art. arXiv:1412.6980, Dec 2014.

T. D. Kitching, M. Viola, H. Hildebrandt, A. Choi, T. Erben, D. G. Gilbank, C. Heymans, L. Miller, R. Nakajima, and E. van Uitert. RCSLenS: Cosmic Distances from Weak Lensing. *arXiv e-prints*, art. arXiv:1512.03627, Dec. 2015.

D. Kodi Ramanah, T. Charnock, F. Villaescusa-Navarro, and B. D. Wandelt. Super-resolution emulator of cosmological simulations using deep physical models. *arXiv e-prints*, art. arXiv:2001.05519, Jan. 2020.

N. Krachmalnicoff and M. Tomasi. Convolutional neural networks on the HEALPix sphere: a pixel-based algorithm and its application to CMB data analysis. *Astronomy and Astrophysics*, 628:A129, Aug. 2019. doi: 10.1051/0004-6361/201935211.

P. La Plante and M. Ntampaka. Machine Learning Applied to the Reionization History of the Universe in the 21 cm Signal. *The Astrophysical Journal*, 880(2):110, Aug. 2019. doi: 10.3847/1538-4357/ab2983.

S. D. Landy and A. S. Szalay. Bias and Variance of Angular Correlation Functions. *The Astrophysical Journal*, 412:64, July 1993. doi: 10.1086/172900.

S. E. Laux. On particle-mesh coupling in monte carlo semiconductor device simulation. *IEEE Trans. on CAD of Integrated Circuits and Systems*, 15:1266–1277, 1996.

A. Lewis and A. Challinor. Camb: Code for anisotropies in the microwave background. *Astrophysics Source Code Library*, pages 02026–, 02 2011.

L. Lucie-Smith, H. V. Peiris, and A. Pontzen. An interpretable machine-learning framework for dark matter halo formation. *Monthly Notices of the Royal Astronomical Society*, 490(1):331–342, Nov. 2019. doi: 10.1093/mnras/stz2599.

A. Maas, A. Hannun, and A. Ng. Rectifier nonlinearities improve neural network acoustic models. *ICML, 30*, 2013.

M. C. March, R. Trotta, P. Berkes, G. D. Starkman, and P. M. Vaudrevange. Improved constraints on cosmological parameters from Type Ia supernova data. *Monthly Notices of the Royal Astronomical Society*, 418(4):2308–2329, Dec 2011. doi: 10.1111/j.1365-2966.2011.19584.x.

J. Martin. The Theory of Inflation. *arXiv e-prints*, art. arXiv:1807.11075, Jul 2018.

M. McKay, R. Beckman, and W. Conover. Comparison the three methods for selecting values of input variable in the analysis of output from a computer code. *Technometrics; (United States)*, 5 1979. doi: 10.1080/00401706.1979.10489755.

J. Merten, C. Giocoli, M. Baldi, M. Meneghetti, A. Peel, F. Lalande, J.-L. Starck, and V. Pettorino. On the dissection of degenerate cosmologies with machine learning. *Monthly Notices of the Royal Astronomical Society*, 487(1):104–122, July 2019. doi: 10.1093/mnras/stz972.

P. Meszaros. The behaviour of point masses in an expanding cosmological substratum. *Astrophysical Journal*, 37(2):225–228, 1974.

H. Minkowski. Raum und zeit. *Physikalische Zeitschrift 10, pp. 75–88*, 1908.

A. Moss. Improved Photometric Classification of Supernovae using Deep Learning. *arXiv e-prints*, art. arXiv:1810.06441, Oct. 2018.

M. Münchmeyer and K. M. Smith. Fast Wiener filtering of CMB maps with Neural Networks. *arXiv e-prints*, art. arXiv:1905.05846, May 2019.

D. Muthukrishna, D. Parkinson, and B. E. Tucker. DASH: Deep Learning for the Automated Spectral Classification of Supernovae and Their Hosts. *The Astrophysical Journal*, 885(1):85, Nov. 2019. doi: 10.3847/1538-4357/ab48f4.

M. Ntampaka, C. Avestruz, S. Boada, J. Caldeira, J. Cisewski-Kehe, R. Di Stefano, C. Dvorkin, A. E. Evrard, A. Farahi, D. Finkbeiner, S. Genel, A. Goodman, A. Goulding, S. Ho, A. Kosowsky, P. La Plante, F. Lanusse, M. Lochner, R. Mandelbaum, D. Nagai, J. A. Newman, B. Nord, J. E. G. Peek, A. Peel, B. Poczos, M. M. Rau, A. Siemiginowska, D. J. Sutherland, H. Trac, and B. Wandelt. The Role of Machine Learning in the Next Decade of Cosmology. *Bulletin of the American Astronomical Society*, 51(3):14, May 2019.

S. Pan, M. Liu, J. Forero-Romero, C. G. Sabiu, Z. Li, H. Miao, and X.-D. Li. Cosmological parameter estimation from large-scale structure deep learning. *arXiv e-prints*, art. arXiv:1908.10590, Aug 2019.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, art. arXiv:1912.01703, Dec 2019.

P. Peebles. The galaxy and mass n-point correlation functions: a blast from the past. 04 2001.

P. J. E. Peebles. *The large-scale structure of the universe.* 1980.

P. J. E. Peebles. Discovery of the hot Big Bang: What happened in 1948. *European Physical Journal H*, 39(2):205–223, Apr 2014. doi: 10.1140/epjh/e2014-50002-y.

P. J. E. Peebles and M. G. Hauser. Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. *The Astrophysical Journal Supplement Series*, 28:19, Nov. 1974. doi: 10.1086/190308.

A. Peel, F. Lalande, J.-L. Starck, V. Pettorino, J. Merten, C. Giocoli, M. Meneghetti, and M. Baldi. Distinguishing standard and modified gravity cosmologies with machine learning. *Physical Review D*, 100(2):023508, July 2019. doi: 10.1103/PhysRevD.100.023508.

W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Z. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401(4):2148–2168, 01 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.15812.x. URL https://doi.org/10.1111/j.1365-2966.2009.15812.x.

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517(2):565–586, June 1999. doi: 10.1086/307221.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R.

Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wand elt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, art. arXiv:1807.06209, Jul 2018.

Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, B. Casaponsa, A. Challinor, H. C. Chiang, L. P. L. Colombo, C. Combet, B. P. Crill, F. Cuttaia, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, R. Fernand ez-Cobos, F. Finelli, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, Y. Giraud-Héraud, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, A. Zacchei, and A. Zonca. Planck 2018 results. V. CMB power spectra and likelihoods. *arXiv e-prints*, art. arXiv:1907.12875, Jul 2019.

G. Pratten, D. Munshiand, Patrick, Valageas, and P. Brax. 3d weak lensing: Modified theories of gravity. *Physical Review D, American Physical Society*, 93(10):103524, May 2016. doi: 10.1103/PhysRevD.93.103524.

A. Riess, S. Casertano, W. Yuan, L. Macri, and D. Scolnic. Large magellanic cloud cepheid standards provide a 1the determination of the hubble constant and stronger evidence for physics beyond λcdm. *The Astrophysical Journal*, 876:85, 05 2019. doi: 10.3847/1538-4357/ab1422.

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astrophysical Journal*, 116(3):1009–1038, Sept. 1998. doi: 10.1086/300499.

L. Roszkowski, E. M. Sessolo, and S. Trojanowski. WIMP dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6): 066201, Jun 2018. doi: 10.1088/1361-6633/aab913.

D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct 1986. doi: 10.1038/323533a0.

S. Sasaki. A New Method to Estimate Cosmological Parameters Using the Baryon Fraction of Clusters of Galaxies. *Publications of the ASJ*, 48:L119–L122, Dec. 1996. doi: 10.1093/pasj/48.6.L119.

G. Schellenberger and T. H. Reiprich. HICOSMO: cosmology with a complete sample of galaxy clusters - II. Cosmological results. *Monthly Notices of the Royal Astronomical Society*, 471(2):1370–1389, Oct. 2017. doi: 10.1093/mnras/stx1583.

J. Schmelzle, A. Lucchi, T. Kacprzak, A. Amara, R. Sgier, A. Réfrégier, and T. Hofmann. Cosmological model discrimination with Deep Learning. *arXiv e-prints*, art. arXiv:1707.05167, July 2017.

R. Scoccimarro, H.Couchman, and J. Frieman. The bispectrum as a signature of gravitational instability in redshift space. *The Astrophysical Journal*, 517(2):531–540, Jun 1999. doi: 10.1086/307220.

E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro. Cosmology and the bispectrum. *Physical Review D*, 74(2):023522, July 2006. doi: 10.1103/PhysRevD.74.023522.

S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning : from theory to algorithms*. 2014. ISBN 9781107057135 1107057132. URL http://www.worldcat.org/search?qt=worldcat_org_all&q=9781107057135.

N. Shirish Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. *arXiv e-prints*, art. arXiv:1609.04836, Sep 2016.

J. Silk. Cosmic black-body radiation and galaxy formation. *Astrophysical Journal vol. 151, pp. 459*, 1969.

V. Simha and S. Cole. Cosmological constraints from applying SHAM to rescaled cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 436(2): 1142–1151, Dec. 2013. doi: 10.1093/mnras/stt1643.

J. Sola and J. Sevilla. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science, vol. 44, no. 3*, pages 1464–1468, 6 1997.

V. Springel. The cosmological simulation code GADGET-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, Dec 2005. doi: 10.1111/j.1365-2966.2005.09655.x.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

M. Takada and B. Jain. The three-point correlation function in cosmology. *Monthly Notices of the Royal Astronomical Society*, 340(2):580–608, Apr. 2003. doi: 10.1046/j.1365-8711.2003.06321.x.

H. Totsuji and T. Kihara. The Correlation Function for the Distribution of Galaxies. *Publications of the ASJ*, 21:221, Jan. 1969.

F. Villaescusa-Navarro, S. Naess, S. Genel, A. Pontzen, B. Wandelt, L. Anderson, A. Font-Ribera, N. Battaglia, and D. N. Spergel. Statistical Properties of Paired Fixed Fields. *The Astrophysical Journal*, 867(2):137, Nov 2018. doi: 10.3847/1538-4357/aae52b.

F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. Kodi Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Allys, A. Brochard, C.-T. Chiang, S. He, A. Pisani, A. Obuljen, Y. Feng, E. Castorina, G. Contardo, C. D. Kreisch, A. Nicola, R. Scoccimarro, L. Verde, M. Viel, S. Ho, S. Mallat, B. Wand elt, and D. N. Spergel. The Quijote simulations. *arXiv e-prints*, art. arXiv:1909.05273, Sep 2019.

S. Weinberg. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity.* 1972.

P. Wilking, R. Röseler, and P. Schneider. Constrained correlation functions from the Millennium Simulation. *Astronomy and Astrophysics*, 582:A107, Oct. 2015. doi: 10.1051/0004-6361/201525906.

A. E. Willem de Sitter. On the relation between the expansion and the mean density of the universe. *Proceedings of the National Academy of Sciences. 18, pp. 213–214*, 1932.

K. C. Wong, S. H. Suyu, G. C. F. Chen, C. E. Rusu, M. Millon, D. Sluse, V. Bonvin, C. D. Fassnacht, S. Taubenberger, M. W. Auger, S. Birrer, J. H. H. Chan, F. Courbin, S. Hilbert, O. Tihhonova, T. Treu, A. Agnello, X. Ding, I. Jee, E. Komatsu, A. J. Shajib, A. Sonnenfeld, R. D. Bland ford, L. V. E. Koopmans, P. J. Marshall, and G. Meylan. H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes. *arXiv e-prints*, art. arXiv:1907.04869, Jul 2019.

Y. Wu, L. Liu, J. Bae, K.-H. Chow, A. Iyengar, C. Pu, W. Wei, L. Yu, and Q. Zhang. Demystifying learning rate polices for high accuracy training of deep neural networks. 08 2019.

X. Zhang, Y. Wang, W. Zhang, Y. Sun, S. He, G. Contardo, F. Villaescusa-Navarro, and S. Ho. From Dark Matter to Galaxies with Convolutional Networks. *arXiv e-prints*, art. arXiv:1902.05965, Feb 2019.

F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta, pp. 110–127*, 1933.