

SCUOLA DI SCIENZE
Corso di Laurea in Astrofisica e Cosmologia

**Cosmological exploitation
of Neural Networks:
constraining Ω_m from the two-point
correlation function of BOSS**

Relatore:
Dott. Federico Marulli

Presentata da:
Niccolò Veronesi

Correlatori:
Ch.mo Prof. Lauro Moscardini
Dott. Alfonso Veropalumbo

Appello III
Anno Accademico 2018/2019

Contents

Abstract	5
Sommario	7
Introduction	9
1 Theoretical cosmological background	13
1.1 Elements of General Relativity	13
1.2 Friedmann-Lemaître-Robertson-Walker metric	14
1.3 Distances in Cosmology and the Hubble’s Law	15
1.4 Friedmann models	17
1.5 Flat, open and closed models	20
1.6 The Λ CDM Model	21
2 Cosmic structures formation	27
2.1 Jeans length	27
2.2 Perturbations in an expanding Universe	29
2.3 Correlation Function and Power Spectrum	31
2.4 Evolution of the Power Spectrum	34
2.5 N-body simulations	37
3 Clustering	41
3.1 The two-point correlation function	41
3.2 Estimators of the correlation function	42
3.3 Measures of two-point correlation function	43
3.4 Geometric distortions	44
3.5 Redshift-space distortions	45
3.6 Baryon Acoustic Oscillations	46
4 Fundamentals of Machine Learning	49
4.1 Supervised Machine Learning	50
4.2 Model’s architecture of Neural Networks	51

4.2.1	Activation functions	52
4.2.2	Categories of layers	53
4.3	Training process	55
4.3.1	Loss	56
4.3.2	Optimization	58
4.4	Validation	61
5	The analyzed dataset	63
5.1	The Sloan Digital Sky Survey	63
5.2	Properties of BOSS	64
5.3	The correlation function of BOSS galaxies	64
5.4	Standard analysis on clustering statistics	68
5.4.1	Bayesian statistical analysis	68
5.4.2	Cosmological constraints from BOSS clustering	69
6	The training, validation and test datasets	73
6.1	Construction of log-normal mock catalogues	73
7	Exploitation of the Neural Networks	79
7.1	The Classification model	79
7.2	Results of the Classification model	82
7.3	The Regression model	85
7.4	Results of the Regression model	89
8	Conclusions	95
8.1	Final results and considerations	95
8.2	Future perspectives	96
A	ML code implementation	99
A.1	Classification algorithm	99
A.2	Regression algorithm	101

Abstract

One of the biggest challenges of the 21st century cosmology is to accurately estimate the cosmological model parameters and to discriminate among alternative cosmological frameworks. The large-scale structure of the Universe provides powerful probes to put constraints on the main cosmological parameters. Galaxy surveys exploit galaxies as observable tracers to map the total matter distribution of the Universe. During the last decade, increasingly large galaxy surveys have been conducted both with ground-based and space-borne telescopes. These instruments are able to produce tens of terabytes of data every day. Next-generation astronomical facilities, such as Euclid, the James Webb Space Telescope and the Large Synoptic Survey Telescope, will dramatically increase the amount of data available for cosmological investigations. This scenario suggests that Machine Learning (ML) techniques will play a key role in future researches, revolutionizing all astronomical data analysis methodologies. ML algorithms can learn the representations of the data they are fed with, and produce outputs the goodness of which depends on the amount of the input data. The ML models considered in this thesis work are often referred to as Artificial Neural Networks, being inspired by the biological neural networks that constitute animal brains. These supervised ML networks provide output responses without following specific instructions, learning from past experiences by judging how they performed on previously analyzed data.

In this thesis work we exploited two alternative ML-based techniques to put constraints on the matter density contrast, Ω_m , from the two-point correlation function of galaxies in the Baryon Oscillation Spectroscopic Survey (BOSS), which is part of the Sloan Digital Sky Survey (SDSS) III. The considered galaxy sample consists of almost 1.2 millions massive galaxies over an effective area of 9329 square degrees, in a comoving volume of 18.7 Gpc³ spanning the redshift range $0.2 < z < 0.75$. The two implemented ML models perform a Multi-Class Classification and a Regression Analysis, and have been trained with a large set of mock two-point correlation function measurements, obtained from log-normal mock catalogues with the same observational selections of the BOSS sample.

The constraints we obtained with the Classification and Regression models are $\Omega_m = 0.30 \pm 0.03$ and $\Omega_m = 0.307 \pm 0.006$, respectively. These results are

remarkably consistent with the ones from Alam et al. (2017), that have been obtained through a standard Bayesian analysis using the same galaxy catalogue. In particular, the Regression ML method implemented in this thesis work provides Ω_m constraints that are competitive with the ones obtained by state-of-the-art standard analyses, providing a new, independent confirmation of the standard Λ Cold Dark Matter (Λ CDM) cosmological framework.

This thesis work demonstrates that supervised ML techniques can be effectively applied to observed cosmological data such as galaxy catalogues, and not only to images or simulations. The main advantage of these techniques with respect to standard analyses is that they are likelihood-free statistical methods: no theoretical model has to be developed for the analyzed statistics, the algorithm itself can learn the representation of the input data that produce the best possible output. Moreover, once the model has been implemented and trained, the application on a new, previously unseen dataset requires a negligible computational time. On the other hand, the need of a huge number of examples to learn the correct representation of the data and the fact that this number exponentially increases when multiple cosmological parameters are considered, represent the main drawbacks of the method. Nevertheless, modern techniques and *free software* codes are nowadays available to address these issues. ML techniques can thus be considered as a powerful, independent alternative to standard cosmological methodologies.

Sommario

Una delle più grandi sfide della cosmologia del ventunesimo secolo è quella di stimare accuratamente i parametri che caratterizzano il modello cosmologico, e quindi di operare una corretta discriminazione tra i vari contesti cosmologici. La struttura a grande scala dell'Universo fornisce un importante metodo d'indagine per porre vincoli sui più importanti parametri cosmologici. Le surveys di galassie utilizzano questi oggetti cosmologici come traccianti per mappare la distribuzione di materia nell'Universo. Nel corso dell'ultima decina di anni, il numero di queste surveys è aumentato, sia utilizzando telescopi da terra, che telescopi spaziali. Questi strumenti sono in grado di produrre decine di terabytes di dati ogni giorno. Gli strumenti astronomici della prossima generazione come, ad esempio, Euclid, James Webb Space Telescope e il Large Synoptic Survey Telescope, aumenteranno drasticamente la quantità di dati disponibile per indagini cosmologiche. Questo scenario suggerisce che le tecniche di apprendimento automatico, meglio conosciuto con il termine inglese Machine Learning (ML), potranno avere un ruolo chiave nelle ricerche future, rivoluzionando tutte le tecniche di analisi di dati astronomici. Gli algoritmi di ML possono imparare delle specifiche rappresentazioni dei dati che vengono loro forniti e riescono, da queste, a produrre degli output, la bontà dei quali dipende dalla quantità dei suddetti dati forniti in input. Spesso ci si riferisce ai modelli di ML trattati in questa tesi con il termine Reti Neurali Artificiali, poichè originariamente furono ispirati dalle reti neurali biologiche che formano il sistema nervoso degli animali. Le reti di apprendimento supervisionato forniscono un output senza seguire specifiche istruzioni, ma imparando dalle passate esperienze, valutando la loro prestazione su dati che hanno precedentemente analizzato.

In questa tesi abbiamo utilizzato due differenti tecniche basate sul ML per porre vincoli sul parametro cosmologico che caratterizza il contrasto di densità della materia, Ω_m , partendo dalla funzione di correlazione a due punti di galassie misurata a partire dal catalogo della Baryon Oscillation Spectroscopic Survey (BOSS), che fa parte della terza fase della Sloan Digital Sky Survey (SDSS). Il campione di galassie preso in considerazione conta quasi 1.2 milioni di galassie massive, su un'area effettiva di 9329 gradi quadrati, in un volume comovente di 18.7 Gpc^3 e in un intervallo di redshift compreso tra $z = 0.2$ e $z = 0.75$. I due modelli di ML che

sono stati implementati sono in grado di eseguire una Classificazione a più classi e una Regressione, e sono stati allenati con un grande numero di funzioni di correlazioni a due punti misurate a partire da cataloghi mock generati assumendo una distribuzione log-normale di contrasti di densità. Questi cataloghi mock sono stati costruiti in modo da avere le stesse selezioni osservative del campione di BOSS.

I vincoli che abbiamo ottenuto con la Classificazione e la Regressione sono, rispettivamente, $\Omega_m = 0.30 \pm 0.03$ e $\Omega_m = 0.307 \pm 0.006$. Questi risultati sono considerevolmente in accordo con quelli di Alam et al. (2017), i quali sono stati ottenuti attraverso una analisi standard bayesiana, usando lo stesso catalogo di galassie. In particolare, l'algoritmo di Regressione che è stato implementato in questo lavoro di tesi ha fornito un vincolo su Ω_m che risulta essere competitivo rispetto a quelli ottenuti attraverso le analisi standard più all'avanguardia, fornendo quindi una nuova e indipendente conferma della validità del modello cosmologico standard Λ Cold Dark Matter (Λ CDM).

Questa tesi dimostra che le tecniche di apprendimento supervisionato possono essere applicate con efficacia a dati osservativi cosmologici, come cataloghi di galassie, e non solo a immagini o simulazioni. Il vantaggio principale di queste tecniche, rispetto all'analisi standard, è che queste consistono in un metodo statistico in cui non è necessario sviluppare un modello teorico dell'oggetto che si vuole analizzare: l'algoritmo impara da solo la rappresentazione dei dati di input che gli permette di fornire in output il miglior risultato possibile. Per questo motivo queste tecniche sono chiamate likelihood-free. Inoltre, una volta che il modello è stato costruito e allenato, la sua applicazione ad un nuovo dataset mai visto prima richiede un tempo di calcolo trascurabile. D'altro canto, il bisogno di un numero di esempi molto grande per poter imparare la corretta rappresentazione dei dati e il fatto che questo numero aumenti esponenzialmente nel caso in cui più parametri cosmologici vengano presi in considerazione, rappresentano i principali svantaggi di questo metodo. Ciononostante tecniche all'avanguardia e specifiche librerie numeriche sono oggi disponibili per fronteggiare queste problematiche. Le tecniche di ML possono quindi essere già oggi considerate come un'alternativa valida e indipendente rispetto ai metodi standard di analisi cosmologica.

Introduction

The cosmological framework that in the recent years has become the standard one for the scientific community is the Λ CDM model. This model is based on Einstein's theory of General Relativity (Einstein 1915), which provides the current best description of the gravitational force, the dominant interaction when large scales, like the cosmological ones, are considered. The Λ CDM model assumes the presence of a Cosmological Constant, Λ , in Einstein's field equations and in the existence of CDM particles, which can interact with each other and with the rest of their environment only gravitationally. Fritz Zwicky was the first one to hypothesize the existence of DM, after applying the virial theorem to the Coma Cluster and obtaining the evidence of unseen mass (Zwicky 1937). The Universe described by the Λ CDM model is experiencing an accelerated expansion, first proved by Perlmutter et al. (1998) and Reiss et al. (1998) using Type Ia SuperNovae as standard candles. The Universe is assumed to be composed for almost 70% of Dark Energy, represented by Λ in a General Relativity framework, and for approximately 25% of CDM. This means that what we are able to measure directly through photons is only 5% of the mass/energy content of the Universe.

The study of the distribution of collapsed structures, like galaxies and galaxy clusters, plays a key role in the cosmological investigation. The large-scale structure of the Universe is the result of an evolution that lasted about 14 billions years, and that was driven by the laws that govern the Universe and by their parameters. Therefore, understanding the evolution of the matter distribution is of fundamental importance to put constraints on the parameters of the Λ CDM model. The primordial fluctuations of matter density started growing, driven by gravity, when matter became neutral for the first time in cosmic history. This cosmic epoch is called recombination. According to the standard cosmological scenario, the evolution of density perturbations started from a Gaussian distribution, fully described by its variance. In configuration space, the variance of the field is called two-point correlation function and it is a function of the modulus of the distance between objects, r . In particular, the correlation function is defined as the excess of probability of finding two objects with a relative comoving distance r with respect to the case these objects were distributed randomly. At large scales,

the matter distribution can still be approximated as a Gaussian distribution, and thus, the two-point correlation function contains most of the information.

The two-point correlation function and its analogous in Fourier space, the power spectrum, have been the centre of many analyses of observed galaxy catalogues (e.g. Totsuji and Kihara 1969, Peebles 1974, Hawkins et al. 2003, Alam et al. 2016 and many others). Measuring the correlation function of galaxy catalogues is not an easy task. Many estimators of this statistic have been proposed, with the aim of minimizing both the scatter and the bias in the measurements (Davis and Peebles 1983, Hamilton 1993, Landy and Szalay 1993). The correlation function is affected by two main sources of distortions, that is dynamic and geometric distortions. The dynamic, or redshift-space distortions (RSD), are introduced if the peculiar velocities of the objects are neglected when measuring their comoving distances along the line of sight using the redshift of their spectra. The geometric distortions are caused instead by a wrong cosmological model assumption when computing the distances between objects, starting from observed coordinates.

The standard way to put constraints on cosmological parameters from the two-point correlation function is by fitting the measured statistic through a likelihood which should account for all kinds of distortions and uncertainties. In this thesis work we present two new likelihood-free ML-based techniques to put constraints on the matter density fraction, Ω_m , starting from the two-point correlation function of the BOSS galaxy catalogue.

ML algorithms can be used to provide reliable outputs that are based on previous experiences, and not on specific instructions (Samuel 1959, Goodfellow 2016). The level of reliability of the outputs increases when the amount of input data increases. This makes ML a powerful tool in a scenario such as the current one, in which real extra-galactic surveys and simulations are producing increasingly large amounts of cosmological datasets (e.g. Cole et al. 2005, Parkinson et al. 2012, Anderson et al. 2014).

This thesis work is organized as follows:

- In Chapter 1 we introduce the fundamental laws that govern the Universe, focusing on the description of the standard Λ CDM model and its observational constraints.
- In Chapter 2 we describe the evolution of the matter density distribution, from the primordial density fluctuations to the present day distribution of collapsed objects.
- In Chapter 3 we introduce clustering as the study of the spatial distribution of cosmic objects, like galaxies and galaxy clusters. Then we investigate which features of this distribution can be used to put constraints on the cosmological parameters.

- In Chapter 4 we describe the fundamentals of ML. We explain what an Artificial Neural Network is and what it is composed of. We then provide information about the training process, showing how it can teach the model to give trustworthy predictions and which are its hyperparameters that can be modified to improve the performances.
- In Chapter 5 we present the dataset we applied our ML models to, that is the BOSS galaxy catalogue. Then we explain how we measured its two-point correlation function, and how the same correlation function has been used to put constraints on cosmological parameters by Alam et al. (2017), using a standard Bayesian approach.
- In Chapter 6 we introduce the method to construct log-normal mock catalogues and its importance in cosmology. Then we present the training set and validation set we constructed.
- In Chapter 7 we give a detailed description of the implemented ML models and of the training processes they went through. Then, we present the results obtained from those trained models.
- In Chapter 8 we summarize the results of this thesis work and discuss about future perspectives.

Chapter 1

Theoretical cosmological background

1.1 Elements of General Relativity

Gravity is the most relevant interaction when large spatial scales are considered and it is so far best described by Einstein's Theory of General Relativity (Einstein 1915).

This theory is able to describe the curvature in the space-time structure induced by a distribution of mass and energy and it is this curvature that governs the motion of whatever has not null mass or energy. In a Relativity framework an event is described by one time coordinate and three spatial coordinates: $x^\mu = (t, x^1, x^2, x^3)$. Starting from this definition it is possible to define the distance between two events, that is:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad . \quad (1.1)$$

The metric tensor that describes a space-time structure where gravity is present is defined starting from its analogous in Special Relativity $\eta_{\alpha\beta}$:

$$g_{\mu\nu} \equiv \eta_{\alpha\beta} \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \quad , \quad (1.2)$$

where ξ^α is the set of coordinates of the reference frame that do not experience the gravitational force and sees in a uniform straight motion what the other reference frame (characterized by the set of coordinates x^μ) sees in free fall.

The Einstein's field equation for the Theory of General Relativity links the metric tensor to the energy-momentum tensor $T_{\mu\nu}$ which is the source term that takes into account the amount of energy and mass that is generating the gravitational field. The field equation is the following:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \quad , \quad (1.3)$$

where G is the universal gravitational constant, c is the speed of light, $R_{\mu\nu}$ is the Ricci Tensor, that comes from the contraction of the Riemann Tensor $R_{\mu\kappa\nu}^{\kappa}$, and R is the Ricci scalar, representing the contraction of the Ricci Tensor. Equation (1.3) actually consists of a set of 10 differential equations whose solutions are 10 different potentials that form the metric tensor. Everything that is on the right side of the field equation is called the source term and is a tensorial description of the distribution of mass and energy we expect is originating the gravitational field; while on the left side there are the terms that are strictly connected to gravity itself and to the curvature it causes in the space-time structure.

1.2 Friedmann-Lemaître-Robertson-Walker metric

The Einstein's notation states that whenever an index is repeated in a covariant coordinate and in a contravariant one a sum over all the values this index can assume must be performed. Equation (1.1) can be therefore written as

$$ds^2 = g_{00}dt^2 + 2g_{0i}dtdx^i + g_{ij}dx^i dx^j \quad , \quad i, j = (1, 2, 3) \quad (1.4)$$

where the time, mixed and spatial terms have been separated.

In order to have a metric tensor that can be used for cosmological purposes it is convenient to introduce the Cosmological Principle: at sufficiently large scales (that today are hundreds of Megaparsecs) the Universe can be considered homogeneous and isotropic, and so no point or direction can be considered privileged. The assumption of the impossibility to have a privileged reference frame automatically eliminates from Equation (1.4) the mixed term.

To determine the value of g_{00} it is sufficient to take into account the motion of a photon that (by definition) has $ds^2 = 0$. This leads to:

$$\begin{aligned} 0 &= g_{00}dt^2 - dl^2 \\ g_{00} &= \frac{dl^2}{dt^2} \\ g_{00} &= c^2 \quad , \end{aligned}$$

where dl^2 is the spatial term of equation (1.4) and its minus sign comes from the signature that has been chosen in this work: (+ - - -).

The analytical expression for the spatial distance, dl^2 , can be determined assuming three different 3D surface geometries: the Euclidean one, the hyperspherical one and the hyperbolic one. There is a general way to express this term in polar coordinates that already takes into account all of these cases:

$$dl^2 = a^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right] \quad , \quad (1.5)$$

where $a(t)$ is called the cosmic scale factor and has the dimensions of a length, while κ is the curvature parameter that can assume only three different values: -1 if the geometry of the Universe is hyperbolic (open Universe), 0 if it is Euclidean (flat Universe) and 1 if it is hyperspherical (closed Universe).

Therefore it is now possible to write the metric as:

$$ds^2 = c^2 dt^2 - a^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right] . \quad (1.6)$$

Equation (1.6) is called Friedmann-Lemaître-Robertson-Walker metric and has to be used if the assumption of homogeneity and isotropy is valid, in a reference frame that can be described in 3D polar coordinates.

1.3 Distances in Cosmology and the Hubble's Law

The proper distance is defined as the distance between two events measured in a reference frame where the separation in time (dt) is null. Without loss of generality, it can be chosen a reference frame where also $d\theta = d\phi = 0$. Putting all these assumptions and choices together, the proper distance (d_{pr}) is defined as follows:

$$d_{pr} = \int_0^r a(t) \frac{dr'}{\sqrt{1 - \kappa r'^2}} = a(t) f(r) ; \quad (1.7)$$

in the last term of Eq. (1.7) the time-dependent part and the spatial-dependent one have been explicitly separated. The analytical form of $f(r)$ depends on the value of κ , thus on the geometry of the Universe:

$$f(r) = \begin{cases} \sinh^{-1}(r) & \text{if } \kappa = -1 \\ r & \text{if } \kappa = 0 \\ \sin^{-1}(r) & \text{if } \kappa = 1 \end{cases} . \quad (1.8)$$

The comoving distance is defined as the proper distance evaluated at the present time. When a time-dependent variable has "0" as subscript it means that it has been calculated at $t = t_0$, the present time. So the comoving distance becomes

$$d_c = a_0 f(r) = a_0 f(r) \frac{a(t)}{a(t)} = \frac{a_0}{a(t)} d_{pr} . \quad (1.9)$$

The comoving distance usually turns out to be useful because it is time-independent, the same cannot be said for the proper distance. Calculating the time derivative of the proper distance, it is possible to find a very important result:

$$\dot{d}_{pr} = \dot{a}(t) f(r) = \dot{a}(t) f(r) \frac{a(t)}{a(t)} = \frac{\dot{a}(t)}{a(t)} d_{pr} = H(t) d_{pr} , \quad (1.10)$$

where $H(t)$ is called Hubble's parameter and the equation itself is known as the Hubble's Law, which states that the relative speed between two cosmic objects is proportional to their relative distance. By evaluating $H(t)$ at $t = t_0$ we obtain H_0 : the Hubble constant. Ivanov et al. (2019) found $H_0 = (67.89 \pm 1.06) \text{ km s}^{-1} \text{ Mpc}^{-1}$ by fitting the galaxy power spectrum of BOSS galaxy catalogue.

Let us now consider a photon emitted at time t_{em} and observed in another position at t_{obs} . For a photon $ds^2 = 0$, and so, starting from Eq. (1.6) and assuming that the reference frame is the one that leads us to Eq. (1.7), we have that

$$\int_{t_{em}}^{t_{obs}} \frac{cdt}{a(t)} = f(r) \quad . \quad (1.11)$$

The same thing can be said for another photon, emitted at time $t_{em} + \delta t_{em}$ and observed at $t_{obs} + \delta t_{obs}$, and so we have also

$$\int_{t_{em} + \delta t_{em}}^{t_{obs} + \delta t_{obs}} \frac{cdt}{a(t)} = f(r) \quad . \quad (1.12)$$

Due to the fact that $f(r)$ does not depend on t , we reach the necessary conclusion that the first member of Eq. (1.11) and the one of Eq. (1.12) must be equal to each other. If we now assume δt_{em} and δt_{obs} to be small enough in order to be able to consider $a(t)$ as a constant during the integrations, the evaluation of the two integrals leads us to

$$\frac{t_{obs}}{a_{obs}} - \frac{t_{em}}{a_{em}} = \frac{t_{obs}}{a_{obs}} + \frac{\delta t_{obs}}{a_{obs}} - \frac{t_{em}}{a_{em}} - \frac{\delta t_{em}}{a_{em}} \quad . \quad (1.13)$$

Considering that $\lambda = c\delta t$, where λ is the wavelength of the signal, it can be derived that

$$1 + z = \frac{a_{obs}}{a_{em}} \quad , \quad (1.14)$$

where $z = (\lambda_{obs} - \lambda_{em})/\lambda_{em}$ is the redshift. If we now consider a photon emitted at time t and observed at the present day, Eq. (1.14) becomes

$$1 + z = \frac{a_0}{a(t)} \quad . \quad (1.15)$$

Equation (1.15) shows us that the redshift of a signal is inherently linked to the cosmic scale factor and therefore to the distance between two objects. The positive sign of the Hubble parameter tells that a increases with the passing of time, and so the redshift can be considered as a measure of both distance and time; the higher the redshift of an object is, the further the object is from the observer.

1.4 Friedmann models

We already found a metric (Eq. (1.6)) that takes into account the assumptions of homogeneity and isotropy; considering the Universe as a perfect fluid, its energy-momentum tensor can be written as

$$T_{\mu\nu} = -pg_{\mu\nu} + (p + \rho c^2)u_\mu u_\nu \quad , \quad (1.16)$$

where p and ρ are the pressure and the density, respectively, and u_i are the elements of the 4-velocity of a fluid's element.

It is now possible to use the Friedmann-Lemaître-Robertson-Walker metric (Eq. (1.6)) and the energy momentum of a perfect fluid (Eq. (1.16)) in Einstein's field equation (Eq. (1.3)) in order to obtain two differential equations useful to derive the time evolution of the cosmic scale factor $a(t)$; these equations are called first and second Friedmann's Law, and their analytical expressions are the following:

$$\ddot{a} = -\frac{4}{3}\pi G \left(\rho + \frac{3p}{c^2} \right) a \quad (1.17)$$

$$\dot{a}^2 + \kappa c^2 = \frac{8}{3}\pi G \rho a^2 \quad ; \quad (1.18)$$

from now on the fact that a depends on time t is given as implicit. Starting from one of the two Friedmann Laws it is possible to obtain the other one using the adiabatic condition $dU = -pdV$, where U is the internal energy of the whole system and V its volume. The same condition can also be written as $d(\rho c^2 a^3) = -pda^3$.

The Friedmann's Laws, firstly presented by Aleksandr Friedmann in 1922 (Friedmann, 1922), are a direct solution of the Einstein's field equation for General Relativity. By that time the universe was thought to be static ($\ddot{a} = \dot{a} = 0$). But if the static conditions are assumed in Eq. (1.17) this new condition arises:

$$\rho = -\frac{3p}{c^2} \quad , \quad (1.19)$$

which is physically not acceptable because both the density and the pressure are positive quantities, by definition.

For this reason Einstein modified his field equation in the simplest way possible: introducing a constant (Λ) multiplied by $g_{\mu\nu}$ in order to have a correct tensorial formalism. Equation (1.3) therefore became

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \quad , \quad (1.20)$$

and Λ took the name of Cosmological Constant.

This new version of the field equation of General Relativity can be written just like the previous one, but with a modified energy-momentum tensor:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}\tilde{T}_{\mu\nu} \quad , \quad (1.21)$$

with $\tilde{T}_{\mu\nu} = T_{\mu\nu} + \frac{\Lambda c^4}{8\pi G}g_{\mu\nu}$. Having the same analytical expression, solving Eq. (1.21) using Friedmann-Lemaître-Robertson-Walker metric (Eq. (1.6)) will be exactly the same as solving Eq. (1.3) and will again lead to Friedmann's first and second Laws, but instead of ρ there will be $\tilde{\rho} = \rho + \rho_\Lambda = \rho + \frac{\Lambda c^2}{8\pi G}$, and instead of p there will be $\tilde{p} = p + p_\Lambda = p - \frac{\Lambda c^4}{8\pi G}$.

When, in 1929, Edwin Hubble discovered the expansion of the Universe (Hubble, 1929), Einstein himself removed the Cosmological Constant from his formula, because the static Universe assumption was not needed any more. The Cosmological Constant was reintroduced (this time in the source side of (1.3), at the right of the equal sign) to explain the accelerated expansion of the Universe proved by Perlmutter et al. (1998) and by Riess et al. (1998) with the observations of Type Ia SuperNovae.

If we consider the second Friedmann Law and divide every term by a^2 we can obtain the following equation:

$$H^2 \left(1 - \frac{\rho}{\rho_{crit}} \right) = -\frac{\kappa c^2}{a^2} \quad , \quad (1.22)$$

where we have defined the critical density as

$$\rho_{crit} \equiv \frac{3H^2}{8\pi G} \approx 1.9 \cdot 10^{-29} h^2 \frac{g}{cm^3} \quad , \quad (1.23)$$

with $h = H/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$.

From Eq. (1.22) it can be easily seen that if the Universe (considered as a perfect fluid) has a density equal to the critical density, it means that κ must be null and so the Universe is to be considered flat; similarly, if $\rho < \rho_{crit}$ the Universe is open, while if $\rho > \rho_{crit}$ the Universe is closed.

So far the Universe has been considered as a perfect fluid. This assumption will not be discarded, but it is important to take into account the different components that are part of this fluid. Each of these components (non-relativistic matter, radiation and relativistic matter and Dark Energy, which is represented by the Cosmological Constant in a General Relativity framework) has a different Equation of State (EoS) that links its pressure to its density:

- Non-relativistic matter (or simply matter): $p = Nk_B T$; this pressure is negligible compared to the one of the other components;

- Radiation and relativistic matter: $p = \frac{1}{3}\rho c^2$;
- Dark Energy: $p = -\rho c^2$, as can be easily derived from the definitions of \tilde{p} and $\tilde{\rho}$.

It is useful to write down the EoS for all of the components in a unique way:

$$p = w\rho c^2 \quad , \quad (1.24)$$

so that the components will be now identified by a specific value of w , that will therefore be 0 for matter, $\frac{1}{3}$ for radiation and relativistic matter and -1 for the Dark Energy.

Starting from the adiabatic condition, it is possible to find a relation between the density of each component, considered separately, ρ_w and a (or the redshift z):

$$\rho_w = \rho_{w,0} \left(\frac{a}{a_0} \right)^{-3(1+w)} \propto (1+z)^{3(1+w)} \quad . \quad (1.25)$$

It is now clear how the different components evolve differently in density through cosmic time. Figure 1.1 shows how the different trends of the densities can be used to divide the cosmic time into three main epochs, each of them dominated by a particular component. The moment in which there has been the matter-radiation crossover point is called the time of equivalence (t_{eq}).

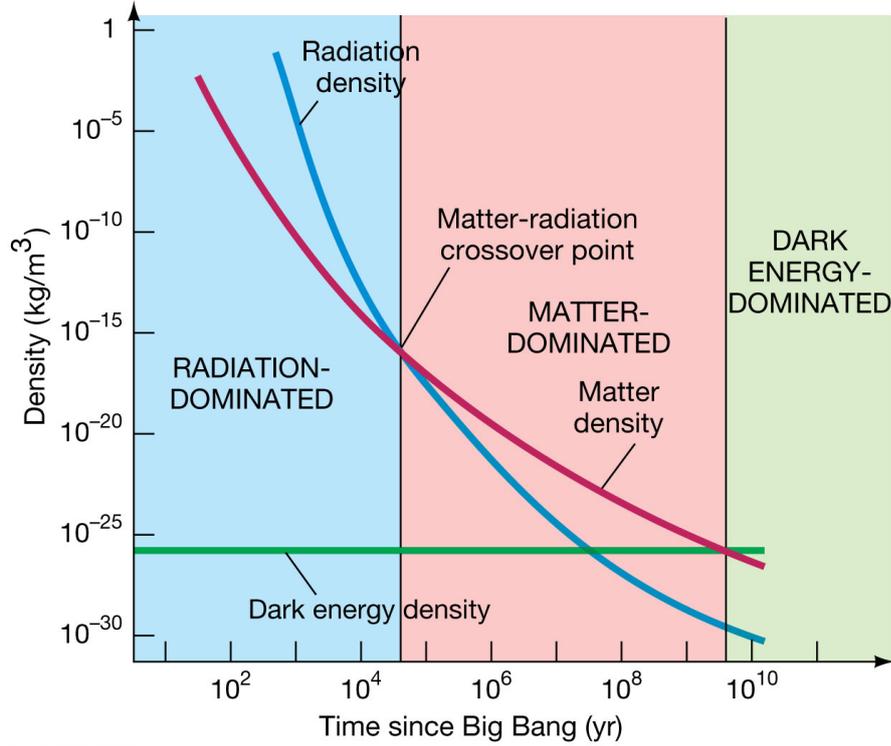
From the combination of Eq. (1.25), Eq. (1.22) and Eq. (1.15), considering the Universe as formed by multiple components, it is possible to obtain a law that describes the dependence of the Hubble parameter on the redshift, z :

$$H^2(z) = H_0^2(1+z)^2 \left[1 - \sum_i \Omega_{0,w_i} + \sum_i \Omega_{0,w_i}(1+z)^{1+3w_i} \right] \quad , \quad (1.26)$$

where it has been defined a new variable, $\Omega_w \equiv \rho_w/\rho_{crit}$, that indicates what fraction of the critical density is equal to the density of a specific component characterized by w . The first two terms of the squared bracket of Eq. (1.26) are called the curvature density parameter. If this parameter is null then it means that $\rho_{tot} = \sum_i \rho_{w_i} = \rho_{crit}$, and so the Universe is flat. For high values of z the curvature density parameter is negligible if compared to $\sum_i \Omega_{0,w_i}(1+z)^{1+3w_i}$. So it can be asserted that every Universe that follows Friedmann's equations can be considered flat at high redshift.

If we re-write the second Friedmann law for a single component, we obtain

$$\ddot{a} = -\frac{4}{3}\pi G\rho(1+3w)a \quad ; \quad (1.27)$$



© 2011 Pearson Education, Inc.

Figure 1.1: Evolution through cosmic time of the densities of the cosmic components. It is clear from the image that the Dark Energy began to be the dominant part only in recent times, so its contributes can be neglected when high redshifts are concerned. Image from <https://pages.uoregon.edu/jimbrau/>. Credits to 2011 Pearson Education, Inc.

if the Universe is dominated by a component with $w > -\frac{1}{3}$ (as we are supposing it is for high redshift, when ρ_Λ is negligible) then $\ddot{a} < 0$ and the Universe is in a phase of decelerated expansion. This means that there has been a precise moment when \dot{a} was null. This is called the Big Bang. The existence of this singularity comes directly from the initial conditions of homogeneity and isotropy.

1.5 Flat, open and closed models

We have already demonstrated how at high redshift every model that follows Friedmann's Laws behaves like a flat one.

However, for lower values of z the curvature becomes more and more important, making it possible to see differences in the evolution of the cosmic scale factor between an open model of the Universe and a closed one.

An open Universe ($1 - \sum_i \Omega_{0,w_i} > 0$) is characterized by an asymptotically

constant growth of $a(t)$. A closed Universe ($1 - \sum_i \Omega_{0,w_i} < 0$) is fated to reach a maximum expansion, and then begin to shrink until it reaches again a singularity similar to the Big Bang, which has been called the Big Crunch. If the Universe maintains its flatness with a dominant component with $w > -\frac{1}{3}$ then will keep on growing, like it happens in a flat model, but the growth rate will be lower. In Figure 1.2 the three models are shown and compared. It can be seen that for

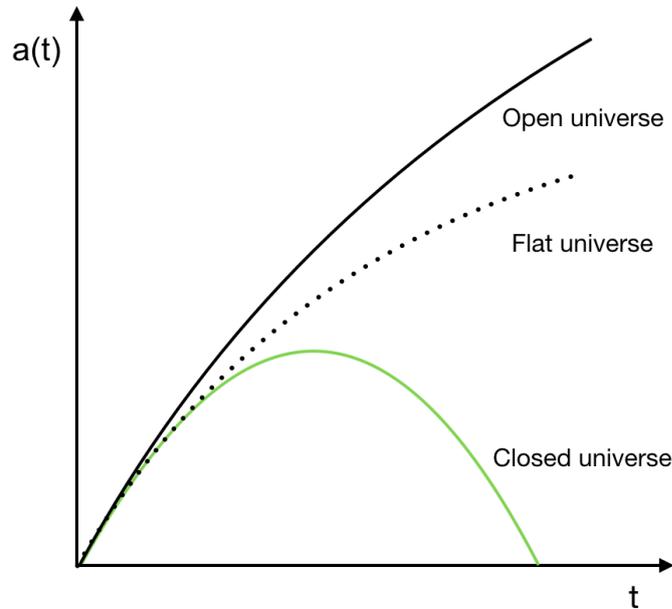


Figure 1.2: Comparison between the trends of the cosmic scale factor $a(t)$ in a closed Universe (green line), flat Universe (black dots) and open Universe (black line).

$t \rightarrow 0$ all the three different models behaves the same as far as the trend of $a(t)$ is concerned; and when the difference becomes evident, at a fixed time t , the size of the closed Universe will always be lower than the size of the flat one, which is lower than the size of the open one.

1.6 The Λ CDM Model

The cosmological model with the lowest number of free parameters, in agreement with observations, within the uncertainties, and with the Theory of General Relativity is called Standard Cosmological Model. According to this model the Universe is flat and it is composed mainly by Λ and CDM; the densities of these two components are of the same order of magnitude at the present time.

The fact that the Universe is flat comes as a necessary final condition of the inflation models that describe a particular moment of the cosmic time when the Universe experienced a huge expansion. An observational proof of the flatness of the Universe comes from the Cosmic Microwave Background (CMB), in particular from the position of the first peak of its angular spectrum. The CMB is a thermal radiation field that permeates the whole Universe, described by a Black Body radiation with $T \approx 2.726K$ at the present time. It was emitted during the recombination, that is the moment when the Universe became neutral for the first time.

From the temperature of the CMB it is also possible to obtain a measure of the radiation density fraction at the present time:

$$\Omega_{0,r} \approx 10^{-5} \quad , \quad (1.28)$$

that results to be very small if compared to the density fractions of other components.

There are multiple ways to evaluate the matter density fraction of the Universe, for example by measuring the radial velocities of objects in a galaxy to evaluate the total gravitational mass of the galaxy, or the radial velocities of galaxies in a cluster in order to obtain the total gravitational mass of the cluster. There are also other techniques such as Gravitational Lensing, that can measure the total mass of an object from the deflection on the path of light that it causes.

All these measurements lead to the same conclusion concerning the matter density fraction:

$$\Omega_{0,m} \approx 0.3 \quad . \quad (1.29)$$

Let us now define as baryonic matter all the matter that interacts with electromagnetic fields, like dust and gas. The density fraction of the baryonic matter, $\Omega_{0,b} \approx 0.04$, results to be only a small fraction of the total density fraction of matter. This fact suggests the existence of another type of matter: the DM, that has only gravitational interactions. It was the Swiss astronomer Fritz Zwicky who in 1937 first proposed the existence of DM to explain the motion of galaxies in the Coma Cluster (Zwicky 1937), but a direct measurement of a DM particle is yet to be done, though there are many hypothesis on how this particle should be. DM can be divided in two big categories: Hot and Cold. The DM is called Hot if at the time of its decoupling it is relativistic, Cold if it is not. The fact that the main part of the matter component of the Universe is Cold implies a “Bottom-Up” scenario for the formation of compact objects, that means that smaller objects are formed first, and then they merge forming bigger ones. This scenario is favored by observations.

If the Universe is flat then the sum of the density fractions of all the components of the Universe must be equal to 1, but if we consider only matter and radiation

we obtain $\Omega_{0,m} + \Omega_{0,r} \approx 0.3$. The remaining 0.7 is the density fraction of the Dark Energy, represented by Λ , the dominant component of the present Universe. The fact that $\Omega_{0,\Lambda} \approx 0.7$ comes not only from the flatness of the Universe, but also from a very important measure already cited in section 1.4. Perlmutter et al. (1998) and Riess et al. (1998) measured the value of what is called the deceleration parameter, $q = -\frac{\ddot{a}a}{\dot{a}^2}$, using Type Ia SuperNovae as standard candles; they found that this parameter is $q_0 \approx -0.55$. This implies that \ddot{a} must be positive, and therefore the expansion of the Universe is accelerated. Moreover it can be demonstrated that

$$q = \frac{\Omega_m}{2} - \Omega_\Lambda \quad , \quad (1.30)$$

and replacing the known values of q_0 and $\Omega_{0,m}$ we obtain

$$\Omega_{0,\Lambda} \approx 0.7 \quad , \quad (1.31)$$

that is exactly what was expected for a flat Universe. The Dark Matter, expressed in terms of the Cosmological Constant, is therefore considered to be the main component of the Universe, and to be also what causes its accelerated expansion.

The estimation of the Hubble constant is an open issue for the Λ CDM model, because it appears to be a tension between the results derived from different observations. The H_0 tension is the discrepancy between the measures of the Hubble constant through CMB fitting and the results that comes from the use of empirical distance ladders such as Classical Cepheids (CCs) (Riess et al. 2019).

Recently, Breuval et al. (2019) adopted the second Gaia release (GDR2) parallaxes of the companions of 23 Galactic CCs as a proxy for the parallaxes of those CCs in order to have a new measure of H_0 . Figure 1.3 shows that, for low offsets between the considered Cepheid and the companion whose parallax measure is used to measure the value of H_0 , this value is consistent with the Planck Collaboration result (Planck Collaboration, 2018). This could be a step in the direction of the solution of the H_0 tension, but in order to be sure, GDR2 parallaxes for the Cepheids are needed.

The Λ CDM model has thus some open issues, but most of the measurements, made in the recent years with sophisticated technologies and techniques able to acquire a constantly increasing amount of data, are in agreement with its predictions.

In the near future, new instruments and telescopes are planned to acquire quantities of high-accuracy data that are without precedents. One example of these missions is Euclid.

The European Space Agency (ESA) mission Euclid (whose launch is scheduled for June 2022) will play an important role in the study of the Cosmological Constant, measuring the shape and the redshift of galaxies and galaxy clusters up to

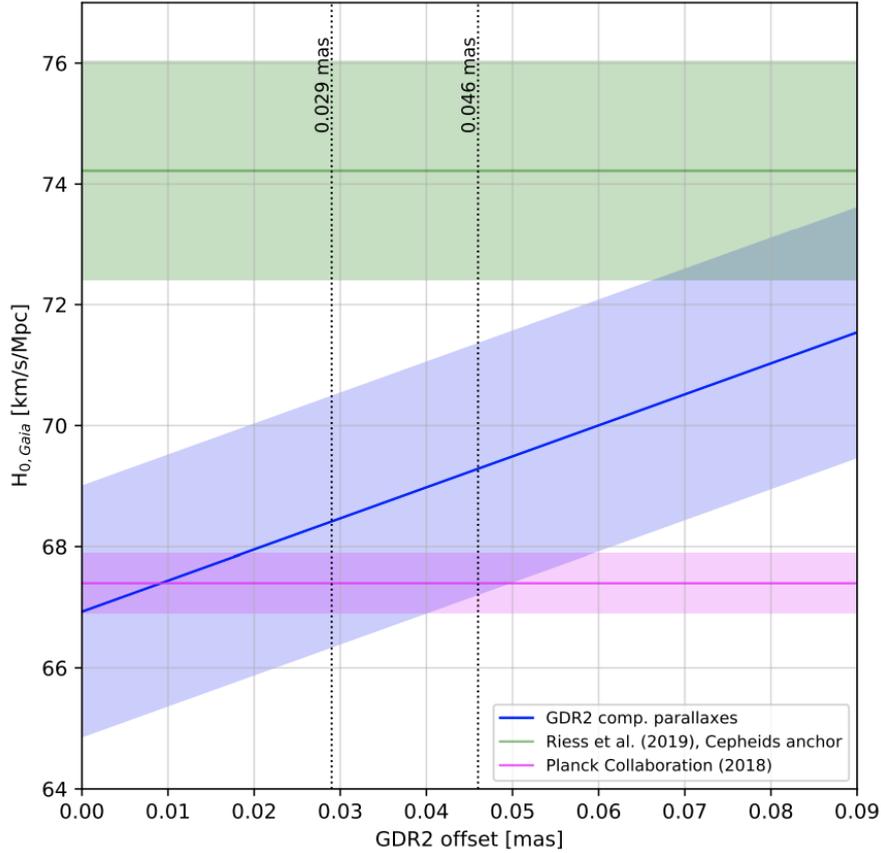


Figure 1.3: Values of the Hubble constant evaluated with three different methods: CMB modelling (pink), Classical Cepheids using HST/FGS parallaxes (green) and CCs with GDR2 parallaxes of the companions, as a function of the offset between the Cepheid and the companion (blue) (Breuval et al. 2019).

$z \approx 2$. The goal is to put constraints on the distance-redshift relationship and to have a better understanding about the formation of cosmic structures (Laurejjs et al. 2011), Amendola et al. 2018). The sky coverage of Euclid will be of 15.000 square degrees of extragalactic sky for wide surveys, during a lifetime of 7 years. Three deep surveys are planned, too. These surveys will reach two magnitudes deeper then the wide ones, for a total of 40 square degrees. The telescope is a three mirror Korsch configuration with a 1.2 m diameter and a 24.5 m focal length.

The visible instrument (VIS) is composed by a 6x6 matrix of 4096x4132 12 micron pixels CCDs and is equipped with a broad band filter that covers from 550 nm to 900 nm. The mean resolution of VIS is about 0.23 arcseconds and the sensitivity will be of 25 mag.

A second instrument, the Near Infrared Spectrometer and Photometer (NISP) is composed of a matrix of 4×4 2040×2040 18 micron pixel TIS detectors. The photometric channel of this instrument is equipped with three broad band filters (Y, J and H), covering a total wavelength range that goes from 900 nm to 2000 nm. The spectroscopic channel is equipped with 4 infrared grisms, 3 red grisms and one blue grims, but with no slits.

Euclid will observe about 10 billion astronomical objects and will measure the spectroscopic redshift of 50 million sources.

Chapter 2

Cosmic structures formation

To draw conclusions about which cosmological model is the closest to the observed one, is of primary importance to understand how the cosmic structures, like galaxies and galaxy clusters, were formed and how fast these overdensities can grow.

The starting point is the recombination ($z \approx 1100$). Since then, the baryonic matter has been able to increase the amplitude of its density fluctuations, driven by gravity, falling into DM halos that had started forming at the time of equivalence.

By the time of recombination the distribution of baryonic matter was highly homogeneous. It can be demonstrated that the amplitude of the perturbations in the density field was of the same order of the temperature fluctuations that can be seen in the CMB:

$$\frac{\delta\rho}{\rho} \propto \frac{\delta T}{T} \approx 10^{-5} \quad . \quad (2.1)$$

The task is now to describe how these perturbations evolved in an expanding Universe, starting from recombination, into what we see at the present time, that is a density field with fluctuations of the order of 10^2 on the scale of collapsed objects.

2.1 Jeans lenght

How small fluctuations can evolve into collapsed structures is well described by the Jeans Theory, first presented at the beginning of the 20th Century by the English physicist James Jeans (Jeans, 1902).

Let us now consider a homogeneous fluid, which has to satisfy the following conditions:

- Conservation of mass, that is guaranteed if the continuity equation is satisfied:

$$\frac{\partial\rho}{\partial t} + \nabla(\rho\vec{v}) = 0 \quad , \quad (2.2)$$

where \vec{v} is the velocity of the considered fluid element;

- Euler equation, which guarantees the conservation of the momentum:

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla) \cdot \vec{v} = -\frac{1}{\rho} \nabla p - \nabla \Phi \quad , \quad (2.3)$$

where Φ is the Newtonian gravitational potential that is present at the position of the fluid element;

- Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho \quad . \quad (2.4)$$

The equation of state of this fluid is $p = p(S, \rho)$, where S is the entropy, but we will further consider only adiabatic perturbations, so $dS/dt = 0$. Therefore the pressure will be considered as a function of the density only.

We can now assume to know a solution of this kind:

$$\begin{cases} \rho = \rho_b = const \\ \Phi = \Phi_b = const \\ p = p_b = const \\ \vec{v} = \vec{0} \end{cases} \quad . \quad (2.5)$$

The set of equations (2.5) (in which the subscript b means background) implies a static Universe with no perturbations of any kind.

This cannot be a real solution because from Eq. (2.4) we have that $\Phi = const \iff \rho = 0$. But we are not really interested in the static background situation, we are interested in the perturbations.

Let us hence add some small perturbations to our static solution, that is almost negligible if compared to the background values of our four variables:

$$\begin{cases} \rho = \rho_b + \delta\rho \\ \Phi = \Phi_b + \delta\Phi \\ p = p_b + \delta p \\ \vec{v} = \delta\vec{v} \end{cases} \quad . \quad (2.6)$$

Imposing that the values of the variables expressed in Eq. (2.6) still are solutions of Eq. (2.2), Eq. (2.3) and Eq.(2.4), and discarding all the second-order and higher-order terms of the perturbations, we obtain the following set of equations:

$$\begin{cases} \frac{\partial \delta\rho}{\partial t} + \rho_b \nabla \delta\vec{v} = 0 \\ \frac{\partial \delta\vec{v}}{\partial t} = -\frac{v_s^2}{\rho_b} \nabla \delta\rho - \nabla \delta\Phi \\ \nabla^2 \delta\Phi = 4\pi G \delta\rho \end{cases} \quad . \quad (2.7)$$

The element v_s that appears in the second line of Eq. (2.7) is the sound speed in the fluid we are considering.

It is now possible to assume that every perturbation has a wavelike analytical expression, $\delta\rho(r, t) = \delta\rho_k e^{ikr+i\omega t}$. This assumption leads to a new system of equations whose determinant is null if

$$\omega^2 = v_s^2 k^2 - 4\pi G\rho_b \quad . \quad (2.8)$$

If we now define the Jeans length as

$$\lambda_J = \frac{2\pi}{k_J} = v_s \sqrt{\frac{\pi}{G\rho_b}} \quad , \quad (2.9)$$

we can see that if the dimension of a perturbation is bigger than λ_J , then the amplitude of this perturbation will grow exponentially with time, otherwise it will keep oscillating with no significant increase nor decrease.

2.2 Perturbations in an expanding Universe

If the expansion of the Universe is taken into account, we can still assume a solution similar to Eq. (2.6), but now the velocity \vec{v} will have two terms: the peculiar motion of the fluid element, \vec{v}_p , and the velocity that comes from the Hubble flow:

$$\vec{v} = \frac{d\vec{r}}{dt} = \frac{d(a\vec{x})}{dt} = a \frac{d\vec{x}}{dt} + \frac{da}{dt} \vec{x} = Hr + a \frac{d\vec{x}}{dt} = Hr + \vec{v}_p \quad . \quad (2.10)$$

In Eq. (2.10) \vec{r} is the proper position of the fluid element, while \vec{x} is the comoving one. The peculiar motion of the fluid element will be treated as a perturbation with respect to the Hubble flow. We can now proceed as we did in the previous section, imposing that our new perturbed solution still satisfies the continuity equation (2.2), the Euler equation (2.3) and the Poisson equation (2.4). The result is a system of three equations, analogous to Eq. (2.7):

$$\begin{cases} \frac{\partial \delta\rho}{\partial t} |_{\vec{x}} + \frac{\rho_b}{a} \nabla_{\vec{x}} \vec{v}_p = 0 \\ \frac{\partial \vec{v}_p}{\partial t} |_{\vec{x}} + H \vec{v}_p = -\frac{v_s^2 \nabla_{\vec{x}} \delta}{a} - \frac{1}{a} \nabla_{\vec{x}} \delta\Phi \\ \frac{1}{a^2} \nabla_{\vec{x}}^2 \delta\Phi = 4\pi G\rho_b \delta \end{cases} \quad , \quad (2.11)$$

where $\delta = \delta\rho/\rho_b$.

Again, it is possible to assume a wavelike analytic expression for all the perturbations:

$$\begin{cases} \dot{\delta}_k + \frac{ikv_{p,k}}{a} = 0 \\ \dot{v}_{p,k} + \frac{\dot{a}}{a}v_{p,k} = -\frac{ik}{a}[v_s^2\delta_k + \delta\Phi_k] \\ \delta\Phi_k = -\frac{4\pi G\rho_b\delta_k a^2}{k^2} \end{cases} . \quad (2.12)$$

By combining the three equations of this system, the following differential equation arises:

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \left[\frac{k^2}{a^2}v_s^2 - 4\pi G\rho_b \right] \delta_k = 0 . \quad (2.13)$$

Equation (2.13) is called dispersion equation and it has different solutions, again depending if the dimension of the perturbation is bigger or smaller than the Jeans length:

- If $\lambda < \lambda_J$ ($k > k_J$) the solution of the differential equation is a wavelike function whose amplitude does not increase through time;
- If $\lambda > \lambda_J$ ($k < k_J$) the spatial part of the exponential that describes the wave keeps on being imaginary, but the time dependent one does not and so the amplitude of the perturbation will change through time.

In this second case, the general solution will be a linear combination of two solutions: the increasing one, δ_+ , and the decreasing one, δ_- .

We are interested in the increasing solution, because it is the one that can lead to the formation of collapsed structures. For a generic cosmological model we have:

$$\delta_+(z) = -H(z) \int_0^z \frac{(1+z')}{a_0^2 H^3(z')} dz' . \quad (2.14)$$

Since Eq. (2.14) has no general solutions, usually an approximated parametric one is used, defining the growth factor f :

$$f \equiv \frac{d \ln \delta_+}{d \ln a} \approx \Omega_m^\gamma + \frac{\Omega_\Lambda}{70} \left(1 + \frac{\Omega_m}{2} \right) , \quad (2.15)$$

where the value of α depends on which gravity model is assumed. If General Relativity is assumed to be the right theory for gravitation then $\gamma \approx 0.545$ (Linder et al. 2003). Constraining the value of γ from observations provides a powerful test for General Relativity. As an example, in Alam et al. (2016), measurements from the Planck satellite mission and galaxy redshift surveys over the last decade have been used for such a purpose and the result was $\gamma = 0.612 \pm 0.072$, compatible with what is predicted by General Relativity. Another parameter that can

be constrained to gain information about the value of γ is $f\sigma_8$, where σ_8 is the normalization of the power spectrum (see section 2.3) at $z = 0$. In Pezzotta et al. (2017) was found $f\sigma_8 = 0.55 \pm 0.12$ at $z = 0.6$ and $f\sigma_8 = 0.40 \pm 0.11$ at $z = 0.86$ by modelling the anisotropic galaxy clustering measured on the VIPERS survey.

As it can be seen from Eq. (2.15), the formation of cosmic structures depends strongly on the density fraction of matter and poorly on the Cosmological Constant. This enforces the concept that this formation is driven by matter (both baryonic and dark) through gravity.

2.3 Correlation Function and Power Spectrum

The first fluctuations in the density field were created by the inflation, that is supposed to have perturbed the structure of the Universe, that is its metric. This perturbation can be seen as a perturbation of the gravitational potential Φ , from a Newtonian point of view. The Poisson equation (2.4) provides the link between Φ and the density fluctuation δ .

The evolution of the density fluctuations is the key phenomenon to be described to model the statistical properties of the large-scale structure of the Universe. A frequentist approach may sound difficult to exploit, because we have only one Universe from which we can draw our samples. Anyhow, the ergodic hypothesis allows to use different volumes of the Universe just as they were an ensemble of Universes, as long as these volumes are not superimposed and big enough to be independent from each other (Ellis, Maartens and MacCallum, 2012).

Primordial density fluctuations are imaginary numbers with random phase, and they have a Gaussian distribution. An analysis of the fluctuations' statistical properties can be done both in real and Fourier space. In the first case we will consider $\delta(\vec{x})$, while in the second $\delta(\vec{k})$ or simply δ_k ; the relation that links these two quantities is

$$\delta(\vec{x}) = \frac{1}{(2\pi)^3} \int \delta(\vec{k}) e^{-i\vec{k}\cdot\vec{x}} d\vec{k} . \quad (2.16)$$

A very important statistic is the Correlation Function $\xi(r)$, defined as follows:

$$\xi(r) = \langle \delta(\vec{x}) \delta(\vec{x} + \vec{r}) \rangle , \quad (2.17)$$

where the angle brackets mean that the correlation function is defined as an average over all the positions \vec{x} and all the points that are distant from \vec{x} by a vector whose magnitude is r . The fact that the correlation function does not depend on the direction of \vec{r} comes directly from the isotropy assumption.

It is also possible to express the correlation function using $\delta(\vec{k})$ instead of $\delta(\vec{x})$, by combining Eq. (2.16) with Eq. (2.17):

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) e^{i\vec{k}\cdot\vec{r}} d^3k \quad , \quad (2.18)$$

where we have defined the power spectrum $P(k)$ (using the three-dimensional Dirac Delta δ_D^3) using the following expression:

$$P(k) \delta_D^3(\vec{k} + \vec{k}') = \frac{\langle \delta(\vec{k}) \delta(\vec{k}') \rangle}{(2\pi)^3} \quad . \quad (2.19)$$

Equation (2.18) shows that the Correlation Function and the Power Spectrum are connected by a Fourier transform, while from Eq. (2.19) we can see that the Power Spectrum is proportional to $\langle |\delta_k|^2 \rangle$, that is the mean of the squared amplitude of all the perturbations characterized by a wave number k , independently of their position \vec{x} .

Inflation models predict that the primordial density fluctuations have a Gaussian distribution, which has, by definition, a null mean. Let us now define the variance of this distribution:

$$\sigma^2 = \langle \delta^2(\vec{x}) \rangle = \frac{1}{V} \int \langle |\delta(\vec{x})|^2 \rangle d^3x \quad , \quad (2.20)$$

where V is the total volume of the Universe and the average is taken from a sample of different realizations of V .

It can be demonstrated, starting from Eq. (2.20) and using the Parseval's relation, that

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^\infty P(k) k^2 dk \quad . \quad (2.21)$$

Most of the time a direct measure of $\delta(\vec{x})$ is impossible to perform and the only available observable is the distribution of cosmic objects in a finite volume.

It is also impossible to obtain the measure of a continuous density field. Therefore a set of small volumes shall be defined. If galaxies are used as tracers of the density fluctuations, their density contrast in the i -th volume can be written as follows:

$$\delta_{gal}(V_i) = \frac{N_{gal}(V_i) - \overline{N_{gal}}(V)}{\overline{N_{gal}}(V)} \quad , \quad (2.22)$$

where $N_{gal}(V_i)$, is the measured number of galaxies in the volume V_i , and $\overline{N_{gal}}(V)$ is the expectation value of the number of galaxies in V .

In a similar way we can define the density contrast for the total mass density fluctuations in volume V_i :

$$\delta_M(V_i) = \frac{M(V_i) - \overline{M}(V)}{\overline{M}(V)} , \quad (2.23)$$

where the expected and measured number of galaxies have been replaced by the total mass (again, expected and measured) in the volume V .

The linear bias factor, b , is used when the observed distribution of galaxies is taken into account to trace the total mass distribution, and it is defined as follows:

$$b = \frac{\delta_{gal}}{\delta_M} . \quad (2.24)$$

The total mass discrete density fluctuation, δ_M , is the convolution between the continuous density fluctuation, $\delta(\vec{x})$, and a window function characterized by a radius R :

$$\delta_M(\vec{x}) = \delta(\vec{x}) \otimes W(\vec{x}, R) . \quad (2.25)$$

One example of window function is the Top Hat function:

$$W(R) = \frac{3}{4\pi R^3} \Theta \left(1 - \frac{|x - x'|}{R} \right) , \quad (2.26)$$

where Θ is the Heaviside step function, x is the magnitude of \vec{x} and x' is the center of the step function.

It is now possible to define the mass variance, using Eq. (2.25) and the convolution theorem, as

$$\sigma_M^2 \equiv \langle \delta_M^2 \rangle = \frac{1}{(2\pi)^3} \int P(k) \hat{W}^2(\vec{k}, R) d^3\vec{k} , \quad (2.27)$$

where $\hat{W}(\vec{k}, R)$ is the Fourier transform of the window function.

The mass variance depends on the value of R in the window function:

- if $R \rightarrow \infty$ ($M \rightarrow \infty$) the window function is filtering over the whole Universe, so the mass variance will be null: $\sigma_M^2 \rightarrow 0$;
- if $R \rightarrow 0$ ($M \rightarrow 0$) the window function is not filtering at all, and the mass variance will be the same as the punctual variance: $\sigma_M^2 \rightarrow \sigma^2$.

2.4 Evolution of the Power Spectrum

The primordial perturbations in the density field were formed during the inflation with no privileged scale. It is then rightful to assume that the primordial power spectrum's analytical expression is a power law:

$$P(k) = Ak^n \quad . \quad (2.28)$$

Combining Eq. (2.21) with Eq. (2.28) we obtain

$$\sigma^2 = \frac{A}{2\pi^2} \int_0^\infty k^{n+2} dk \quad . \quad (2.29)$$

In Section 2.2 we learnt that the growth of perturbations is described by the increasing solution of Eq. (2.13), that is Eq. (2.14). This increasing solution does not depend on the scale of the perturbation, so in the linear growth regime we have

$$\delta_k(z) = \delta_{k,i} \delta_+(z) \quad , \quad (2.30)$$

where $\delta_{k,i}$ is the initial value of the density fluctuation characterized by the wavenumber k . Considering that $P(k) \propto \langle |\delta_k|^2 \rangle$ the evolution through time of the power spectrum will be

$$P(k, t) = P_i(k) \delta_+^2(t) \quad . \quad (2.31)$$

Combining Eq. (2.21) with Eq. (2.31) it is possible to obtain

$$\sigma^2(t) \propto \delta_+^2(t) k^{n+3} \quad . \quad (2.32)$$

Taking into account that $k \propto R^{-1}$ and that $M \propto V \propto R^3$, Eq. (2.32) becomes

$$\sigma^2(t) \propto \delta_+^2(t) M^{-\frac{n+3}{3}} \quad . \quad (2.33)$$

If we now consider a fixed scale R and take into account Eq. (2.33) and the perturbation of the Newtonian gravitation field, we have

$$\delta\Phi \propto \frac{G\delta M}{R} \propto \frac{G\delta\rho R^3}{R} \propto \sigma R^2 \propto M^{\frac{n-1}{6}} \quad . \quad (2.34)$$

Inflation models tell us that the squared amplitude of a certain primordial fluctuation of the gravitational field should not depend on the mass of this perturbation. Thus the spectral index, n , has to be equal to 1, and the analytical form of the primordial power spectrum, or Zel'dovič Spectrum, becomes:

$$P(k) = Ak \quad , \quad (2.35)$$

where A is a free parameter of inflation models and has to be constrained by observations.

The primordial power spectrum expressed in Eq. (2.35) evolves through time as it is shown in Eq. (2.31), with no dependencies on the scale; this means that all perturbations grow with the same rate, as it is shown in Figure 2.1.

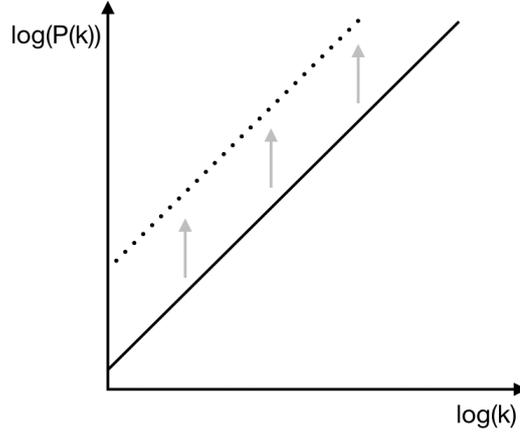


Figure 2.1: Time evolution of the Zel'dovich Spectrum. The solid line is the primordial power spectrum, while the dotted line is the power spectrum at a certain time $t > 0$. This Figure shows how the time evolution does not depend on the scale, and the power spectrum remains parallel to itself. Due to the fact that the spectral index is equal to one, the line forms a 45° angle with the x-axis.

It is convenient here to introduce the Cosmological Horizon, $R_H(t)$, which is the radius of the sphere which contains all the events that can be causally connected with the center of the sphere:

$$R_H(t) = a(t) \int_0^t \frac{c}{a(t')} dt' . \quad (2.36)$$

For scales bigger than the Cosmological Horizon the only possible interaction is the gravitational one, so all the density fluctuations bigger than $R_H(t)$ grow in amplitude through time.

Before the equivalence, only the perturbations bigger than the Cosmological Horizon were able to grow because the pressure of the radiation (the dominant component of the Universe by that time) hindered the collapse of smaller fluctuations.

The Cosmological Horizon is an increasing function of time. So as time went on (from the end of the inflation to the equivalence) always less fluctuations were

able to grow in amplitude and only the largest ones kept on growing like in Figure 2.1.

This phenomenon modified the shape of the power spectrum, by reducing its value for the scales smaller than the Cosmological Horizon evaluated at the equivalence, $R_H(t_{eq})$.

The Transfer Function, $T(k)$, is introduced to describe all these scale-dependent modifications in the shape of $P(k)$:

$$P(k, t_{eq}) = AkT^2(k) \quad , \quad (2.37)$$

where $T^2(k)$ is proportional to k^{-4} for the small scales ($k > 2\pi/R_H(t_{eq})$) and to k^0 for the big ones ($k < 2\pi/R_H(t_{eq})$).

Figure 2.2 shows the Power Spectrum at the present day (Tegmark et al. 2004). The figure compares data from different observational probes with a line that represents the best fit of the data points coming from SDSS galaxy clustering, assuming the linear CDM power spectrum of Eisenstein and Hu (1999), with fixed baryon fraction $\Omega_b/\Omega_m = 0.17$ and Hubble parameter $h = 0.72$.

After the equivalence, every density fluctuation bigger than the Jeans Scale can grow, so the power spectrum keeps on evolving as stated in Eq. (2.31) for every scale, as long as it is in the linear regime ($\sigma^2 < 1$).

The first scales that enter the non-linear regime are the smallest ones. The non-linear regime is very hard to treat analytically. Approximate analytical solutions can be obtained if the following three main assumptions are made:

- Null initial velocity field;
- Flat Universe composed only by matter;
- Perfectly spherical collapse.

Outside the linear regime the distribution of density fluctuations is no more Gaussian. Once a certain scale k reaches and crosses the edge of the linear regime ($\delta_k \approx 1$) the fluctuations can rapidly grow in amplitude up to values of the order of 10^2 .

This growth requires a numerical approach to be described. Even in the case of the analytical solutions of a spherical collapse, the value reached by δ_k strongly depends on the cosmological model that has been assumed.

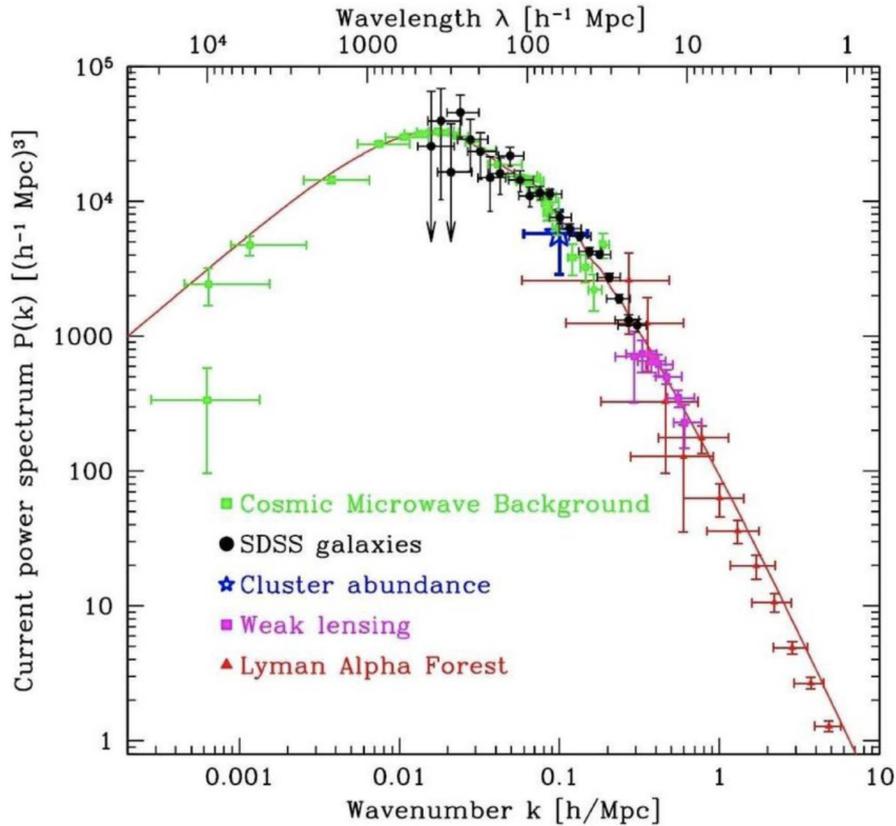


Figure 2.2: The matter power spectrum $P(k)$ at the present time. For values of k close to $0.01 Mpc/h$ it is possible to see the change in the shape, because that one is considered to be the scale of the Cosmological Horizon by the time of recombination. In the plot are shown results coming from different observables at different scales. Credits to Tegmark et al. (2004).

2.5 N-body simulations

A widely used numerical approach to describe the non-linear evolution of cosmic structures is the one of N-body simulations. Firstly, the number of particles that are used as tracers of the density field is set. Then the initial conditions have to be defined, that are the positions of the particles that describe the density field of the Universe at high redshifts. Finally, the physical laws the particles must follow have to be decided and implemented.

The main physical law the particles must follow is the gravitational force. The

three equations that the $i - th$ particle shall satisfy for each time step dt are

$$\begin{cases} \vec{F}_i = GM_i \sum_{j \neq i} \frac{M_j}{r_{ji}^3} \vec{r}_{ji} \\ \vec{a}_i = \frac{\vec{F}_i}{M_i} = \frac{d\vec{v}_i}{dt} \\ \vec{v}_i = \frac{d\vec{x}_i}{dt} \end{cases}, \quad (2.38)$$

where \vec{F}_i is the gravitational force acting on the $i - th$ particle, M_i the mass of that particle, \vec{a}_i its acceleration and \vec{v}_i its velocity, \vec{r}_{ji} is the distance between the $i - th$ and the $j - th$ particles, and r_{ji} its amplitude.

In the followings we present four different methods to compute \vec{F}_i :

- Particle-particle: the force on the $i - th$ particle is computed as the sum of the forces exerted by each of the others particles, computed separately. This is the most precise and accurate method, but it is certainly very expensive from the computational point of view;
- Particle-Mesh: in this case the volume in which the simulation takes place is divided into cells, forming a grid. The most relevant quantities (for example the number density of particles) are evaluated for every cell, and not for every particle. The Poisson equation (2.4) is solved in the Fourier space. The force on the $i - th$ particle in the real space is then computed using a fast Fourier transform algorithm. If compared to the Particle-Particle method, this method is less precise (spatial resolution is lost because of the use of a grid) but faster;
- Particle-Particle Particle-Mesh: this third method is a hybrid of the previous ones. It is necessary to define a certain distance R . The gravitational force on the $i - th$ particle will then be calculated with the Particle-Particle method for distances smaller than R , otherwise a Particle-Mesh approach is used. For obvious reasons, both accuracy and speed of this method lie between the ones of the Particle-Particle and the Particle-Mesh methods. The bigger the value chosen for R is, the closer the Particle-Particle Particle-Mesh method will be to the Particle-Particle one;
- Hierarchic tree: the volume of the simulation is divided into regions so that every region contains a certain number of particles and the gravitational force on the $i - th$ particle exerted from distant particles is calculated using the coordinates of the centre of mass of the region these particles are in.

If other phenomena such as hydrodynamic interactions are required to be taken into account, then other kinds of laws that govern the behaviour of particles must

be implemented. Hydrodynamic simulations are the best choice when one is interested in galaxy formation or baryonic matter flows, but they are very expensive from the computational point of view.

Finally, there is also an alternative way to simulate the distribution of galaxies, which consists in taking a DM simulation and placing galaxies inside the DM halos or sub-halos following a Halo Occupation Distribution model. The role of Halo Occupation Distributions is to provide descriptions of how galactic matter is distributed inside halos of given size and mass (Zheng et al. 2005). These distributions are built using real data or hydrodynamical simulations.

Chapter 3

Clustering

We now focus on the study of how cosmic structures, such as galaxies or galaxy clusters, are distributed in space. These objects can be used as tracers of the underlying total matter density distribution. In this chapter the role of the two-point correlation function will be discussed, as long as the distortions that can affect it.

3.1 The two-point correlation function

Let us consider N point masses uniformly distributed in a volume V . For every comoving position \vec{x} in V we can define the number density of objects in that position as

$$n(\vec{x}) = \bar{n} + \delta n(\vec{x}) = \bar{n}[1 + \delta(\vec{x})] \quad , \quad (3.1)$$

where $\bar{n} = N/V$ is the average density and $\delta(\vec{x})$ is the fractional number density fluctuation.

The two-point correlation function (defined in Eq. (2.17)) in terms of the fractional number density fluctuation can hence be written as

$$\xi(|\vec{x}_1 - \vec{x}_2|) = \langle \delta(\vec{x}_1)\delta(\vec{x}_2) \rangle = \frac{\langle n(\vec{x}_1)n(\vec{x}_2) \rangle}{\bar{n}^2} - 1 \quad . \quad (3.2)$$

If we now consider a sub-volume $\delta V(\vec{x})$ centered in \vec{x} , then the number of objects in that sub-volume will be $n(\vec{x})\delta V(\vec{x})$ and so the value of $n(\vec{x})\delta V(\vec{x})/N$ can be interpreted as the probability that one object is in $\delta V(\vec{x})$.

Combining Eq. (3.1) with Eq. (3.2), the joint probability of having an object in $\delta V(\vec{x}_1)$ and another one in $\delta V(\vec{x}_2)$ can be written as follows:

$$P_{12} = \left\langle \frac{n(\vec{x}_1)\delta V(\vec{x}_1)}{N} \frac{n(\vec{x}_2)\delta V(\vec{x}_2)}{N} \right\rangle = [1 + \xi(r)] \frac{\delta V(\vec{x}_1)}{V} \frac{\delta V(\vec{x}_2)}{V} \quad , \quad (3.3)$$

where $r = |\vec{x}_1 - \vec{x}_2|$.

The probability of having an object in $\delta V(\vec{x}_i)$ is $\delta V(\vec{x}_i)/V$, so from Eq. (3.3) it is clear that the conditional probability of having an object in $\delta V(\vec{x}_2)$, given that another object is in $\delta V(\vec{x}_1)$, is

$$P_{2|1} = [1 + \xi(r)] \frac{\delta V(\vec{x}_2)}{V} . \quad (3.4)$$

The correlation function thus represents the excess of probability, compared to a random uniform distribution, of finding two objects (for example galaxies) at a distance r from each other (Vittorio 2017).

It is clear from this last definition that if $\xi(r) > 0$ then galaxies will tend to cluster on the scale of r , otherwise, if $\xi(r) < 0$, galaxies will be anti-correlated on that scale and will avoid to be at distance r from each other.

The expectation number of galaxies within a range r from another galaxy can be computed as

$$\langle N(< r) \rangle = \int_V dP_{2|1} = \int_V \bar{n} dV [1 + \xi(r)] = \frac{4}{3} \pi r^3 \bar{n} + 4\pi \int_0^r \xi(r') r'^2 dr' . \quad (3.5)$$

If $r \rightarrow \infty$, $\langle N(< r) \rangle$ is expected to be equal to the left term of the last member of Eq. (3.5), so $4\pi \int_0^\infty \xi(r') r'^2 dr' = 0$. It is then clear that there is an integral constraint which tells that for every positive value of $\xi(r)$ there must be another one that will be negative in order to compensate it. One last constraint can be derived from Eq. (3.4): there can be no value of r for which $\xi(r) < -1$ because a probability cannot be negative.

3.2 Estimators of the correlation function

It is now time to look at how the correlation function can be estimated from a sample of galaxies. The first step is to create a random distribution of objects with the same selection effects of the analyzed catalogue. A not sufficiently accurate random catalogue will lead to a wrong evaluation of the correlation function, so particular attention is needed in its creation.

Let us indicate the number of objects in the real catalogue with N_D and the number of objects in the random catalogue with N_R . The numbers of pairs will therefore be $N_{DD} = N_D(N_D - 1)/2$ and $N_{RR} = N_R(N_R - 1)/2$, for the real and the random catalogue, respectively. The number of pairs that can be obtained combining the two catalogues is $N_{DR} = N_D N_R$. Finally, let $DD(r)$ be the number of pairs at distance r present in the real catalogue, $RR(r)$ its analogous for the random catalogue and $DR(r)$ the number of objects in the random catalogue with a distance r from the objects of the real catalogue.

Here we present three of the most used estimators for $\xi(r)$:

- The first proposed estimator, also called the natural estimator, is the one proposed in 1983 by Davies and Peebles (Davies and Peebles, 1983):

$$\xi(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1 ; \quad (3.6)$$

- Ten years later, in 1993, two other estimators were proposed. The first of them was proposed by Hamilton (1993) to avoid the main defect the estimator presented in Eq. (3.6) had. This defect was that the uncertainty of that estimator at large scales was limited by the uncertainty in the mean density. Hamilton therefore proposed the following estimator, proving that it did not suffer from such limitation:

$$\xi(r) = \frac{N_{DR}^2}{N_{DD}N_{RR}} \frac{DD(r)RR(r)}{DR(r)^2} - 1 ; \quad (3.7)$$

- The other estimator firstly proposed in 1993 by Landy and Szalay (1993) is the following:

$$\xi(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2 \frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} . \quad (3.8)$$

In their article they demonstrated how this estimator has a Poissonian variance and should be used any time the error on the pair counts is assumed to be Poissonian, too.

Hamilton's estimator and Landy and Szalay's one have been demonstrated to perform almost indistinguishably on the Hubble volume simulation by the Virgo Supercomputing Consortium (Kersher et al. 2000). For further details see Keihänen et al. (2019).

3.3 Measures of two-point correlation function

The first galaxy two-point correlation function measures came from Totsuji and Kihara, (1969) and, independently, from Peebles, (1974). These measures pointed out that for scales between 10 kpc/ h and 10 Mpc/ h the correlation function can be well approximated by a single power law:

$$\xi(r) \propto \left(\frac{r}{r_0} \right)^{-\gamma} . \quad (3.9)$$

The power law behaviour was extended by Gott and Turner, (1979) and Maddox et al. (1990) to smaller and bigger scales, respectively.

The values of the parameters of Equation (3.9) depend on the selection that is used on the tracers. For example, as far as galaxies are concerned, the values of γ and r_0 , that describe the spatial distribution of such objects, depend on the redshift (e.g. Marulli et al. 2013), the luminosity (e.g. Davis et al. 1988, Marulli et al. 2013), the stellar mass (e.g. Li et al. 2006, Marulli et al. 2013), the colour (e.g. Willmer et al. 1998), the environment (e.g. Abbas and Sheth, 2006) and the morphological (e.g. Davis and Geller, 1976) and spectral type (e.g. Brown et al. 2003).

Generally speaking it can be said that luminous and massive galaxies with morphologies dominated by the bulge and red colours cluster more strongly, in agreement with the scenario of hierarchical growth of cosmic structures (Kaiser, 1984).

As an example of comparison between two different analyses of the galaxy correlation function, let us consider the works by Hawkins et al. (2003) and Zehavi et al. (2004). In both cases the two-point correlation function has been fitted with a single power law up to $r = 30$ Mpc/ h . Hawkins et al. (2003) obtained, as best-fit values, $\gamma = 1.67 \pm 0.003$ and $r_0 = (5.05 \pm 0.26)$ Mpc/ h , while the results from Zehavi et al. (2004) are $\gamma = 1.80$ and $r_0 = 5.77$ Mpc/ h .

The differences in the best-fit parameters are due to the different galaxy selection of the two surveys used in the experiments: Two degrees Field Galaxy Redshift Survey for Hawkins et al. (2003) and SDSS for Zehavi et al. (2004), which are blue-magnitude and red-magnitude selected surveys, respectively.

3.4 Geometric distortions

In a typical observed galaxy catalogue the positions of the objects are expressed with three coordinates: Right Ascension (RA), Declination (Dec) and redshift. The first two coordinates are the angular ones, while the redshift tells us how far the galaxy is on the line of sight (Equation (1.10)).

In order to estimate the correlation function we need to calculate the distances between objects starting from their coordinates, but this needs an assumption on the underlying geometry of the Universe, and hence on the background cosmological model (Marulli et al. 2012).

The distance r between two objects can be separated into r_{\perp} and r_{\parallel} : the component perpendicular to the line of sight and the one parallel to it, respectively. Each of these components will be affected by the assumption of a cosmological model.

The connection between $r_{\perp,1}$ and $r_{\perp,2}$, where the subscripts “1” and “2” indicate

two different assumed cosmological models, is

$$r_{\perp,1} = \frac{D_{A,1}(z)}{D_{A,2}(z)} r_{\perp,2} \quad , \quad (3.10)$$

where D_A is the angular diameter distance, defined from the comoving distance

$$D_C(z) = c \int_0^z \frac{dz'}{H(z')} \quad (3.11)$$

as follows:

$$D_A(z) = \frac{c}{H_0(1+z)\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k} D_C(z)\right) \quad , \quad (3.12)$$

with $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$.

Analogous calculations can be done for the component of r which is parallel to the line of sight. Assuming two different cosmological models, the relation in this case is

$$r_{\parallel,1} = \frac{H_2(z)}{H_1(z)} r_{\parallel,2} \quad . \quad (3.13)$$

If the real shape of an object, or of a statistic is known, but a different one is observed because of geometric distortions, it is possible to put constraints on the background cosmological parameters. This cosmological analysis is known as the Alcock and Paczyński test (Alcock and Paczyński 1979). Specifically, this test consists in searching for the values of the parameters that turn the distorted shape of what is observed into the shape it is assumed it should have.

One example can be the iso-correlation contours (i.e. the curves along which the two-point correlation function has the same value) as a function of r_\perp and r_\parallel . Under the assumption of isotropy, these contours should be circles, but the observed ones may be not, because r_\perp and r_\parallel do not depend in the same way on the assumed cosmological model. Hence, an Alcock Paczyński test can be performed to spot what cosmological model assumption can turn the observed shape of these curves into circles.

3.5 Redshift-space distortions

When dealing with observed galaxies, the redshift is not caused only by the Hubble flow, but also by the peculiar motion those objects might have, for example due to the gravitational interaction with other galaxies in a galaxy cluster. Therefore the observed redshift, z_{obs} , differs from the cosmological one, z .

The observed redshift can be expressed as a function of the cosmological one as follows:

$$z_{obs} = z + \frac{v_{\parallel}}{c}(1+z) + \frac{\sigma_v}{c} \quad , \quad (3.14)$$

where v_{\parallel} is the component of the peculiar velocity parallel to the line-of-sight and σ_v is the random error of the observed redshift.

The cosmological redshift is known in simulations, but it is generally now known in the case of observed catalogues of objects. In this case the observed redshifts are used to compute the comoving distances of all the objects through Eq. (3.11), neglecting the last two terms of Eq. (3.14). This approximation introduces the so called redshift-space distortions (RSD) (Kaiser 1987).

These distortions can be seen, for example, when plotting the iso-correlation contours of $\xi(r_{\perp}, r_{\parallel})$ in the $(r_{\perp}, r_{\parallel})$ space, as it is shown in Figure 3.1, where the coordinate system $(s_{\perp}, s_{\parallel})$ is used instead of $(r_{\perp}, r_{\parallel})$ for the redshift space (the space where RSD are present). For low values of r_{\perp} the contours are lengthened along the r_{\parallel} axis because of the peculiar motion of cosmic objects due to their gravitational interaction with their surroundings. These distortions of the iso-curvature contours for low values of the perpendicular component of the distance are often referred to as ‘‘Fingers of God’’ due to their peculiar shape (Jackson 1972). The flattening of the iso-correlation contours on large scales is due to the dynamic flow of matter towards overdense regions, the so called bulk flow.

In linear theory the RSD can be expressed as a function of $\beta \equiv f(z)/b(z)$, where $f(z)$ is the growth factor and $b(z)$ the linear galaxy bias (Kaiser, 1987). The ratio between the redshift-space and real-space correlation functions can be approximately expressed using the large-scale limit of the Kaiser model:

$$\frac{\xi(s)}{\xi(r)} = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2 \quad . \quad (3.15)$$

3.6 Baryon Acoustic Oscillations

Baryon Acoustic Oscillations (BAO) are an oscillation pattern in the power spectrum that has been generated before decoupling ($z_{dec} \approx 1100$) by the sound waves in the baryon-photon fluid (Eisenstein and Hu 1998).

Before decoupling, baryons were not able to freely follow the gravitational collapse because they were still coupled with radiation, whose pressure opposed such a collapse. In that fluid, density fluctuations in the form of sound waves were able to propagate up to the scale that took the name of sound horizon, $r_s \approx 150$ Mpc, before they were dissipated. Such a scale remained imprinted in the present matter distribution and can be measured from clustering.

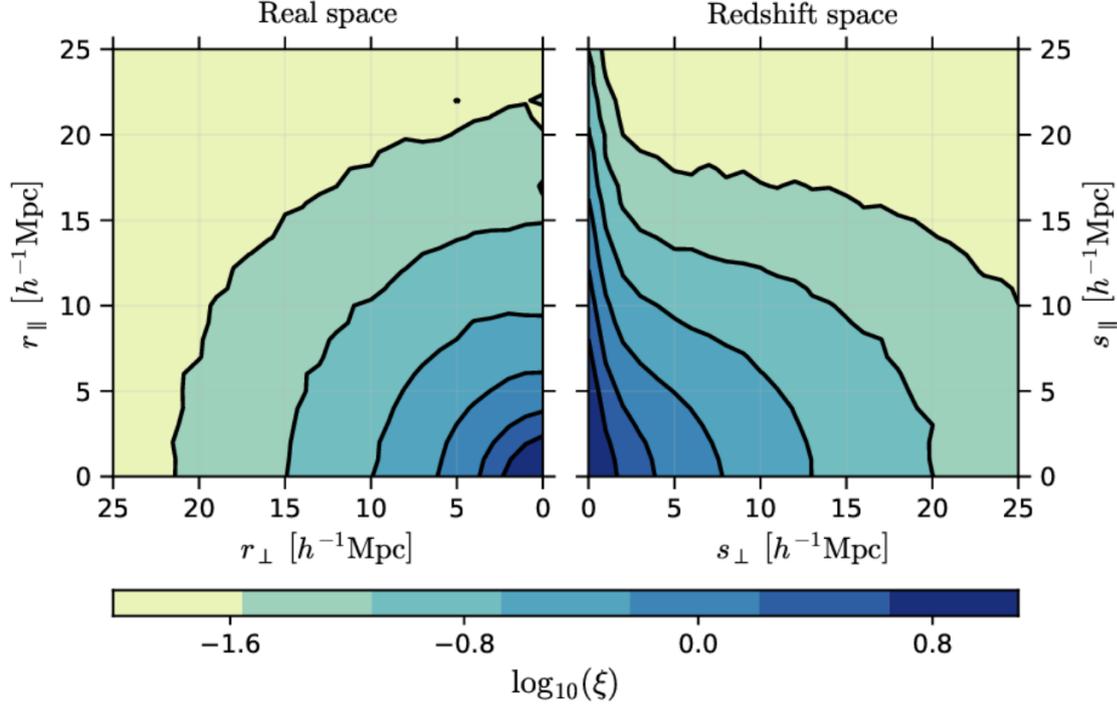


Figure 3.1: Example of iso-correlation contours in real space (left panel) and redshift space (right panel) for DM particles in a simulation at $z=0$. This simulation follows the formation of the large-scale structure in a cubic box with a side of $1200 \text{ Mpc}/h$ and considering 1024^3 particles (Pillepich et al. 2010). In the right panel both the large-scale and small-scale effects of the RSD are visible. These are the gravity-driven flow of matter towards overdense regions at large scales and the stretching of the contours along the line of sight at small scales, due to the peculiar motion of objects, gravitationally interacting with their surroundings. Credits to Kuruvilla and Porciani (2018).

The BAO can therefore be used as a standard ruler to extract cosmological information by modelling the power spectrum and the two-point correlation function (Sanchez et al. 2013).

As an example case, in Veropalumbo et al. (2016) the BAO of galaxy clusters was used for the first time to constrain the redshift-distance relation. It has also been used to put constraints on the parameters of the cosmological model: $H_0 = 64_{-9}^{+14} \text{ km}/(\text{s}\cdot\text{Mpc})$, $\Omega_k = -0.015_{-0.36}^{+0.34}$ and $\omega = -1.01 \pm 0.44$.

Figure 3.2 shows that BAO appears in the two-point correlation function as a single peak at $s \approx 100 \text{ Mpc}/h$, that represents an excess of clustering compared to its adjacent scales.

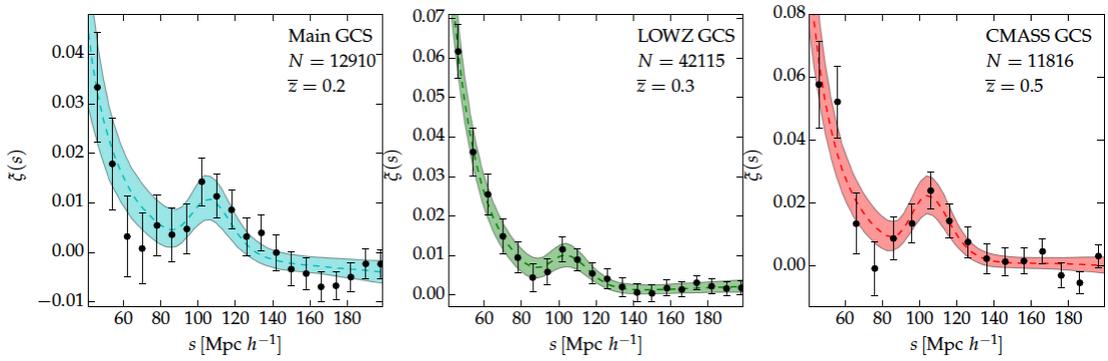


Figure 3.2: Three examples of two-point correlation functions in redshift space, calculated from three different galaxy cluster catalogues: Main Galaxy Cluster Sample (left panel), LOWZ Galaxy Cluster Sample (central panel), CMASS Galaxy Cluster Sample (right panel). The errorbars are computed with the lognormal mock method, that will be presented in detail later in this work, and the shaded areas are the 68% posterior uncertainties that came from a Monte Carlo Markov Chain analysis. In each panel, N is the number of objects of the catalogue and \bar{z} the median redshift. Credits to Veropalumbo et al. (2016).

Chapter 4

Fundamentals of Machine Learning

The term ML was used for the first time by Arthur Samuel (1959) and it refers to the development of algorithms, programs and models that computers can use to draw conclusions starting from previous experience instead of specific instructions (Goodfellow, 2016).

ML techniques rely both on the ability of the programmer, in building the fittest model for its specific purpose and feeding it with the correct representation of the data, and on the quantity of data that it is used to train the model.

During the last decades, increasing large data samples became available, concerning large-scales structures traced by galaxies and galaxy clusters (Cole et al. 2005, Parkinson et al. 2012, Anderson et al. 2014). Ground-based and space telescopes produce every day tens of terabytes of data, the complexity and quality of which are without precedents (Ravanbakhsh et al. 2017).

A few years from now, new instruments like Euclid, the Large Synoptic Survey Telescope and the James Webb Space Telescope will increase even more the amount of data that will be available for future researches. ML can therefore play a key role in modern cosmology.

In fact, ML is already used for many astrophysical and cosmological purposes, for example in image recognition in Gravitational Lensing studies (Ostrovski et al. 2016, Hezaveh et al. 2017), in measuring photometric redshift using galaxy images (Hoyle, 2016), in morphological galaxy classification (De La Calleja and Fuentes, 2004), in measuring the dynamical mass of galaxy clusters (Ntampaka et al. 2015) and in exoplanet transit detection (Schanche et al. 2018).

Image and pattern recognition is not the only task that a ML algorithm can be used for. Every kind of data can be used as input for the model.

In this chapter we will look at how a ML model is built, and how it works.

4.1 Supervised Machine Learning

Supervised ML is the task of learning a function that maps a vector \vec{x} of the input space to an output based on previously seen input-output pairs (Russel 2016).

The alternative to supervised learning is unsupervised learning, that concerns models that are able to subdivide the data into small groups, according to peculiarities and resemblances that the model finds, without any previously given information about the output that should be produced. Unsupervised learning will not be described in detail in this work.

The main elements of supervised ML are: Labels, Features, Examples and Models.

Labels are what we want to predict, the output of the model. It can be a continuous variable (for example the value of y in the case of a linear regression) or a discrete one (for example in a classification algorithms).

The features are the input variables. The choice of the features to feed the algorithm with is very important. Features have to be a good representations of our data, every aspect of our dataset must be well described by the features we choose. If the features we use are not able to give a proper overview of the dataset's peculiarities, then, after the training is done, the model will not be able to generalize and to give correct predictions over previously unseen inputs. For instance, if the purpose is to constrain cosmological parameters starting from a catalogue of observed galaxies, then useful features might be the coordinates of those galaxies, or their masses, or their absolute magnitude, but not their names.

With the term Example we denote a particular instance of data, \vec{x} . For example, let's say we want to predict the mass of a galaxy as a function of its position in the cluster it lives in, then an example will be a vector containing the three coordinates of a particular galaxy.

Examples can be divided into two categories: labeled examples and unlabeled examples. Labeled examples are composed by both features and labels. These are used to train the model, so they have to represent all the possible scenarios in order to have a trained model able to generalize.

Once the model has been trained, unlabeled examples (composed only by features) are used to make predictions on their label. If the label of an unlabeled example is known by the user, then it is possible to test the performance of the model comparing this label with the prediction.

The last key element of supervised ML is the model. The model is the element that provides the relationship between the features and the labels. As it will be

better described in the next section, a model is composed by layers, and each layer has a certain number of parameters and plays a different role in the creation of the relationship mentioned above. The ensemble of layers of a model is called architecture.

Deep Learning is a class of ML algorithms that uses complex architectures, formed by multiple layers, to progressively extract higher level features from the input data (Deng et al. 2014).

Part of the ML analysis consists in understanding which model's architecture is the fittest for the purpose. If the model is too simple it will not be able to learn much from the features, so it will generalize well to previously unseen examples, but the performance will not be as good as it might be. Otherwise, if the model is over-complicated (or, using the ML language, is too deep), then it will perform very well on the training set, but will not be able to generalize. This phenomenon is called overfitting.

Once the model is created, it goes through two different steps: the training and the inference. During the training the model learns the representations that connect the input data to the labels. Once the model is trained, it can be used to do inference and make predictions on unlabeled examples (the test set).

It is common to refer to ML models as Artificial Neural Networks, or, simply, Neural Networks (NNs). This is because these models are collections of units, or nodes, that are similar to the biological neural networks that constitute animal brains. For the same reason, the units of the model, that are the single elements able to perform operations on the inputs and are connected to each other, are also called neurons.

4.2 Model's architecture of Neural Networks

When creating a model, it is very important to choose properly its architecture. The best architecture depends on the purpose the NN is created for.

There is always an input layer, the one that just introduces the features to the model, and an output layer, the one that performs the last operations and gives the output values of the predictions. Every other layer that lies between the input and the output ones is called hidden layer.

In order to have an idea of how layers work, let's take as an example the simplest model possible: a model with no hidden layers. We also assume that our input is a list of examples and that every example is composed by four features, our example will therefore be $\vec{x} = (x_1, x_2, x_3, x_4)$. Figure 4.1 is a graphical representation of this model.

The single node of the output layer (the green dot) performs the following

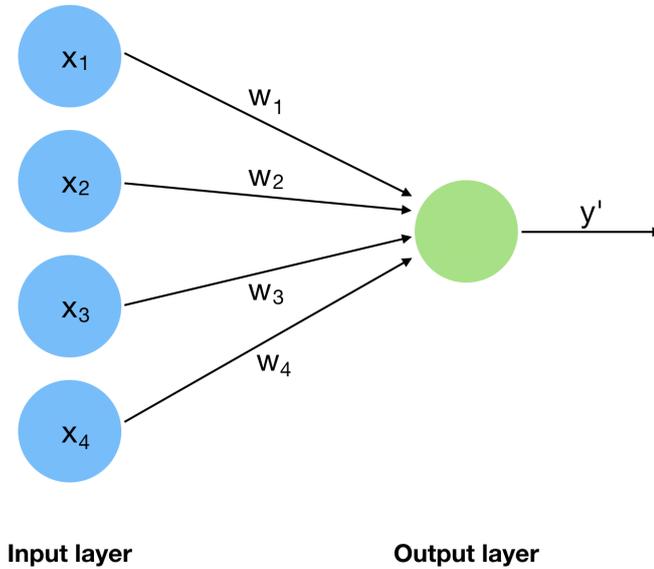


Figure 4.1: Representation of a model with no hidden layers and four features as input.

operation, to produce the prediction y' :

$$y' = a \left(\sum_i \vec{w}_i \cdot \vec{x}_i \right) , \quad (4.1)$$

where \vec{w}_i is a vector containing weights (also known as kernel matrix) and a is the activation function of the layer.

Both the weights and the activation function are crucial elements of the training process.

The weights are parameters of the model, the objects that are going to be modified at every step of the iterative process the training consists in, to produce the prediction y' , that has to be as close as possible to the label y of our example.

The weights are not the only parameters the model can have. Every layer can have an extra parameter: the bias, b . In this case Eq. (4.1) becomes

$$y' = a \left(\sum_i \vec{w}_i \cdot \vec{x}_i + b \right) . \quad (4.2)$$

4.2.1 Activation functions

From Equation (4.2) it is clear that if no activation function (a) is involved, then the output layer is only able to perform linear operations. Even if the model was

formed by several layers and every layers was composed by several nodes (or units) the result would always be a linear function of the input, because a combination of linear functions is still a linear function. Thus, the role of the activation function is almost always to introduce non-linearity in the model.

Some examples of activation functions are the following:

- Rectified Linear Unit (ReLU) (Glorot et al. 2011): this function returns 0 if the value of the input is negative, otherwise it returns the value of the input itself. This activation function is widely used in several applications and model architectures;
- Hyperbolic tangent: given an input p this function returns $\tanh(p)$;
- Sigmoid function: this function is defined as $S(p) = 1/(1 + e^{-p})$. This activation function is used in the output layer of Classification algorithms with two possible outcomes (also known as Binary Classifications). For example, this kind of activation function can be used in algorithms that classify galaxies as spherical or elliptical, or in algorithms trained to understand if an image of a galaxy is or is not deformed by Gravitational Lensing;
- Softmax function. This function takes as input a vector \vec{p} of n real numbers and gives as output a vector $\vec{\sigma}$ that is also composed by n real numbers. Softmax function is defined as follows:

$$\sigma(\vec{p})_i = \frac{e^{p_i}}{\sum_{j=1}^n e^{p_j}} \quad , \quad (4.3)$$

where p_i is the i – th element of \vec{p} .

From its definition it is clear that the output of the softmax function is a vector of real numbers, all belonging to the interval $(0, 1)$. Moreover, the sum of these elements is the unity.

These properties makes the softmax function the perfect candidate for the output layers of non-binary classification models. When we want to know to which of n classes an unlabeled example belongs to, the i – th element of the output of the softmax function can be considered as the probability for the example to be in the i – th class.

4.2.2 Categories of layers

In order to build the best model for our scientific purpose, it is important to choose a convenient number of layers, how many units these layers should have and what

type of layers they should be. There are several categories of layers. Here we present the most used ones:

- **Dense layer:** this layer is the most used one and it simply performs the operation described in Equation (4.2). It is a fully connected layer, which means that it is connected with all the units of the previous layer. Both the ML models we implemented use this type of layer. Figure 4.2 shows how a model built using only dense layers looks like.

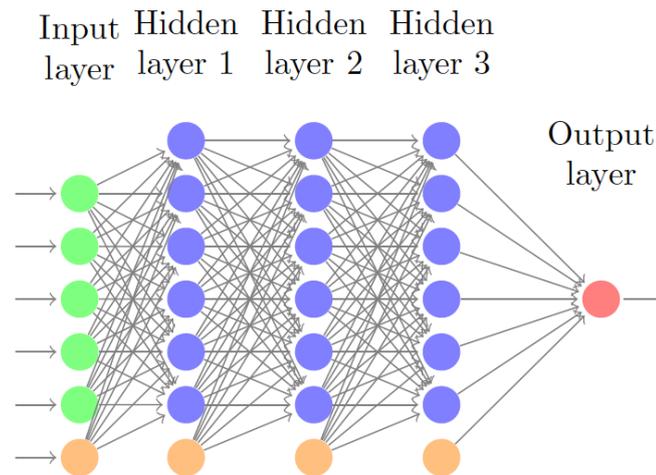


Figure 4.2: Model built using only dense layers. In this case there are three hidden layers, and each of them has 6 units (blue) and a bias (orange). Credits to Rezaie et al. (2019).

- **Dropout layer:** this layer randomly sets a specific fraction of input units to 0 at each iteration of the training. The value of this fraction is decided during the construction of the model. The following layer therefore will not have all the information that comes from the previous one. This is useful in deep and large models to prevent overfitting. We used this layer in the Classification model we implemented, that will be described in detail further in this work.
- **Distribution Layer:** this layer was presented, with the name of Distribution-Lambda layer, for the first time at the 2019 TensorFlow Dev Summit (6-7 March 2019) together with the library it belongs to: TensorFlow Probability. This layer takes as input n values and gives as an output a distribution instance parameterized by those n inputs. In our Regression model we used this layer to have as output a Gaussian distribution as prediction of the value of Ω_m for each example.

- Convolutional layers: these types of layers, as the Pooling layers and the Flattening layers, have not been used in the ML algorithms presented in this work, but they are described here to give a proper overview on the wide range of types of layers that can be used in ML. Convolutional layers apply filters to the input tensors. If, for example, the input of this layer is a 2D image, this image is seen as a matrix, and every element of this matrix is a number indicating the brightness of one pixel. Convoluting this image means replacing the value of every element of it with the result of the dot-product of two matrices. The elements of the first of these matrices are the values of the pixel we are replacing and of the pixels that are next to it. The second matrix is the filter (or kernel) we want to apply at the image. This process can, for example, clean the image from noise, if an averaging filter is applied, or it can ephasize features of the image, like contours and repeated patterns.

Convolutional layers can work on 1, 2 or 3D inputs. These kind of layers are of central importance in image classification algorithms.

- Pooling layers. Together with Convolutional Layers, these layers are the core of Convolutional Neural Networks (CNNs) that are the best models for image recognition and classification. One example of the structure of a CNN that performs handwritten digits recognition is shown in Figure 4.3.

Pooling layers take as input a matrix and perform down sampling, reducing the dimension of such a matrix, improving the speed of the model without any major loss of information. If our input data is a 2D image, a pooling layer subdivides this image in small blocks of pixels and replaces every block with a single real number, that can be for instance the average value of the pixels that are in that block (average pooling) or the maximum value of those pixels (max pooling);

- Flattening layers: these layers flatten the input, turning it into a 1D vector.

Once the structure of the model has been decided and built, it is time to feed it with the input features and to start the training (or learning) process, that will be described in detail in the following section.

4.3 Training process

Setting up the fittest architecture is not the only thing that has to be done in order to obtain good results from ML. The training process is characterized by some choices and hyperparameters that are going to be explained in the next section.

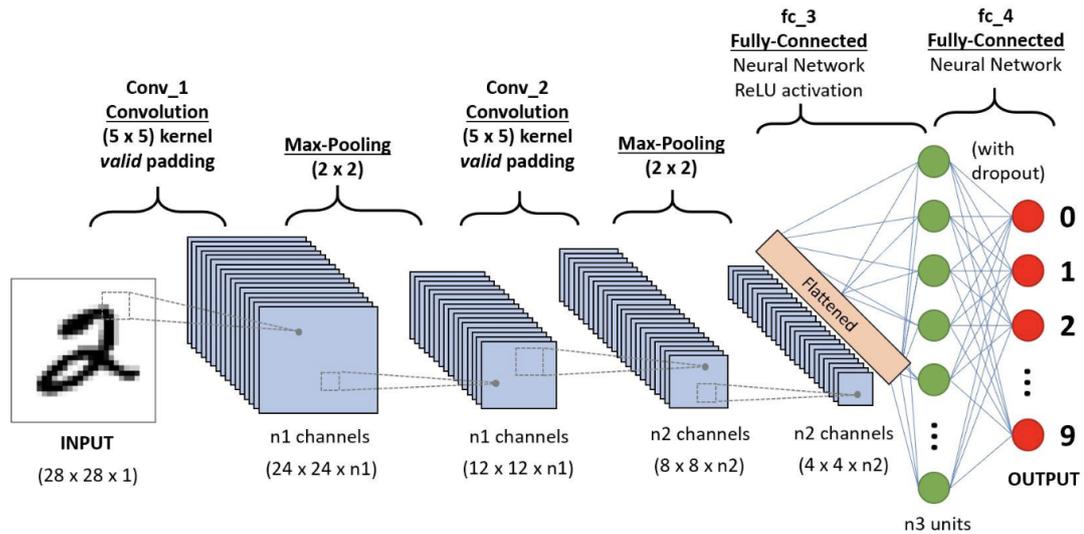


Figure 4.3: Structure of a CNN that performs a classification of handwritten digits. The input consists in a (28×28) matrix containing the values of brightness of every pixel. The blue squares are representations of the inputs that pass through two convolutional layers and two pooling layers. The output of the second pooling layers is then flattened and used as input for the final part of the model, that consists of two dense layers. The output of the model is a vector of ten elements. Every element is the probability that the input represents a specific digit. Credits to <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.

4.3.1 Loss

The goal of the training process is to find good values for all the weights and biases. The best values will be the ones that lead the predictions for the labeled examples to be as close as possible to the labels. The measure of how bad the predictions are is called loss, J . The smaller the loss, the better the model is to predict on the labeled examples.

Once the input features have gone through all the layers of the model and a prediction has been made for each example, we have to find a way to compute the total loss. There are many ways to do it, here we present the most used loss functions:

- Mean Squared Error. This loss function is widely used in linear regression algorithms, and the operation that performs is

$$J = \frac{1}{N} \sum_{i=1}^N (y(\vec{x}_i) - y'(\vec{x}_i))^2 \quad , \quad (4.4)$$

where N is the number of examples present in our dataset, $y(\vec{x}_i)$ is the label of the i -th example and $y'(\vec{x}_i)$ its prediction;

- Mean Absolute Error. This loss function, like the next two ones, is very similar to Mean Squared Error. The expression for Mean Absolute Error is

$$J = \frac{1}{N} \sum_{i=1}^N |y(\vec{x}_i) - y'(\vec{x}_i)| \quad ; \quad (4.5)$$

- Mean Absolute Percentage Error:

$$J = \frac{1}{N} \sum_{i=1}^N \frac{100 \cdot |y(\vec{x}_i) - y'(\vec{x}_i)|}{y(\vec{x}_i)} \quad ; \quad (4.6)$$

- Mean Squared Logarithmic Error:

$$J = \frac{1}{N} \sum_{i=1}^N [\log(y(\vec{x}_i)) - \log(y'(\vec{x}_i))]^2 \quad ; \quad (4.7)$$

- Binary Cross Entropy: in information theory entropy, $H(\vec{p})$, is defined as the average amount of information that comes from one sample drawn from a given distribution and it is defined as follows:

$$H(\vec{p}) = - \sum_i p_i \log_2(p_i) \quad , \quad (4.8)$$

where p_i is the probability associated to the i -th element of the sample. If two probability distributions, the real one, p , and the predicted one, q , are taken into account over the same set of events, the cross-entropy can be defined as a function of both these distributions:

$$H(\vec{p}, \vec{q}) = - \sum_i p_i \log_2(q_i) \quad , \quad (4.9)$$

where p_i and q_i are the probabilities of the i -th event, according to the real and the predicted probability, respectively. If the predicted probability is different from the true one, then the cross entropy will be higher than the entropy. The difference between entropy and cross entropy is called relative entropy, or, more commonly, Kullback-Leibler Divergence (Kullback and Leibler, 1951).

Binary Cross Entropy loss function is the most used in Binary Classification, where the label of each example can be either 0 or 1. In these models the output layer has a sigmoid activation function, that returns the probability the example has to have 1 as label.

The analytic expression of binary cross entropy is:

$$J = -\frac{1}{N} \sum_{i=1}^N y(\vec{x}_i) \cdot \log[p(y(\vec{x}_i))] + (1 - y(\vec{x}_i)) \cdot \log[1 - p(y(\vec{x}_i))] \quad , \quad (4.10)$$

where $p(y(\vec{x}_i))$ is the predicted probability for the $i - th$ example to have 1 as label;

- **Categorical Cross Entropy:** this loss function is very similar to the previous one, but it is used in non-binary classifications, where the model has to choose which of M classes best represents the example. If the labels are integer numbers (like almost always are), then it is common to use the Sparse Categorical Cross Entropy, defined by the following expression:

$$J = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_j(\vec{x}_i) \cdot \log[p(y_j(\vec{x}_i))] \quad , \quad (4.11)$$

where $y_j(\vec{x}_i)$ is a binary indicator: 1 if j is the correct label for \vec{x}_i , 0 if not. This loss is used in models that have as activation function of the output layer a softmax function, and $p(y_j(\vec{x}_i))$ is the predicted probability for \vec{x}_i to belong to class j .

If the labels are continuous, then the sum over j turns into an integral between the minimum label and the maximum one.

It is now time to understand how the model can improve during the training, and how it is supposed to reach the lowest loss possible. First of all, let us recall that the training is an iterative process. The same example passes through all the layers several times. Every time all the examples are processed is called “epoch”. Usually the training process consists in thousands of epochs, and at the end of every epoch the weights are updated in order to reach a lower loss during the following one.

4.3.2 Optimization

The updating of the weights is called optimization. There are many optimizers that can be used in ML. Here some examples are presented:

- Gradient Descent: this optimizer is very basic and can be used as a starting point to understand the functioning of all the other ones.

This optimizer computes the gradient of the loss, as a function of all the weights of the model. The gradient, by definition, points towards the direction of the steepest increase in the loss function. This optimizer updates the weights of the model in order to be a step away from the previous ones, in the direction of the negative gradient. This is repeated at the end of every epoch, until the model eventually converges to a minimum loss.

The length of the step is a fraction of the magnitude of the gradient. This fraction is called learning rate or, sometimes, step size. The choice of the value of the learning rate is not an easy task to do and often requires a little bit of trial and error. If the learning rate is too big, then the model goes rapidly towards the minimum of the loss function, but the risk is to never reach this minimum exactly. If the learning rate is too small, then the training process can require too many epochs to reach the minimum of the loss. Moreover, the risk of a low step size is to fall into a local minimum, without being able to leave it.

A correct learning rate is therefore the one that can lead the model to the lowest loss possible, without taking too many epochs.

If the training set is too big to compute the loss for every example, it is possible to calculate it only on one randomly sampled example for each epoch (Stochastic Gradient Descent) or on a subsample of the dataset (mini batch Stochastic Gradient Descent). Here we have introduced the term “batch” to indicate a subsample of the dataset that can contain any number of examples from one to N .

The mathematical expression for Gradient Descent optimization is

$$\vec{w}_{e+1} = \vec{w}_e - \eta \nabla J(\vec{w}_e) \quad , \quad (4.12)$$

where \vec{w}_e is a vector containing the values that weights have in the $e - th$ epoch, and η is the learning rate.

- Adaptive Gradient Algorithm (AdaGrad)(Duchi et al. 2011) is an adaptive learning rate method. This means that every weight is varied individually. The analytic expression does not differ much from Eq. (4.12):

$$w_{i,e+1} = w_{i,e} - \frac{\eta}{\sqrt{G_{i,e} + \epsilon}} \frac{\partial J}{\partial w_{i,e}} \quad , \quad (4.13)$$

where $w_{i,e}$ is the value of the $i - th$ weight at epoch e , ϵ is a parameter that has to be set during the construction of the model (just like the learning

rate) and $G_{i,e}$ is defined as follows:

$$G_{i,e} = \sum_{j=1}^e \left(\frac{\partial J}{\partial w_{i,j}} \right)^2 . \quad (4.14)$$

From Eq. (4.13) and Eq. (4.14) it can be seen that the effective learning rate for each weight depends on the training itself. The more one weight had been updated up to a certain epoch, the lower its effective learning rate will be in that specific epoch. The AdaGrad optimizer is therefore capable of taking into account that not every weight behaves in the same way during the training, and it is most likely that every weight gets to its best value on a different epoch.

- Adadelta (Zeiler 2012) is basically the same as AdaGrad, but $G_{i,e}$ is no more the sum of squared gradients over all epochs, but the sum is now restricted to a window of n iterations. This means that the denominator of the first fraction in Equation (4.13) can no more tend to infinite, and so the effective learning rate will never become negligible, even after many epochs.
- Root Mean Square propagation (RMSprop) (Tieleman and Hinton 2012): this optimizer is very similar to AdaGrad. Infact Eq. (4.13) is still valid, while the definition of $G_{i,e}$ is modified as follows:

$$G_{i,e} = \beta G_{i,e-1} + (1 - \beta) \left(\frac{\partial J}{\partial w_{i,j}} \right)^2 , \quad (4.15)$$

where the new hyperparameter β has to be set just like η and ϵ , and it is usually close to 0.9.

- Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014) is an optimizer that uses estimates of the first and second moments of the gradient of the loss function in order to update the effective learning rate. Adam uses exponentially moving averages to estimate the moments: m_e and v_e for the first and second moments, respectively, at epoch e . The moving averages are computed as follows:

$$m_e = \beta_1 m_{e-1} + (1 - \beta_1) \nabla J(\vec{w}_e) \quad (4.16)$$

$$v_e = \beta_2 v_{e-1} + (1 - \beta_2) \nabla [J(\vec{w}_e)]^2 , \quad (4.17)$$

where β_1 and β_2 are two hyperparameters set during the construction of the model. Usually those hyperparameters are set to be $\beta_1 = 0.9$ and $\beta_2 = 0.999$, as it was done in the original paper that presented this optimizer.

It can be demonstrated that m_e and v_e are biased estimates, but they can be corrected performing the following operations:

$$\hat{m}_e = \frac{m_e}{1 - \beta_1^e} \quad (4.18)$$

$$\hat{v}_e = \frac{v_e}{1 - \beta_2^e} \quad (4.19)$$

where \hat{m}_e and \hat{v}_e are the unbiased estimators.

Once the unbiased estimates have been defined, it is easy to understand the way the weights are updated:

$$\vec{w}_e = \vec{w}_{e-1} - \eta \frac{\hat{m}_e}{\sqrt{\hat{v}_e + \epsilon}} \quad , \quad (4.20)$$

where η and ϵ play the same role they played in the previously presented optimizers.

Now we know the layers that a model can be composed of, the loss functions that are able to measure how wrong our predictions are at every epoch, and how to optimize the model in order to reduce that error.

Using the fittest layers, with the correct loss function and optimization is mandatory in order to have a well performing model.

4.4 Validation

So far we have talked about the training set of labeled examples and the test set of unlabeled examples. Note that, as far as the test set is concerned, the labels are not given to the model, but might be known by the user.

As we have discussed in the previous sections, there are several choices that have to be done and hyperparameters to be set during the construction of the model. These choices, however, should not be done based on how the model performs on the training set, because, for instance, if we build a huge and deep model, it will most likely perform very well on the set it has been trained on, but will not be able to generalize to previously unseen examples.

This means that the number of layers, the value of the learning rate and so on, have to be chosen based on how well the model performs on the test set. But this again leads to some contradictions: modifying the model depending on the predictions on the test set sounds like indirectly feeding the model with the set, and this is not good if we want an absolute generalization.

Here is where validation comes up. The training set can be split into two subsamples: one of them keeps on being the training set the model learns the

weights on, and the other will be the validation set, that will not be used to update the weights, but to make choices about the architecture of the model and the optimization parameters.

If we want to summarize what have been said so far we can say that the dataset is divided into three different sets:

- The training set is used to train the model. This means that the optimization of the loss function and the updating of the weights is performed based on this set;
- The validation set is used to improve the structure of the model;
- The test set, composed by unlabeled examples, is used to make predictions on previously unseen examples and, if the labels of this set are known by the user, to check whether the training and validation processes worked properly or not.

In the next chapters we will present how the concepts of ML that have been introduced here have been applied in order to build models that can put constraints on the value of Ω_m , starting from the BOSS galaxy catalogue data.

Chapter 5

The analyzed dataset

In this work we exploited supervised ML methods providing as input training data the correlation functions of thousands of mock galaxy catalogues. These mock catalogues have been created following the characteristics of the BOSS galaxy catalogue, using a technique that is going to be presented in the next chapter.

This catalogue has been chosen because it is public and contains the coordinates of a large number of observed galaxies ($\approx 1.200.000$), distributed in a large sky volume. Moreover, BOSS has been the object of other previous projects (e.g. Sanchez et al. 2013, 2016a, 2016b, Li et al. 2016, Alam et al. 2017, Ntelis et al. 2017, Salazar-Albornoz et al. 2017, Satpathy et al. 2017, Slepian et al. 2017, Ivanov et al. 2019) that used standard analysis techniques on clustering statistics to put constraints on cosmological parameters. In particular, we compared our results with the results obtained by Alam et al. (2017), that are presented in detail in this chapter.

5.1 The Sloan Digital Sky Survey

BOSS is part of SDSS, which is an imaging and spectroscopic redshift survey that used a 2.5m modified Ritchey-Chrétien altitude-azimuth optical telescope located at the Apache Point Observatory in New Mexico, USA (Gunn et al. 2006).

The data we have worked on are from the Data Release 12 (DR12), that is the final data release of the third phase of the survey (SDSS-III), that lasted from 2008 to 2014 (Alam et al. 2015).

BOSS is not the only survey of SDSS-III. In fact there are three more, that are not considered in this work (Eisenstein et al. 2011):

- APO Galactic Evolution Experiment (APOGEE) used high resolution infrared spectroscopy to survey 100,000 red giant stars across the full range of

the Galactic bulge, bar, disk and halo;

- Multi-object APO Radial Velocity Exoplanet Large-area Survey (MARVELS) monitored the radial velocities of 11,000 bright stars in order to detect giant gaseous planets orbiting around them;
- SEGUE-2 is the second version of the original Sloan Extension for Galactic Understanding and Exploration (SEGUE-1) that took spectra of almost 230,000 stars, of all the spectral types, in order to investigate the structure of the Milky Way. SEGUE-2 added 119,000 more spectra to the survey.

5.2 Properties of BOSS

BOSS mapped the spatial distribution of luminous red galaxies and quasars to investigate the scales imprinted by the BAO.

The observations were performed from fall 2009 to spring 2014, with a 1000-fiber spectrograph with a resolution $R \approx 2000$. The wavelength range goes from 360 nm to 1000 nm, and the coverage of the survey is almost 10.000 square degrees. The spatial coverage of BOSS is shown in Figure 5.1.

The catalogue from BOSS DR12, used in this work, contains the positions in observed coordinates: RA, Dec and redshift, of 1.198.004 galaxies.

Figure 5.2 shows how these objects are distributed along the three coordinates.

For this thesis work it has been used also a random catalogue with the same geometrical characteristics and the same angular and redshift selection functions of the data catalogue, but with a number of objects that is ten times bigger than the observed galaxies. The choice of having a more populated random catalogue has been done to minimize the impact of Poisson errors in random pair counts.

Both the data and the random catalogue have been created considering the survey footprint, veto masks and systematics of the survey, for example fiber collisions and redshift failures (Reid et al. 2015).

5.3 The correlation function of BOSS galaxies

All the measurements of correlation functions we used to train, validate and test our NNs, that are going to be explicitly described in the next chapter, have been done using the Landy and Szalay estimator (see Eq. (3.8)) from mock catalogues that have the same geometric structure of BOSS. The errors on the measures are computed as the Poissonian uncertainties on the number of pairs, both for the data catalogue and the random catalogue. These measurements have been performed

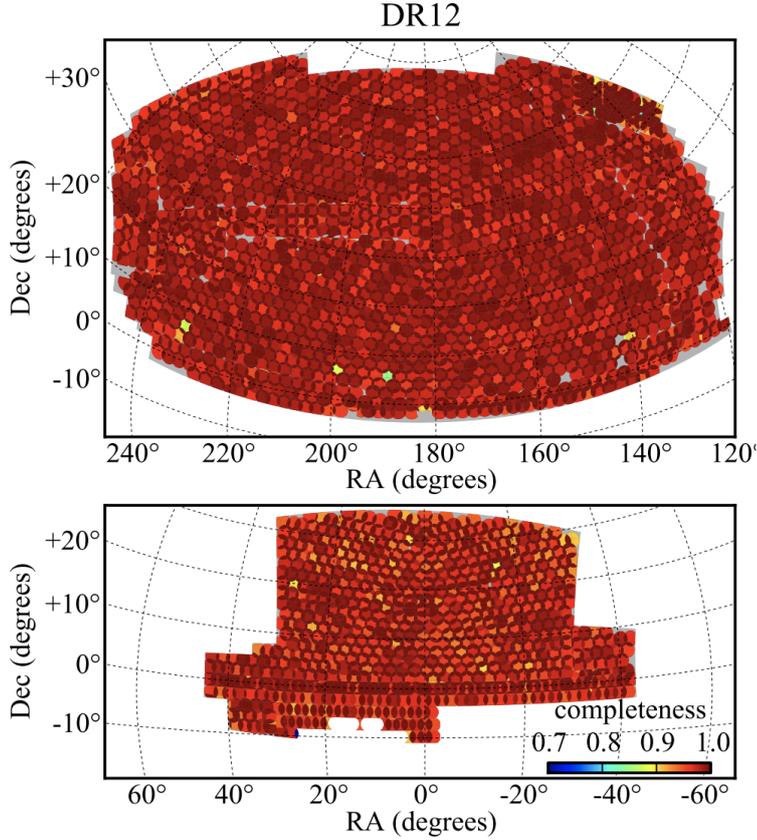


Figure 5.1: Footprints of the Northern and Southern galactic caps of BOSS DR12 sample. Credits to Alam et al. (2015).

using the CosmoBolognaLib, that are a set of Open Source C++ and Python numerical libraries for cosmological calculations (Marulli et al. 2016).

Figure 5.3 shows an example of such measures, obtained with a fitting range of s from $25 \text{ Mpc}/h$ to $145 \text{ Mpc}/h$, divided in 24 linear bins. A redshift selection has been applied on both the data and random catalogue, in the range between $z = 0.4$ and $z = 0.6$. The two-point correlation function is calculated assuming a fiducial cosmology whose parameters are the ones that result from the 2018 measurements of Planck satellite (Planck Collaboration 2018), but with $\Omega_m = 0.31$ instead of $\Omega_m = 0.3153$ (keeping $\Omega_{tot} = 1$). In the same figure, our measurements are compared with the ones of Satpathy et al. (2017), that have been chosen as comparison because are performed with the same catalogues, with the same estimator, in the same distance and redshift ranges and with the same bin size, and because they are publicly available. In particular, Satpathy et al. (2017) performed their measurements assuming a fiducial Λ CDM cosmology with $\Omega_m =$

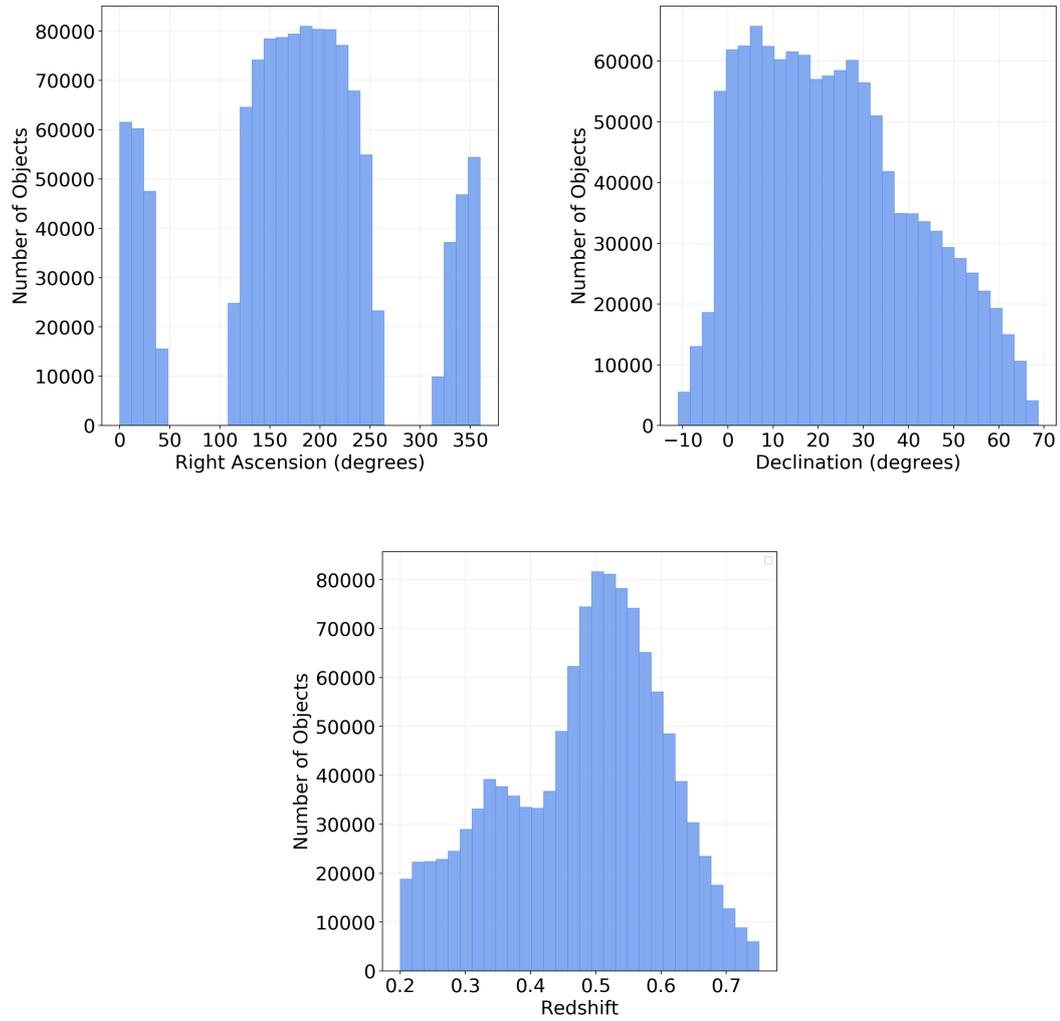


Figure 5.2: Histograms representing the distribution of objects in BOSS DR12 catalogue as a function of Right Ascension (top left panel), Declination (top right panel) and redshift (bottom panel).

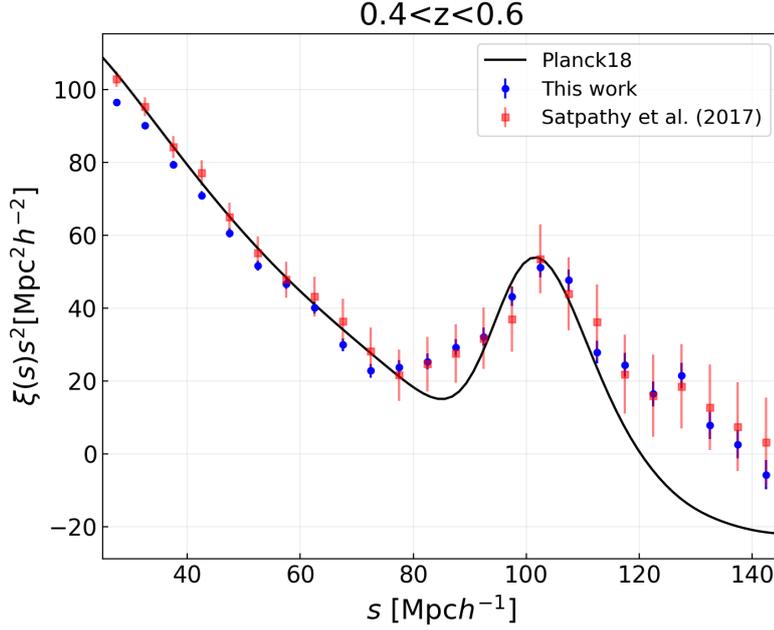


Figure 5.3: Different measures of the two-point correlation function obtained with the Landy and Szalay estimator, from BOSS DR12 catalogue, considering only objects with redshift between 0.4 and 0.6. The blue dots (with Poissonian error bars) have been measured assuming the fiducial cosmology of Planck’s 2018 results with $\Omega_m = 0.31$ instead of $\Omega_m = 0.3153$. The red squares are the measures obtained in Satpathy et al. (2017) assuming a fiducial Λ CDM cosmology with $\Omega_m = 0.31$, $H_0 = 0.676$, $\Omega_\Lambda = 0.69$, $\Omega_b h^2 = 0.022$ and $\sigma_8 = 0.8$. The red error bars have been obtained from the covariance matrix computed from mock catalogues. The black line represents a model for the galaxy correlation function obtained from a DM model multiplied by the Kaiser factor and the squared linear bias. The cosmology represented by this model is the one parameterized by Planck results of 2018. Both the measures have been done in 24 linear bins of s from 25 Mpc/h to 145 Mpc/h.

0.31, $H_0 = 0.676$, $\Omega_\Lambda = 0.69$, $\Omega_b h^2 = 0.022$ and $\sigma_8 = 0.8$. The error bars are obtained from the square root of the diagonal elements of the covariance matrices computed from the Multi-Dark Patchy mock catalogues (Kitaura et al. 2013). The black line in Figure 5.3 is the theoretical model of the galaxy correlation function in redshift space, computed as the correlation function of DM in real space multiplied by the Kaiser factor (see Equation (3.15)) to model RSD, and by the squared linear bias. Specifically, the model, parameterized by $b\sigma_8$ and $f\sigma_8$, has the following analytical expression:

$$\xi_{gal}(s) = \left[(b\sigma_8)^2 + \frac{2}{3}f\sigma_8 + \frac{1}{5}(f\sigma_8)^2 \right] \frac{1}{\sigma_8^2} \xi_{DM}(r) . \quad (5.1)$$

In our case $\xi_{DM}(r)$ is obtained by Fourier transforming the power spectrum modeled with the Code for Anisotropies in the Microwave Background (CAMB)(Lewis

and Challinor, 2011) and assuming Planck18 fiducial cosmology. CAMB is a Boltzmann solver to calculate the evolution of density perturbations in an assumed cosmological framework. The values of f and σ_8 are set by the cosmological assumption and $b\sigma_8$ is the result of a fit of BOSS correlation function through Monte Carlo Markov Chain, as explained in details in the next Section.

5.4 Standard analysis on clustering statistics

5.4.1 Bayesian statistical analysis

The NNs developed in this work have been used as a likelihood-free alternative to the standard Bayesian analysis, to put new constraints on the matter density contrast parameter, Ω_m , using the measured correlation function of the BOSS galaxy catalogue.

In the following, we provide a detailed description of the standard approach generally exploited by cosmological analyses, as a comparison to the new technique introduced in this work.

The first step of the standard approach is to choose a statistic to analyze. The two-point correlation function is often chosen because, under the assumption of a Gaussian distribution of the density field, it encloses all the spatial information, being the variance of a Gaussian field, that has null mean.

Once the statistic have been chosen, a likelihood has to be implemented. For a better understanding of what the likelihood is, and what is its role in this analysis, we introduce the Bayes' theorem:

$$P(\vec{\theta}|\vec{D}) = \frac{\mathcal{L}(\vec{D}|\vec{\theta})p(\vec{\theta})}{\sum_i \mathcal{L}(\vec{D}|\vec{\theta}_i)p(\vec{\theta}_i)} \quad , \quad (5.2)$$

where $\vec{\theta}$ is the vector that contains the values of the model parameters we are fitting our data, \vec{D} , with. $P(\vec{\theta}|\vec{D})$ is referred to as the posterior probability for the model given the data. It represents a probability distribution on the parameters, based on how the model fits the input data. The main goal of this kind of analyses is to fully characterize the posterior distribution. The prior probability, $p(\vec{\theta})$, is also a probability distribution of the parameters, but it is only based on the knowledge we had on them before the data, \vec{D} , were considered. The prior probability in an experiment can be, for example, the posterior probability of previous experiments. Finally, $\mathcal{L}(\vec{D}|\vec{\theta})$ is the likelihood, that represents the probability of getting the observed data, given the model parameterized by $\vec{\theta}$. The likelihood is the same

probability that is used in a frequentist approach. It includes the model itself, together with a description of the uncertainties.

As far as a measure of the two-point correlation function is concerned, the first thing the likelihood has to contain is the model of this statistic for DM in real-space. One example of this can be obtained by Fourier transforming the power spectrum modeled with CAMB. In the likelihood there must be also the bias of the galaxies the correlation function of which we have measured, the RSD, the geometric distortions and the covariance matrix of the measure. The implementation of the likelihood is of central importance in the standard analyses.

The standard Bayesian analysis consists in an iterative process that has the role of sampling the posterior probability distribution. This kind of process can be done through the Monte Carlo Markov Chain (MCMC) technique. A chain is a series of random numbers, where the conditional probability of an element given the others can be explicitly expressed. A chain itself can be seen as a single random object. A Markov chain is a chain in which the conditional probability of every element, x_{n+1} , depends only on the previous one, x_n . This probability, $p(x_{n+1}|x_n)$, is called the transition kernel. The elements of the Markov chain are points in the parameter space. At the n -th step of the fitting process, we are in the point characterized by the element x_n of the chain. The fit then moves to a new point, characterized by x_{n+1} , taking a step in a random direction of the parameter space. This step can be either accepted or rejected. To be accepted the new element x_{n+1} has to fulfill a specific requirement, for example, the likelihood in that point of parameter space must be higher than the likelihood of the previous element of the chain. During the whole process the chain cannot exit from the limits on the values of parameters that are set by the prior distributions.

As the fitting algorithm goes on, the chain will converge to a small region of the parameter space, where the posterior reaches the highest possible values. The values of the parameters that are elements of the Markov chain consist of the sample of the posterior probability. The mean, or the median, of the posterior probability of each parameter can then be considered as the best-fit value of that specific parameter.

To initialize the chain it is necessary to make an initial guess of the values of the parameters, and this guess might often be away from the region the chain will converge to. Therefore, it is common practice to start considering the sampling of the posterior after a certain number of steps, called the burn-in period.

5.4.2 Cosmological constraints from BOSS clustering

Alam et al. (2017) combined individual measurements of the two-point correlation functions and power spectra of the BOSS DR12 galaxy catalogue presented in

seven companion papers (some of which were by that time still unpublished) to put constraints on cosmological parameters. In particular, the BAO scale has been measured in the anisotropic two-point correlation function in Ross et al. (2016) and in Vargas-Magaña et al. (2018) and using the anisotropic power spectrum in Beutler et al. (2016b). The full shape of the two-point correlation function has been analysed in Satpathy et al. (2017) and in Sanchez et al. (2016a). The equivalent analysis in the Fourier space has been done using the power spectrum in Beutler et al. (2016a) and in Grieb et al. (2017).

To estimate the covariance matrix of their clustering measurements, two different types of mock catalogues have been used, created with two distinct methods: Multi-Dark Patchy (Kitaura et al. 2013) and Quick Particle Mesh (White et al. 2013). The former method simulates the growth of density perturbations with a combination of second-order Lagrange perturbation theory and a stochastic halo biasing scheme based on high-resolution N-body simulations. The Quick Particle Mesh method uses instead low-resolution particle-mesh simulations to evolve the density field, then selects particles from this field to match the statistics of the DM haloes. Both these methods then use halo occupation methods to construct galaxy distributions that match the observed redshift-space clustering and the angular selection function of BOSS catalogue. A total of 2000 catalogues were used in the analysis, 1000 for each method.

As an example, we illustrate here the likelihood used in Satpathy et al. (2017). In this paper the correlation function has been used as the analyzed statistic. The correlation function multipoles have been considered to be Gaussian distributed, and the parameter dependence of the covariance matrix has been ignored.

The Gaussian likelihood, \mathcal{L} , as a function of parameters \vec{p} , can be written as follows:

$$\mathcal{L}(\vec{p}) \propto \exp \left[\frac{-\chi^2(\vec{p})}{2} \right] , \quad (5.3)$$

where

$$\chi^2 = (\xi_{data} - \xi_{model}) \Sigma^{-1} (\xi_{data} - \xi_{model})^T . \quad (5.4)$$

In Eq. (5.4) Σ^{-1} is the inverse of the covariance matrix, that in Satpathy et al. (2017) was obtained from the Multi-Dark Patchy mocks, and ξ_{model} is the theoretical value of the correlation function the data have been compared with. The analytical expression that has been used to evaluate the value of ξ_{model} is the following:

$$1 + \xi_{model}(s_{\parallel}, s_{\perp}) = \int \exp \left\{ \frac{-[s_{\parallel} - r_{\parallel} - \mu v_{12}(r)]^2}{2[\sigma_{12}^2(r, \mu) + \sigma_{FOG}^2]} \right\} \cdot \frac{1 + \xi(r)}{\sqrt{2\pi\sigma_{12}^2(r, \mu)}} dy . \quad (5.5)$$

The elements present in Eq. (5.1) are the following:

- s_{\parallel} and r_{\parallel} are the line of sight separations in redshift space and real space, respectively;
- $r = \sqrt{r_{\parallel}^2 + s_{\parallel}^2}$ is the the pair separation in real space,
- $\mu = r_{\parallel}/r$ is the cosine of the angle between the line-of-sight separation in real space and the pair separation vector \vec{r} ;
- the parameter σ_{FOG} (where *FOG* stands for Fingers of God) serves to describe an isotropic dispersion that modifies the scale-dependence of the moments on small scales (Reid et al. 2012);
- the parameters $v_{12}(r)$ and $\sigma_{12}(r)$ represent the mean infall between pairs of matter tracers and the velocity dispersion along the line of sight, respectively.

In Satpathy et al. (2017) the temperature anisotropy data from Planck 2015 release has been used to compute the Planck likelihood, \mathcal{L}_{Planck} . The two likelihoods have been multiplied to get a joint constraint on the parameters:

$$\mathcal{L}_{Total}(p) = \mathcal{L}(p) \times \mathcal{L}_{Planck}(p) . \quad (5.6)$$

The combination of the seven different analyses in Alam et al. (2017) has a higher constraining power than each individual measure, because they have been done on the same catalogues (data and random), but independently and using different clustering statistics and modelling assumptions. The measurements, the likelihoods and the posterior distributions obtained by BAO and full-shape measurements have been combined into sets of consensus constraints using the method presented in Sanchez et al. (2016b).

The posterior has been sampled with a MCMC using a version of the workhorse CosmoMC code (Lewis and Bridle, 2002). The model spaces include variations in the matter density $\Omega_m h^2$, the baryon density $\Omega_b h^2$, the amplitude and the spectral index of the primordial power spectrum and the optical depth at recombination. Note that the choice of the analytical expression of the likelihood, the sampling of the posterior and all the uncertainties on the results that come from them are avoided in the likelihood-free ML approach presented in this work.

In Alam et al. (2017) the BAO and full-shape measurements of the BOSS catalogue have been combined with the CMB anisotropy data from Planck satellite's 2015 measurements (Planck Collaboration, 2016)¹. Assuming a standard Λ CDM model, the combination of BOSS and Planck data leads to the following constraint on the matter density:

$$\Omega_m = 0.311 \pm 0.006 , \quad (5.7)$$

¹Power spectra for both temperature and polarization have been used, using the likelihoods `plik_dx11dr2_HM_v18_TTTEEE` and `lowTEB` for the high and low multipoles, respectively.

this is the value that we used as comparison for the constraints obtained by the NNs presented in this thesis work.

Chapter 6

The training, validation and test datasets

The reliability of the results of a NN relies on the input features and examples the model was trained with. In this chapter we are going to explain how the mock catalogues of the training, validation and test set have been created.

6.1 Construction of log-normal mock catalogues

Let us define a random field in a given volume as a field whose value at the position \vec{r} is a random variable (Peebles 1993, Xavier et al. 2016). One simple example is the Gaussian random field, $N(\vec{r})$. In this case the one-point Probability Density Function (PDF) is a Gaussian (or Normal) distribution, fully characterized by the mean μ and the variance σ^2 :

$$f_1(N) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(N - \mu)^2}{2\sigma^2} \right] . \quad (6.1)$$

If n positions are considered, instead of just one, Eq. (6.1) becomes

$$f_n(\vec{N}) = \frac{1}{(2\pi)^{n/2} |\mathbf{M}|^{1/2}} \exp \left[-\frac{1}{2} \sum_{i,j} \mathbf{M}_{ij}^{-1} N_i N_j \right] , \quad (6.2)$$

where $N_i = N(\vec{r}_i)$, and \mathbf{M} is the covariance matrix, the elements of which are defined as follows:

$$\mathbf{M}_{ij} = \langle (N_i - \mu)(N_j - \mu) \rangle . \quad (6.3)$$

Equation (6.2) is the definition of the Multivariate Gaussian distribution.

As far as the primordial matter density contrasts are concerned, it is correct to approximate the PDF for $\delta(\vec{x})$ with a Gaussian distribution, with null mean and

the correlation function as variance. The same definitions can be used in Fourier space as well, with δ_k that will thus have null mean and $P(k)$ as variance.

In more general cases, the Gaussian random field is only an approximation of the real random field that may present features, such as significant skewness and heavy tails (Xavier et al. 2016).

Coles and Barrow (1987) showed how to construct non-Gaussian fields through non-linear transformations of a Gaussian field. One example of such fields is the log-normal random field (Coles and Jones 1991), obtained through the following transformation:

$$L(\vec{r}) = \exp[N(\vec{r})] \quad . \quad (6.4)$$

The log-normal transformation results in the following one-point PDF:

$$f_1(L) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\log(L) - \mu)^2}{2\sigma^2}\right] \frac{dL}{L} \quad , \quad (6.5)$$

where μ and σ^2 are the mean and the variance of the underlying Gaussian field N , respectively.

Similarly to what has been done for the Gaussian distribution, we can also determine a Multivariate version for the log-normal random field, that is defined as follows:

$$f_n(\vec{L}) = \frac{1}{(2\pi)^{n/2} |\mathbf{M}|^{1/2}} \exp\left[-\frac{1}{2} \sum_{i,j} \mathbf{M}_{ij}^{-1} \log(L_i) \log(L_j)\right] \prod_{i=1}^n \frac{1}{L_i} \quad , \quad (6.6)$$

where \mathbf{M} is the covariance matrix of the N-values.

We are now going to present the algorithm that has been used to generate the mock catalogues needed for the ML training process. As for the measurements of the two-point correlation functions, also for the construction of these mock catalogues the CosmoBolognaLib have been used (Marulli et al. 2016).

The algorithm takes as input:

- one data catalogue;
- one random catalogue;
- the bias of the data catalogue;
- the cosmological model we want the output catalogues to follow.

The first two inputs are used to define a grid with the same geometric structure of the data catalogue and a visibility function that comes from the pixel density of the random catalogue. The assumed cosmology is used to compute the distances

starting from the observed coordinates and to model, together with the bias of the data catalogue and its average redshift, the power spectrum that will serve as the logarithm of the variance for the log-normal random field generation. The density field is sampled from this random field.

Once the algorithm has associated to all the grid cells their density value, we can extract from each of them a certain number of points that depends on the density of the catalogue and on the visibility function. These points compose the output mock catalogue.

The main application for log-normal mock catalogues is the fast generation of mocks for covariance matrix estimate, in particular in anisotropic clustering analyses (e.g. Veropalumbo et al. 2016, Blot et al. 2019).

The log-normal technique allows to generate density fields (and then, catalogues) with the required characteristics very rapidly, specially if compared to N-body simulations, which, nevertheless, would have been more accurate and with more trustworthy higher-order statistics.

The cosmological parameter this work focuses on is Ω_m . Therefore, this was the only parameter that has been modified during the construction of all the different categories of mock catalogues. For all the other parameters, it has been chosen a fiducial cosmology from Planck satellite's 2018 measurements, as it has already been explained in Section 5.3. Every time the value of Ω_m was modified, Ω_Λ was also changed, in order to keep $\Omega_{tot} = 1$.

When a new category of mock catalogues, characterized by $\Omega_m = \Omega_{m,i}$, was constructed, a new linear galaxy bias had to be used. The galaxy bias was obtained by fitting the correlation function of BOSS, measured in the scale range $8 < s[\text{Mpc}/h] < 50$, that is the same range that is going to be used to train and test the NNs. The fit was performed on the scale range $30 < s[\text{Mpc}/h] < 50$. We used a Gaussian likelihood (see Eq. (5.3) and Eq. (5.4)), considering as theoretical model for the two-point correlation function the one expressed in Eq. (5.1), and a Poissonian covariance matrix. The model is parameterized by $f\sigma_8(z)$ and $b\sigma_8(z)$. From now on the dependence on the redshift will no more be made explicit. The value of every redshift-dependent parameter has been calculated using the mean redshift of the data catalogue, $\bar{z} = 0.481$. The value of $f\sigma_8$ was set by the cosmological model assumed during the measure of the two-point correlation function and the construction of the theoretical model. The other parameter of the model, $b\sigma_8$, is thus the only free parameter. We used a uniform prior between 0 and 2.7 for it. The posterior of $b\sigma_8$ has been sampled using a MCMC with 10.000 steps and a burn-in period of 100 steps. The mean value of the posterior is then taken and multiplied by σ_8 , which is again set by the cosmological model that has been assumed, to obtain the value of b that has been used during the construction of log-normal catalogues.

As an illustrative example, Figure 6.1 shows different measures of correlation functions, obtained on the scale range $8 < r[\text{Mpc}/h] < 50$, in 30 logarithmic bins of s . The blue dots and the green ones are the measures performed using two different classes of log-normal mocks. For each class, the measures of 50 mock catalogues are represented. These two classes are the ones characterized by the lowest and the highest values of Ω_m that have been considered in this work, that are 0.24 (blue dots) and 0.38 (green dots). The solid black line is the theoretical model for the correlation function, calculated with the same cosmological assumptions the class of log-normal mocks characterized by $\Omega_m = 0.24$ were created with. The dotted line is the analogous of the solid one, but calculated assuming $\Omega_m = 0.38$. The average values of the correlation functions of each class of log-normal catalogues stay close to their correspondent theoretical model. This is because these mock catalogues represent galaxy catalogues that actually follow the cosmology that was assumed during their construction. According to the ML models we are about to describe, a mock catalogue characterized by $\Omega_m = \Omega_{m,i}$ is an example of how galaxy catalogues would be if $\Omega_{m,i}$ was the true value. The red and the yellow crosses are two measures of the correlation function of the BOSS catalogue obtained assuming $\Omega_m = 0.24$ and $\Omega_m = 0.38$, respectively. Neither of the two measurements of BOSS correlation function follows the model that is characterized by the same cosmological assumptions. This is because both 0.24 and 0.38 are bad guesses for the value of Ω_m . Moreover, the two measures of BOSS correlation function have the same trend, and are only slightly different from each other. These differences are caused by the geometric distortions introduced by the two assumptions on Ω_m . As it will be shown in the next Chapter, the NNs presented in this work are able to predict the same value of Ω_m when applied to the correlation function measured from BOSS, thus not being significantly affected by the geometric distortions.

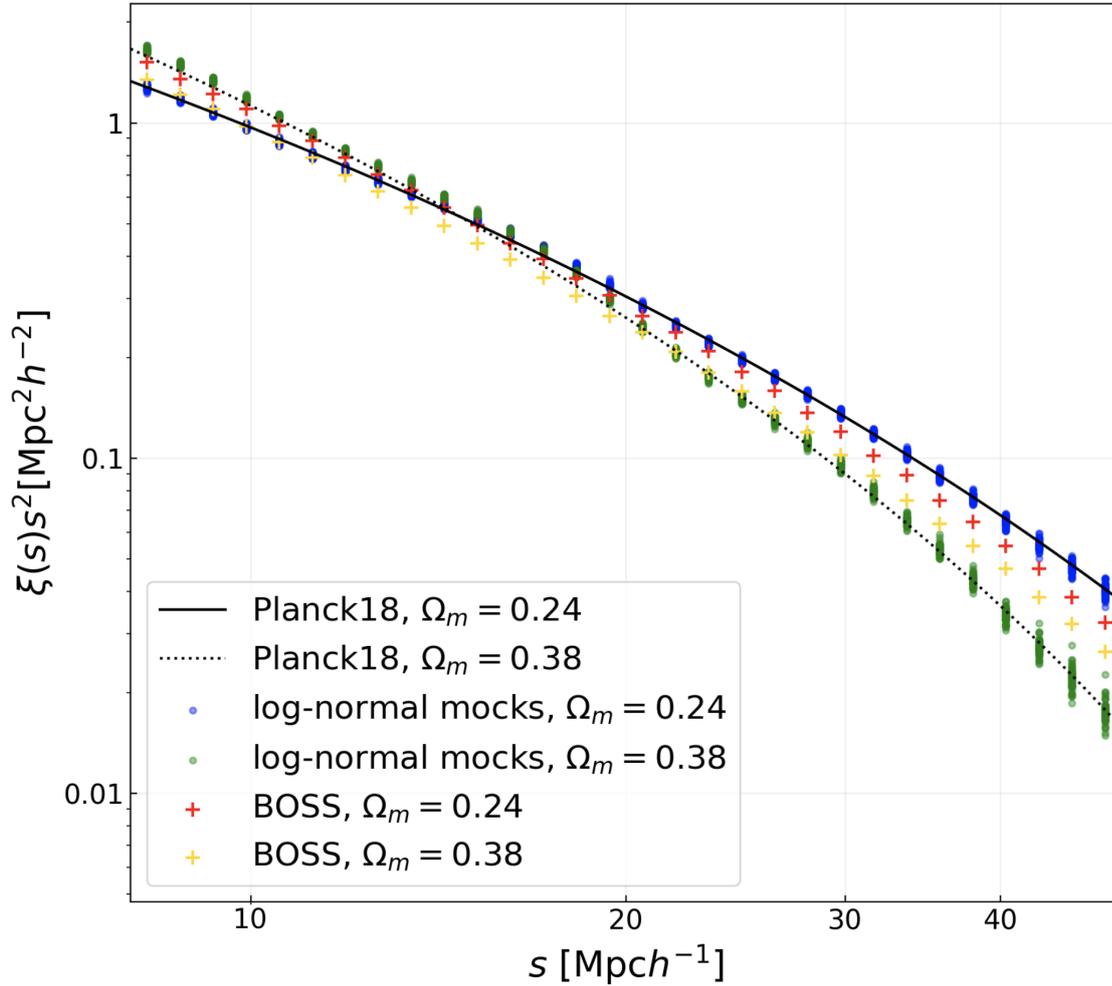


Figure 6.1: Different measures of the correlation function evaluated from 8 Mpc/h to 50 Mpc/h in 30 logarithmic bins of s . The dots represent the measures obtained with 50 log-normal mock catalogues constructed with $\Omega_m = 0.24$ (blue dots) and 50 constructed with $\Omega_m = 0.38$ (green dots). The black lines represent the theoretical models calculated assuming the same cosmological parameters as the measures. The solid line is characterized by $\Omega_m = 0.24$, while the dotted one by $\Omega_m = 0.38$. The red and yellow crosses show the BOSS correlation function obtained assuming $\Omega_m = 0.24$ and $\Omega_m = 0.38$, respectively.

Chapter 7

Exploitation of the Neural Networks

Two different ML models have been exploited in order to put constraints on Ω_m using the correlation function measured from BOSS. The first model performs a Multi-Class Classification, while the second one performs a Regression. These models are implemented in Keras (Chollet 2015) with a Tensorflow backend (Abadi et al. 2016).

In this chapter we are going to explain the architectures of the two models, the details of the training processes and the results that come out from them.

The architectures, the numbers of epochs and the values of the hyperparameters that are going to be presented are the result of a trial-and-error process. A wide variety of configurations has been tested. We kept the models' structures that had the best performances in avoiding overfitting and in providing reliable predictions on the test sets.

7.1 The Classification model

The Classification ML model we implemented for this thesis work takes as input the values of the measured correlation function, calculated in 30 logarithmic bins of s , from 8 Mpc/ h to 50 Mpc/ h . The model then generates as output, for each example, a vector containing the confidences of how likely the example is represented by each of the different classes the model is trained for.

In our case there are 11 classes, each of them characterized by a different value of Ω_m , from 0.26 to 0.36, separated by $\Delta\Omega_m = 0.01$.

The architecture of the implemented NN is the following:

- Input layer that feeds the 30 values of the measured correlation function to the model;
- Dense Layer with 8 units and ReLU activation;

- Dropout Layer with 0.125 as dropout rate;
- Dense Layer with 4 units and ReLU activation;
- Dense Layer with 11 units and softmax activation.

Figure 7.1 is a schematic representation of the architecture of this model. This

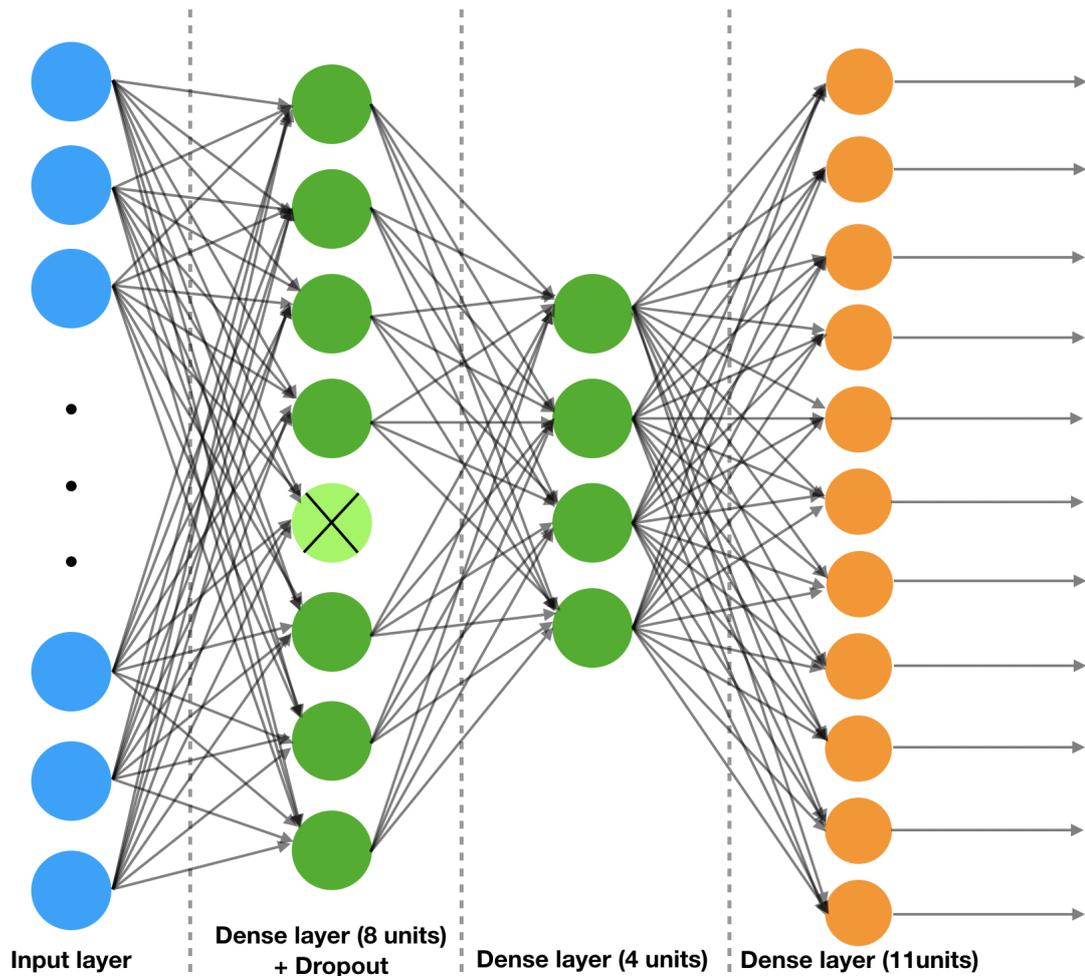


Figure 7.1: Schematic representation of the Classification model. The input layer is represented by blue dots, the hidden layers by green dots and the output layer by orange dots. The Dropout is represented by the cross on the fifth unit of the first hidden layer. In fact, at each epoch a random unit of that layers is chosen and it is not used to produce the predictions for that epoch. This helps to avoid overfitting.

architecture has been chosen because it is the simplest one we tested which is able to provide accurate cosmological constraints. Deeper models have been tried, but

no differences were spotted in the output predictions, and they often incurred in overfitting. This ML architecture is an example of Deep Learning NN, because multiple layers have been used, and each of them learned a different level of representation of the input features.

The training set considered in this work consists of 88.000 measures of correlation functions, that is 8.000 for each class, of which 2.000 measures have been used for the validation set. Classifications ML methods need big datasets to train, because every aspect of each class must be learned from the data. As far as the computation time is concerned, it was not convenient to perform the measures on catalogues with too many objects. Therefore the mock catalogues used to train and test this classification algorithm were obtained by slicing BOSS catalogues (data and random) and selecting only the following intervals of Right Ascension and Declination:

$$\begin{aligned} 3 < \text{RA} < 3.5 \\ 0 < \text{Dec} < 0.1 \end{aligned} .$$

All the resulting mock catalogues have approximately 40.000 objects each.

The learning process was performed using the Sparse Categorical Cross Entropy loss function, with Adam optimization (see Chapter 4). The optimization parameters that have been used are $\eta = 0.001$, $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We tried also with different values for the parameters of the optimization, but these changes produced no significant effects on the outputs. Besides the loss function, also the accuracy of the model has been monitored through epochs, as an indicator of the goodness of the training. The accuracy here is defined as the ratio between the number of times that the maximum value of the output vector of confidences corresponds to the right class (i.e. to the label of the example) and the total number of labeled examples the model has been fed with.

The model has been trained for 40 epochs, in batches of 50 examples each, and has then been used to make predictions on two different sets of unlabeled examples. The first one, that from now on will be referred to as the test set, is composed of 550 examples of mock catalogues, 50 for each class. The labels of this set were known, even if not by model, so these measures have been used to check whether the training process was successful or not.

The trained model has then been applied to a set consisting of 11 measures of BOSS correlation functions, calculated assuming the same values of Ω_m that were used to generate the train set. Measuring the correlation function of the same catalogue of observed objects assuming several cosmological models is important to quantify the impact of geometric distortions on the NN cosmological constraints.

7.2 Results of the Classification model

The predictions of our Classification model reached an accuracy of 28% on the training, the validation and the test sets.

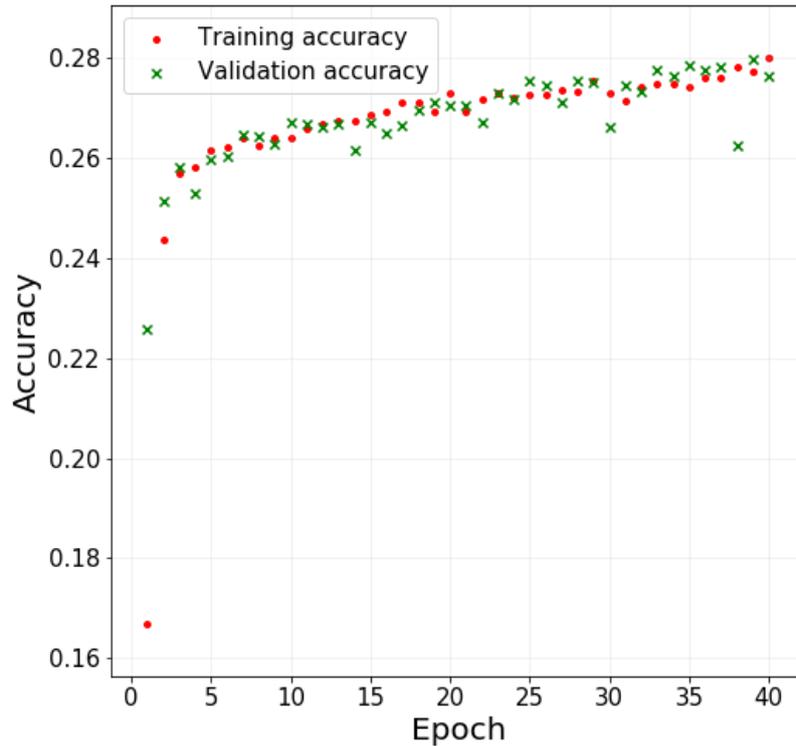


Figure 7.2: Training accuracy (red dots) and validation accuracy (green crosses) as functions of the epochs.

Figure 7.2 shows the trend of the training accuracy (red dots) and of the validation accuracy (green crosses) as a function of the epochs. The fact that the training accuracy is not systematically higher than the validation one means that the learning process worked properly, without overfitting the training set. To prevent overfitting, the learning process has been stopped after the first 40 epochs, because after them there was no further improvement in the accuracy, and the model has already converged.

The performance on the test set is summarized in Figure 7.3, that shows for each class, characterized by $\Omega_m = \Omega_{m,true}$, how the predictions on the 50 examples

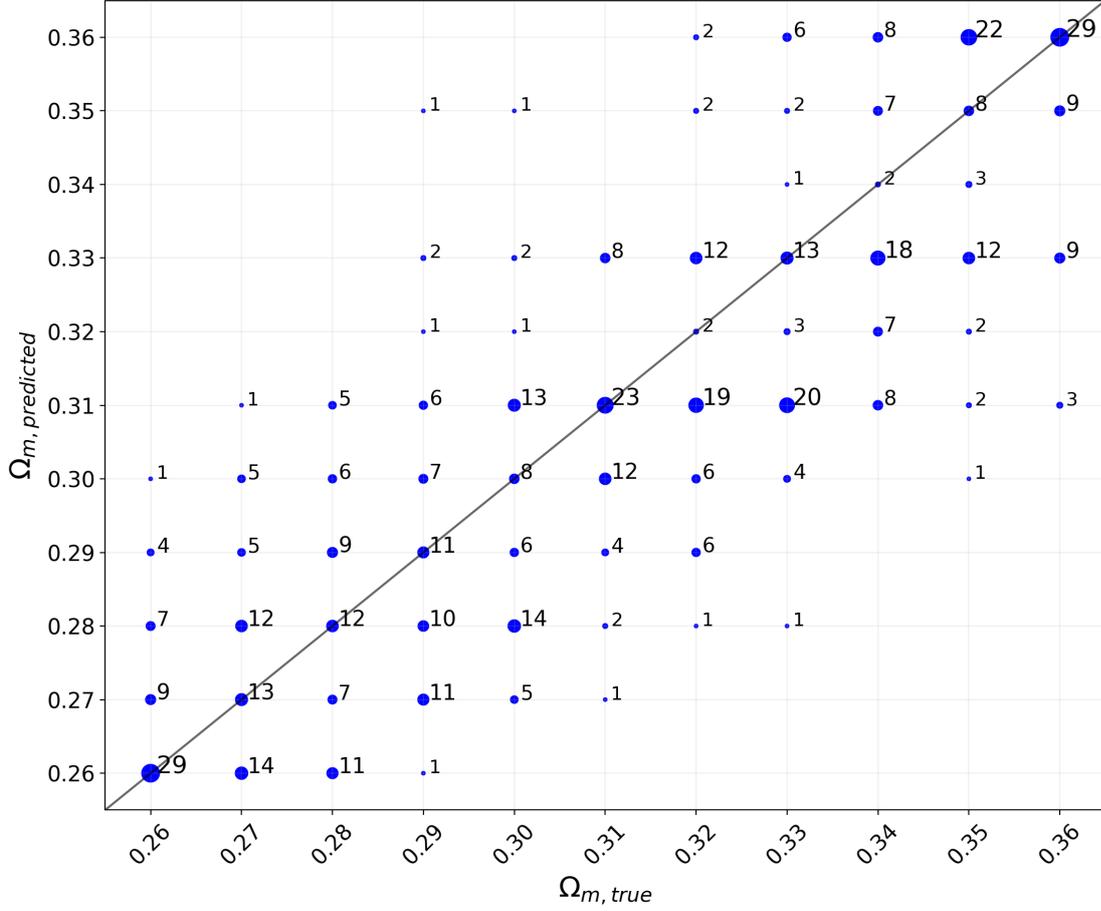


Figure 7.3: Values of Ω_m predicted by the Classification ML model as functions of the true value of the labels of the examples in the test set. The black line represents the bisector of the first quadrant.

of the test set are distributed. For example, for 23 out of 50 examples characterized by $\Omega_{m,true} = 0.31$, the model predicted the right value, for 12 examples of the same class the model predicted $\Omega_{m,predicted} = 0.30$.

The main reason we developed this NN for is to apply it to a set of measured correlation functions that come from observed data, the BOSS catalogue. For each of the 11 measures, corresponding to the 11 different assumptions on Ω_m , the 11 values of confidence the prediction is composed of have been fitted with a Gaussian function. Figure 7.4 is an example of one of these fits, corresponding to the prediction on the correlation function measured assuming $\Omega_m = 0.27$. The red line represents the best-fit model for the prediction, that is represented by the blue dots. The best-fit values are $\mu = 0.30$ and $\sigma = 0.03$. We can therefore

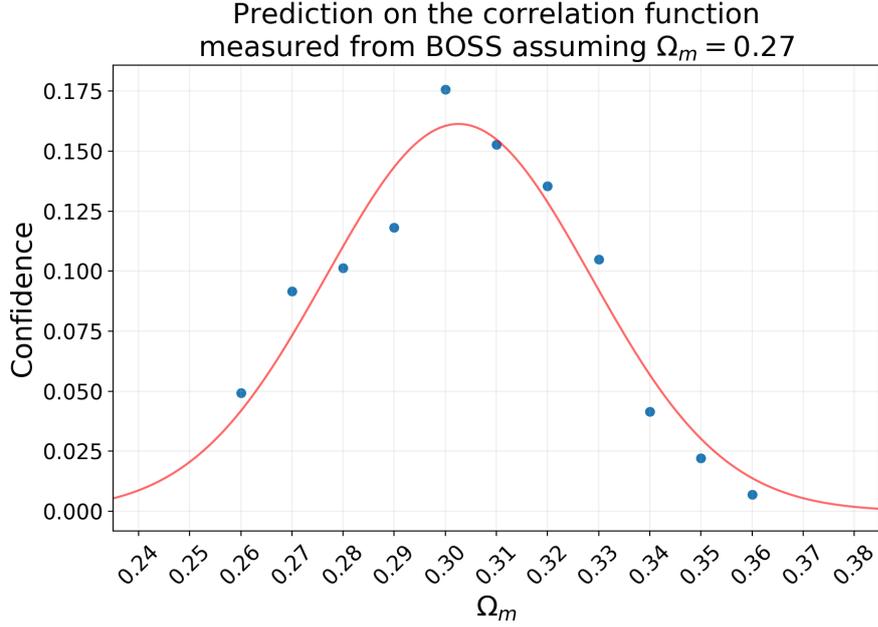


Figure 7.4: Prediction on Ω_m from the correlation function measured from BOSS, assuming $\Omega_m = 0.27$. The blue dots are the output of the model, while the red line shows Gaussian fit. The best-fit values for this fit are $\mu = 0.30$ and $\sigma = 0.03$, where μ is the mean of the Gaussian, and σ its standard deviation.

assert that the value of Ω_m that the model associates to this specific example is $\Omega_m = 0.30 \pm 0.03$.

An analogous procedure has been done also for all the ten others measures of BOSS correlation function. Each result is compatible with all the others, within the uncertainties. Figure 7.5 shows the predictions on Ω_m from the eleven correlation functions of BOSS as a function of the value of Ω_m assumed for the measurements.

The constraint we are able to give using this Classification model, applied on the correlation functions measured from BOSS is thus:

$$\Omega_m = 0.30 \pm 0.03 \quad , \quad (7.1)$$

where the expectation value and the uncertainty have been calculated averaging all the means and the standard deviations of the best fits.

Even if this result is compatible with the one obtained by Alam et al. (2017) (see Eq. (5.7)), the uncertainty we have is much larger. For this reason, we investigated also a different ML method based on the so-called Regression approach that, as we will see, can be used to put much stronger constraints on Ω_m .

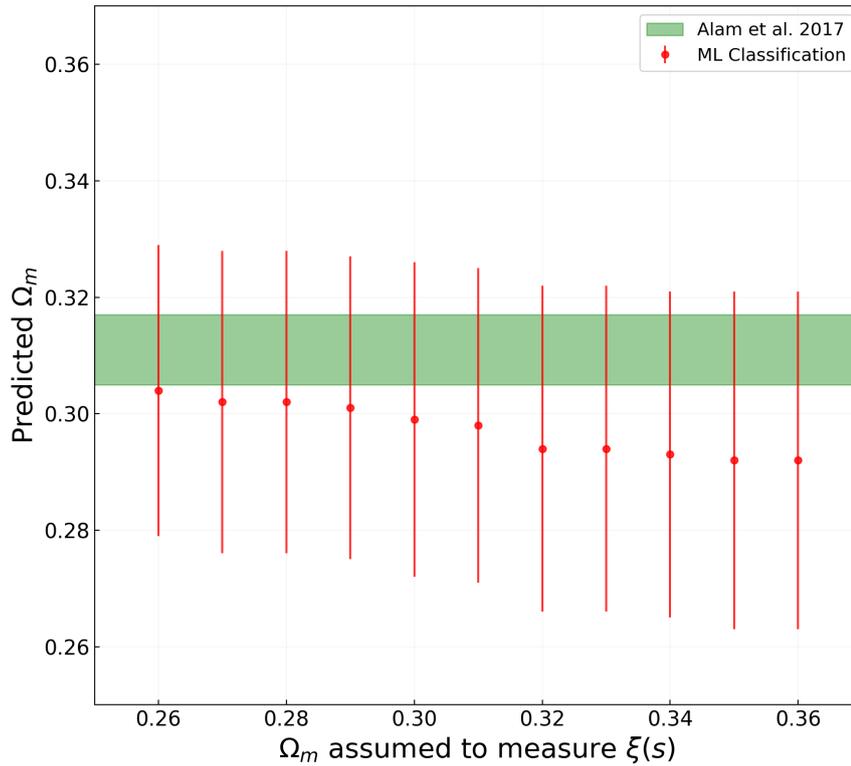


Figure 7.5: ML Classification’s predictions (red dots and bars) on Ω_m from the measured correlation function of BOSS, as a function of the value of Ω_m assumed during the measurements. There is a slightly decreasing trend of the predictions, though not statistically significant. The results are compared to the Ω_m constraint provided by Alam et al. (2017) (green stripe).

7.3 The Regression model

The main difference between a Classification model and a Regression model is that, while the former produces as output a vector of values, that are the confidences of the associations between the example and every class, the latter produces as output one single distribution function, that, in our case, is the predicted probability distribution of Ω_m .

The Regression ML model we are about to introduce has been trained and tested with measures of correlation functions in the range of s that goes from 8 Mpc/ h to 50 Mpc/ h . In this range the measures have been performed in 30 logarithmic bins of s .

The architecture of the implemented Regression NN is the following:

- Input layer that feeds the 30 values of the measured correlation function to

the model;

- Dense layer with 2 units;
- Distribution layer that uses the two outputs of the Dense layer as mean and standard deviation to parameterize a Gaussian distribution, which is given as output.

Figure 7.6 is a schematic representation of the architecture of this Regression model. This architecture was chosen with the same criteria adopted to choose the

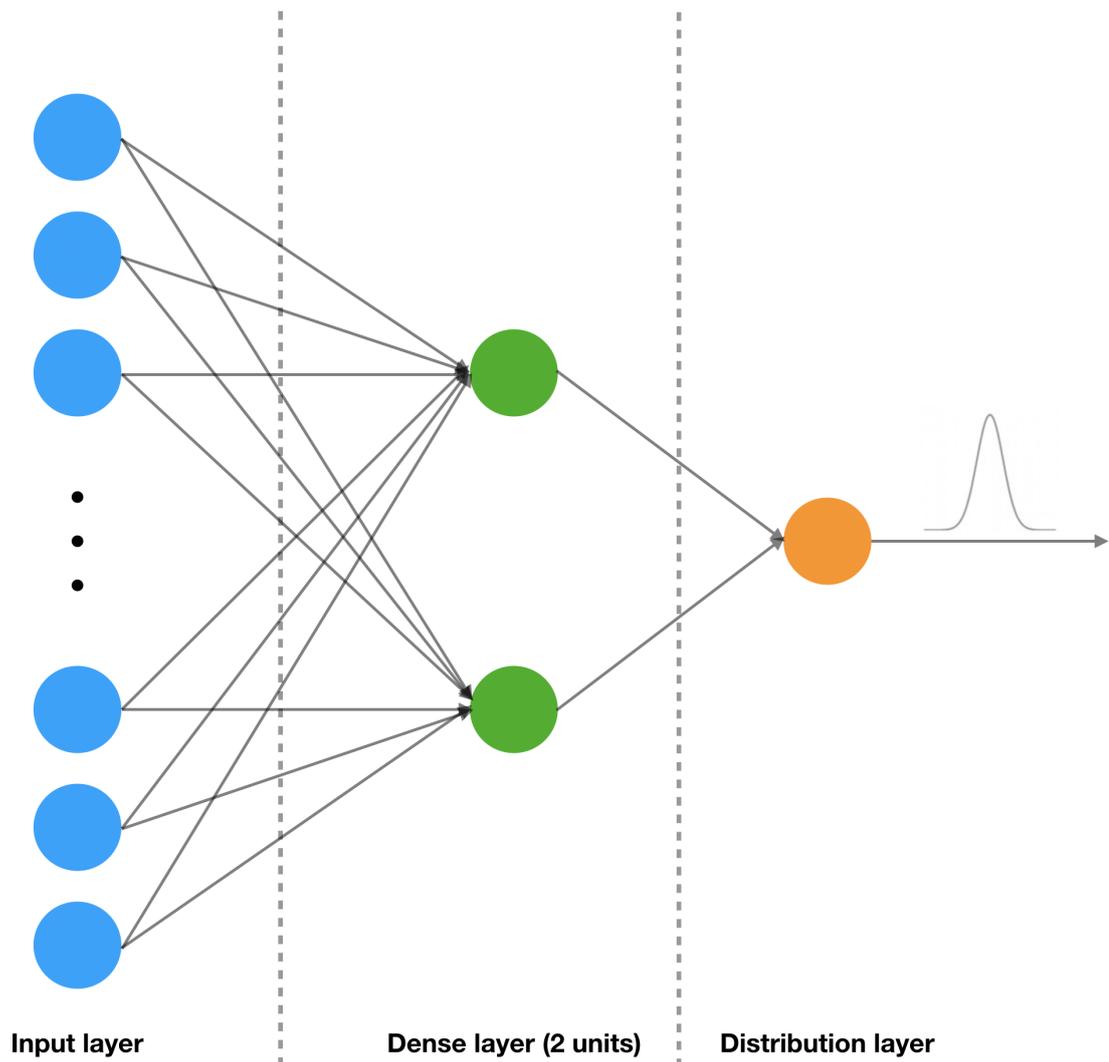


Figure 7.6: Schematic representation of the Regression model. The input layer is represented by blue dots, the hidden layer by green dots and the output layer by the orange dot.

architecture of the Classification NN. In fact it was the simplest model we tested which was able to produce accurate outputs, without overfitting.

One of the advantages that Regression models have with respect to Classification ones, is that they need a lower number of examples to produce trustworthy predictions. The training set and the validation set have been constructed separately. The former consists of 2.000 examples, while the latter of 800. Regression ML models work better if the range of possible labels is well represented in the training and validation sets, so it was not convenient to use just the 11 classes that have been used in the Classification model.

We constructed mock catalogues with 40 different values of Ω_m : 29 from $\Omega_m = 0.24$ to $\Omega_m = 0.38$ with $\Delta\Omega_m = 0.005$ and 11 from $\Omega_m = 0.2825$ to $\Omega_m = 0.3325$, still separated from each other by $\Delta\Omega_m = 0.005$. The later 11 categories of mock catalogues have been constructed to improve the density of labeled examples in the region that has proven to be the one where the predictions on the set composed of measures of BOSS correlation function were more likely to be, during the first attempts of training this NN. All the mock catalogues used during the training and validation have the same dimension of BOSS, i.e. no selection has been done on the three coordinates.

The training set consists of 50 correlation function measures for each of the 40 values of Ω_m , while the validation has been done on 20 measures for each value of Ω_m .

The loss function used during the training process is a custom one we created for this specific purpose. The analytical expression is very similar to the Binary Cross Entropy. If we define p_i as the value that the Gaussian prediction on a given example has at the value of the true label of that example, then our loss function is the following:

$$J = - \sum_{i=1}^N p_i \log p_i \quad , \quad (7.2)$$

where N is the number of examples.

We used the Adam optimization during the training process, that consisted of three different steps:

- 750 epochs with $\eta = 0.002$;
- 150 epochs with $\eta = 0.001$;
- 100 epochs with $\eta = 0.0005$.

During all the three steps we kept $\beta_1 = 0.9$ and $\beta_2 = 0.99$, and the training set was not divided into batches. Variations in the parameters of the optimization did not lead to significantly different outputs.

Gradually reducing the learning rate during the training helps the model to find the correct global minimum of the loss function (Ntampaka et al. 2019). The first epochs, having a higher learning rate, lead the model towards the area of the global minimum, while the last ones, having a lower learning rate, and therefore being able to take smaller and more precise steps in the parameter space, have the task to lead the model towards the very bottom of that minimum.

In this model, minimizing the loss function means both to drive the means of Gaussian distributions we have as output towards the values of the labels, and to reduce as much as possible the standard deviation of those distributions.

As in the case of Classification, the model was tested on a test set composed of measures obtained from mock catalogues and then applied to the set of the measures of BOSS correlation function obtained assuming all the values of Ω_m used in the construction of the mock catalogues that are in the training set.

The test set is composed of 80 different measures obtained from mock catalogues with 16 different randomly generated values of Ω_m : ten values are between 0.27 and 0.35, three are between 0.24 and 0.27 and three are between 0.35 and 0.38.

For each of these Ω_m random values, five mock catalogues have been constructed and their correlation functions measured.

During the training process the interval of labels, $0.24 < \Omega_m < 0.38$, has been mapped into $[0, 1]$ through the following linear operation:

$$L = \frac{l - 0.24}{0.14} , \quad (7.3)$$

where l is the original label and L is the one that belongs to the $[0, 1]$ interval.

Once the mean, $\mu_{[0,1]}$, and the standard deviation, $\sigma_{[0,1]}$, of the predictions have been given as output of the model, we converted them to match the original interval of the labels. The operation we performed on the mean to obtain a useful prediction is therefore the following:

$$\mu = (\mu_{[0,1]} \cdot 0.14) + 0.24 , \quad (7.4)$$

where $\mu_{[0,1]}$ is the prediction the model gives as output in the interval $[0, 1]$, and μ is its value converted to the correct interval $[0.24, 0.38]$.

As far as the standard deviation is concerned, from the theory of propagation of uncertainties we have that

$$\frac{d\mu_{[0,1]}}{d\sigma_{[0,1]}} = \frac{d\mu}{d\sigma} , \quad (7.5)$$

where σ is the standard deviation associated to μ . So the uncertainty on μ has been obtained multiplying $\sigma_{[0,1]}$ by 0.14.

7.4 Results of the Regression model

As far as Regression is concerned, for all the output Gaussian distributions we obtained as predictions on the test sets, the mean and the standard deviation have been extracted to represent the prediction on Ω_m and the uncertainty associated to it, respectively. The standard deviation represents the aleatoric, or statistical, uncertainty (Matthies 2007, Der Kiureghian and Ditlevsen 2009). The Aleatoric uncertainty reflects the noise inherent to the data (Kendall and Gal 2017, Russel and Reale 2019).

Our model is thus able to associate to every point of the input space an uncertainty on the output that depends on the intrinsic scatter of the measures of the correlation functions, and on how similar to each other are the measures relative to different values of the labels.

The uncertainty we are going to associate to every example of the test sets is thus related on how confident the model is about its example-prediction association.

The first result of the Regression model we present is the predictions on the test set, consisting of mock catalogues.

Figure 7.7 shows the predictions on the test set, compared with the true values of Ω_m the mocks in the test were created with. Those predictions, together with the errors associated to them, have been fitted with a straight line, defined by the following equation:

$$\Omega_{m,pred} = a \cdot \Omega_{m,true} + b \quad . \quad (7.6)$$

The best-fit values that were found are $a_{best} = 0.997 \pm 0.014$ and $b_{best} = 0.001 \pm 0.004$. These values are compatible with the angular coefficient and the intercept of the bisector of the first quadrant, which demonstrates that this model is able to make reliable predictions when fed with measures from mock catalogues characterized by previously unseen values of Ω_m .

The linear best-fit model is represented by the black line in Figure 7.7.

As in the case of the Classification model, the main task we wanted this model to fulfill was to make predictions on Ω_m from measures of the two-point correlation functions of the real BOSS galaxy catalogue. We remind that these predictions were required to be not significantly dependent on the geometrical distortions that come from the assumption of a particular cosmology during the measure.

We applied our trained model to 40 measures of the correlation function of BOSS, every measure has been done with a different assumption on the value of Ω_m .

Figure 7.8 shows the predictions on Ω_m from the correlation function of BOSS,

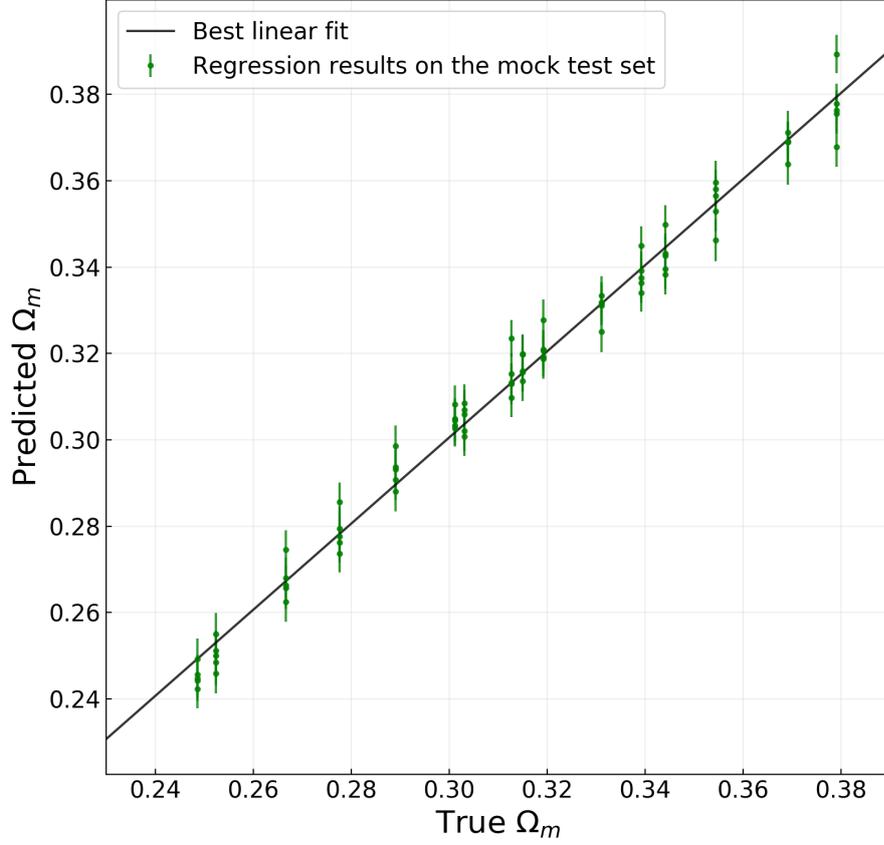


Figure 7.7: Predictions and related aleatoric uncertainties on the test set consisting of measures from mock catalogues (green dots and bars) compared with the best linear fit (black line). The best-fit values are 0.997 ± 0.014 for the angular coefficient and 0.001 ± 0.004 for the intercept.

compared with the Ω_m constraints provided by Alam et al. (2017) (see Eq. (5.7)). The predictions of the Regression model have been fitted with the following linear function:

$$\Omega_{m,pred} = a \cdot \Omega_{m,assump} + b \quad , \quad (7.7)$$

where $\Omega_{m,pred}$ is the prediction of the model and $\Omega_{m,assump}$ is the assumed value of Ω_m used to measure the correlation function.

The best fit-values of the parameters are $a_{best} = 0.01 \pm 0.02$ and $b_{best} = 0.305 \pm 0.007$. The facts that the angular coefficient of this fit is compatible with zero and that every prediction is in agreement with all the others, within the uncertainties, tell us that the model is indeed able to make reliable predictions over observed data, without being biased by the geometric distortions.

We now check the trend of the aleatoric uncertainties as a function of the value of Ω_m assumed during the measure of the correlation functions.

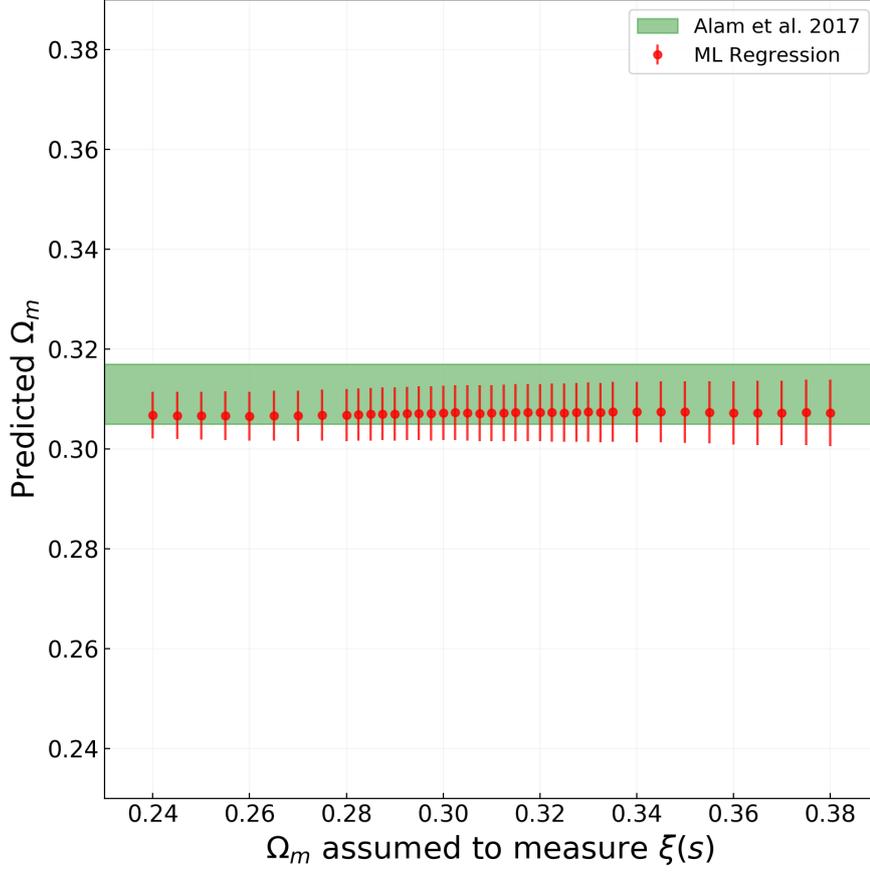


Figure 7.8: ML predictions (red dots and bars) on Ω_m from the measured correlation function of BOSS, as a function of the Ω_m assumed for the measurements. The predictions are compared to the Ω_m constraints provided by Alam et al. (2017) (green stripe).

Figure 7.9 shows this trend for both the test set and the set of measures that come from the observed BOSS catalogue.

It is important here to remember that a fundamental difference exists between a mock catalogue and BOSS: the former reflects exactly the cosmological model they were created with, while the latter uses an assumption over the underlying cosmology just to compute distances and then measure the two-point correlation function.

The values on the horizontal axis of Figure 7.9 thus represent the true values of Ω_m of the examples, the aleatoric uncertainties of which are represented by the green dots. On the other hand, the same value of the horizontal axis represents also the assumption made on the cosmological model to create the examples, the aleatoric uncertainties of which are depicted by the red dots.

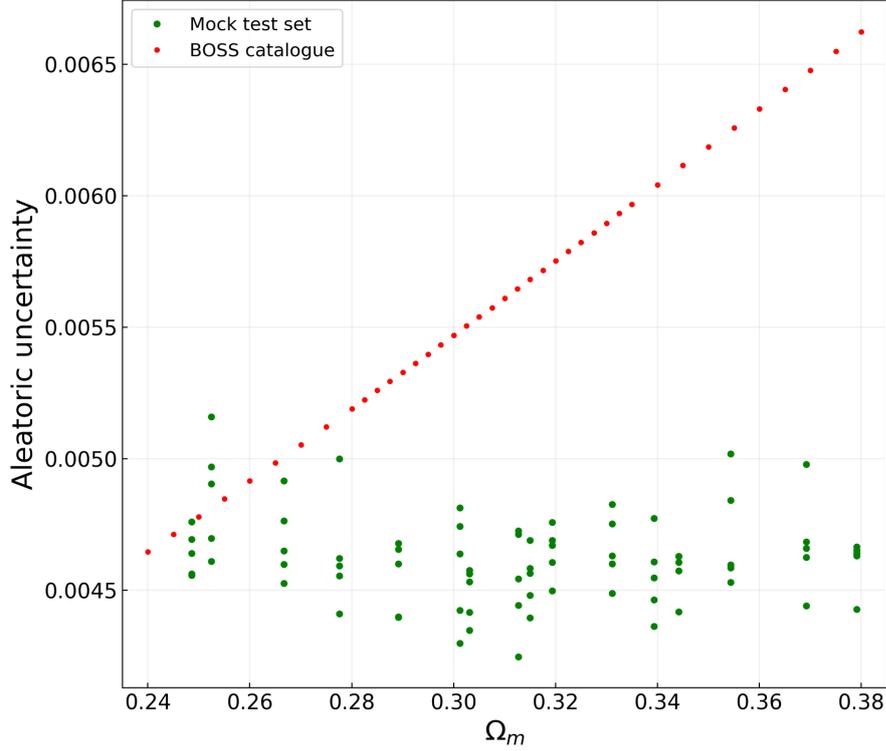


Figure 7.9: Aleatoric uncertainties on the Ω_m predictions from the mock test set (green dots) and from the set composed of measures of the BOSS correlation function (red dots), as a function of the value of Ω_m assumed during the measure of the test examples.

The aleatoric uncertainties on the predictions of the mock test set have been fitted with the following linear model:

$$\sigma = a \cdot \Omega_{m,true} + b \quad , \quad (7.8)$$

where σ is the aleatoric uncertainty on the prediction, and $\Omega_{m,true}$ is the value of Ω_m the catalogue has been created with.

The best-fit values are $a_{best} = -0.0006 \pm 0.0005$ and $b_{best} = 0.0048 \pm 0.0002$. The fact that a_{best} is compatible with 0 tells us that the model is able to associate every two-point correlation function measured starting from a mock catalogue to its prediction with the same uncertainty.

A linear fit has been performed also on the aleatoric uncertainties of the predictions on the BOSS set. In this case, the best-fit value for the angular coefficient is 0.01415 ± 0.00003 , while the one for the intercept is 0.00123 ± 0.00001 . Even in this case the angular coefficient is not far from zero, but it is clear from Figure 7.9 that a different trend exists between the red dots and the green ones. This

might be caused by the fact that the clustering in log-normal mock catalogues do not represent perfectly, at all the selected scales, the clustering of BOSS galaxies. Figure 7.10 shows the values of BOSS correlation function measured assuming $\Omega_m = 0.307$ (red dots) compared with the two-point correlation function of the model that has been assumed to construct the log-normal mocks, characterized by the same value of Ω_m . From the values of the residuals shown in the bottom panel of Figure 7.10 it is clear that there is no perfect match. We conclude that the values of the two-point correlation functions measured from BOSS represent points in the input space slightly different from the points occupied by the measures obtained from the mock catalogues. Despite this difference, that causes the different trends of the aleatoric uncertainties, the model is able to predict a reliable value of Ω_m for BOSS, which is not affected by the geometric distortions.

As far as the BOSS catalogue is concerned, the final result of this Regression model can be calculated as the mean of the Ω_m predictions.

This final result is:

$$\Omega_m = 0.307 \pm 0.006 \quad , \quad (7.9)$$

where the error has been calculated as the value the linear fit of the aleatoric uncertainty on BOSS assumes at $\Omega_m = 0.307$.

This result is remarkably consistent with the one obtained by Alam et al. (2017), that has exactly the same uncertainty, but a slightly higher expectation value.

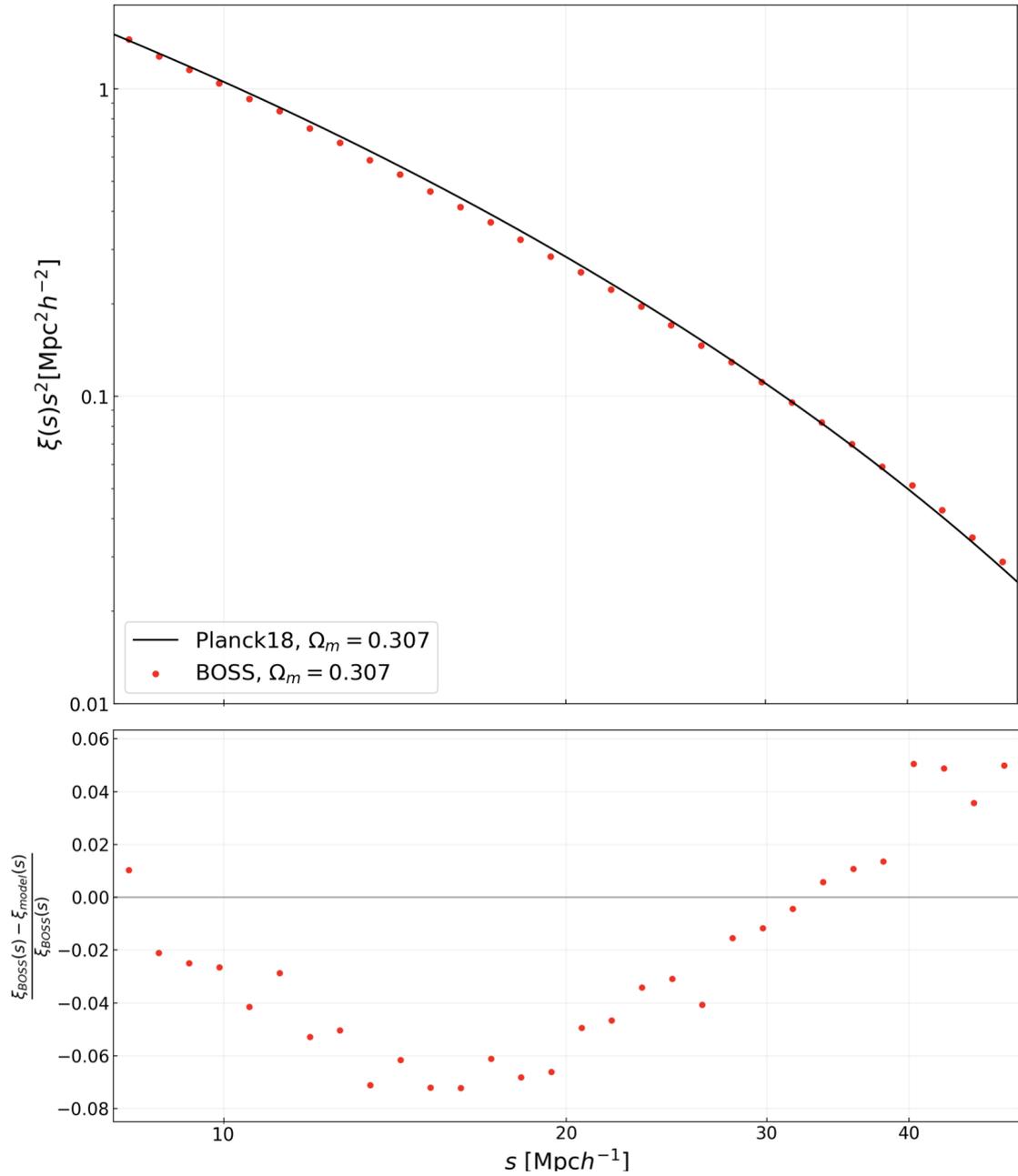


Figure 7.10: In the upper panel is plotted BOSS correlation function measured assuming $\Omega_m = 0.307$ (red dots) compared with the model calculated with the same value of Ω_m . The lower panel shows the residuals between the two.

Chapter 8

Conclusions

8.1 Final results and considerations

In this work we presented two different ML models that provide a new statistical method to obtain constraints on Ω_m from observed galaxy catalogues, without having to implement a likelihood to model the analyzed statistics. Specifically, we considered the two-point correlation function of galaxies as a summary statistic of the large-scale structure of the Universe. The applicability of these ML methodologies to observed surveys, and not only to simulations, is what makes this work innovative, and what took the biggest effort to be implemented.

We implemented and trained two supervised NNs: a Multi-Class Classification algorithm and a Regression algorithm. The steps performed in both cases are the following:

- Measure the large-scale galaxy bias that is used to construct the log-normal mock catalogues. Specifically, we estimated it as the mean of the posterior probability distribution assessed by performing a Bayesian fitting of the measured two-point correlation function of BOSS galaxies, assuming different values of Ω_m .
- Construct the log-normal mock catalogues, and measure their two-point correlation functions. The latter statistics have been used as train, validation and test sets for the NNs.
- Build the architecture of the ML model, choosing in particular how many layers it should be composed of, how many units these layers should have, which loss function to use, which optimization algorithm to be used to minimize the loss function and the values of its hyperparameters.

- Train and validate the model, paying particular attention to avoid overfitting while reaching reliable predictions on Ω_m on the train and validation sets.
- Test the trained model on the test set composed by measures of two-point correlation functions of log-normal mock catalogues, to see if the model can make reliable predictions also on previously unseen examples.
- Apply the trained model to measurements of the two-point correlation function of BOSS galaxies.

The state-of-the-art constraints on the matter density fraction of this thesis work have been compared to the ones obtained by Alam et al. (2017), that is:

$$\Omega_m = 0.311 \pm 0.006 \quad . \quad (8.1)$$

The first of the implemented NNs considered in this work, i.e. the Classification model, has been trained and validated over 88.000 log-normal mock catalogues, characterized by 11 different values of Ω_m , in the range from $\Omega_m = 0.26$ to $\Omega_m = 0.36$. This model gives as output the following constraint when applied to the BOSS galaxy catalogue:

$$\Omega_m = 0.30 \pm 0.03 \quad . \quad (8.2)$$

This result is consistent with the one from Alam et al. (2017), though the estimated uncertainty is too large to be considered competitive.

The second NN we presented is a Regression model, which has been trained and validated with a total amount of 2.800 examples, characterized by 40 different values of Ω_m , from $\Omega_m = 0.24$ to $\Omega_m = 0.38$. The output of the Regression model applied to the BOSS galaxy catalogue is:

$$\Omega_m = 0.307 \pm 0.006 \quad , \quad (8.3)$$

which is in a remarkable agreement and competitive with the constraints from Alam et al. (2017).

8.2 Future perspectives

In this thesis work, we demonstrated that NNs are powerful tools that can be used for cosmological purposes, not only applied on simulations and mock catalogues, but also on observed data. The two ML models considered represent first examples of this kind of application. There is no doubt that in the future they could be extended and improved.

One significant improvement is to use different and more accurate mock catalogues than the log-normal ones exploited in this work, in both the training and

validation phases. In particular, using N-body simulations to create the training and validation sets, we could exploit also higher-order statistics, for example the three-point correlation function, as features to feed the models with. The higher the number of features is, the more accurate the predictions of the model will be.

Multi-labeled Regression models can also be implemented in the future to make predictions on multiple cosmological parameters at the same time. Of course, in order to do that, bigger datasets will be required. In this work, for example, to train and validate the Regression model, we used 2800 examples in total. If we wanted the model to make predictions on n cosmological parameters, a set of 2800^n examples would be required, because each of them would have to be labeled with a vector containing n numbers, and not just one.

New types of ML methods have been growing in importance during the last years. One example is the Bayesian NN. The difference between this NN and the classic ones is that the weights are no more represented by a single value, but by a distribution. The user has to set a prior over all the weights and what the model does during the training is sampling their posterior. A prediction of this model is calculated by sampling a random value from the posterior of each weight. Therefore, in order to have a trustworthy output of the model for a specific example, the average value of a large number of predictions on that same example should be calculated. The dispersion of these predictions can be used as the uncertainty on the output. This kind of approach can surely be used in the next future to improve the results obtained in this thesis work.

The analysis we presented in this work can also be extended to larger scales, up to $150 \text{ Mpc}/h$, but in that case particular attention has to be put to be sure that the mock catalogues can correctly represent the observed data at the BAO scales.

In this work we used as features the values of measured two-point correlation function. A similar analysis could be done using the density map of the catalogues, or even directly the observed coordinates of the galaxies. Surely this approach would be less affected by the human interpretation of the data, such as the particular choice of using the two-point correlation function, and the assumptions on the cosmological model made during the measurements. However, we note that using the coordinates of the galaxies of a sufficient number of catalogues would require a significant amount of RAM and the learning process would need a huge amount of computation time.

We are now planning to test the Regression model with measurements of the correlation function obtained by increasingly smaller catalogues, in particular by selecting increasingly smaller ranges of RA and Dec in the BOSS sample. The trend of the aleatoric uncertainties as a function of the number of objects contained in the test catalogues will then be fitted to extrapolate the value of the uncertainty

that is expected if the model is applied to measures obtained from catalogues with more galaxies than BOSS, thus providing forecasts for next-generation cosmological analyses.

Appendix A

ML code implementation

While the construction of log-normal catalogues and the measurements of the two-point correlation functions have been performed using the built-in classes and functions of the CosmoBolognaLib (Marulli et al. 2016), the implementation, the training, the validation and the test of the NNs have been done using Keras (Chollet 2015) with a Tensorflow backend (Abaldi et al. 2016). Keras is an open-source NN library written in Python, capable of running on top of multiple backends, such as TensorFlow, Microsoft Cognitive Toolkit, R, Theano and PaidML. Keras is concerned with modularity and simplicity of development, and it was developed as part of the research effort of project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System). TensorFlow is an open-source library for dataflow and differentiable programming across a wide range of tasks. It was developed by the Google Brain team for internal Google use and it was first released on February, 2017. The codes that are about to be reported have been implemented using a Google Colaboratory (Colab) notebook. Google Colab is a free Jupyter notebook environment that requires no setup, and runs entirely in the clouds. It is optimized to run ML trainings and supports Graphic Processing Units (GPUs) and Tensor Processing Units (TPUs).

A.1 Classification algorithm

Here is presented the code used to build, train and test the Classification algorithm, once the data have been loaded on the Colab repository as numpy arrays, together with the labels.

```
1 ### Import the required modules  
2  
3 import tensorflow as tf  
4 from tensorflow import keras
```

```
5 import numpy as np
6 import matplotlib.pyplot as plt
7
8 ### Build the Neural Network
9
10 model=keras.Sequential()
11 model.add(keras.layers.Dense(8,activation=tf.nn.relu,input_shape
    =(30,)))
12 model.add(keras.layers.Dropout(0.125))
13 model.add(keras.layers.Dense(4,activation=tf.nn.relu))
14 model.add(keras.layers.Dense(11, activation=tf.nn.softmax))
15
16 ##### Compile the model
17
18 model.compile(optimizer='Adam', loss='
    sparse_categorical_crossentropy',metrics=['accuracy'])
19
20 ### Train the model
21
22 history=model.fit(train_data, train_labels, epochs=40,batch_size
    =50,validation_data=(val_data, val_labels),verbose=1)
23
24 history=history.history
25
26 ### Visualize the loss as a function of the epochs
27
28 acc = history['acc']
29 val_acc = history['val_acc']
30 loss = history['loss']
31 val_loss = history['val_loss']
32 epochs = range(1, len(acc) + 1)
33 plt.figure(figsize=(8,8))
34 plt.scatter(epochs, loss, color='red',label='Training loss',s=15)
35 plt.scatter(epochs, val_loss,color='green',marker='x', label='
    Validation loss')
36 plt.title('Training and validation loss')
37 plt.show()
38
39 ### Visualize the accuracy as a function of the epochs
40
41 plt.figure(figsize=(8,8))
42 plt.scatter(epochs, acc, color='red',label='Training accuracy',s
    =15)
43 plt.scatter(epochs, val_acc,color='green', label='Validation
    accuracy',marker='x')
44 plt.title('Training and validation accuracy')
45 plt.show()
46
47 ##### Check the performance on the test set
```

```
48
49 results = model.evaluate(test_data, test_labels)
50
51 ### Make predictions on the test set and the BOSS set
52
53 test_pred = model.predict(test_data)
54 test_pred_maxes=list()
55 for i in range(550):
56     test_pred_maxes.append(np.argmax(test_pred[i]))
57 test_pred_maxes=np.asarray(test_pred_maxes)
58
59 BOSS_pred = model.predict(BOSS_data)
60 BOSS_pred_maxes=list()
61 for i in range(11):
62     BOSS_pred_maxes.append(np.argmax(BOSS_pred[i]))
63 BOSS_pred_maxes=np.asarray(BOSS_pred_maxes)
```

The full version of the code is available at the following link:

https://colab.research.google.com/drive/1_p_pRQQ7cUxcxwqe0n6eMpLuTStl6hPp.

A.2 Regression algorithm

Just as for the Classification algorithm, we provide here the code used to build, train, validate and test the Regression NN.

```
1
2 ### Import the required modules
3
4 import tensorflow as tf
5 from tensorflow.python import tf2
6 if not tf2.enabled():
7     import tensorflow.compat.v2 as tf
8     tf.enable_v2_behavior()
9     assert tf2.enabled()
10 import tensorflow_probability as tfp
11 tfd = tfp.distributions
12 import numpy as np
13 import matplotlib.pyplot as plt
14 tfd = tfp.distributions
15
16 ### Move the labels to the [0,1] interval
17
18 norm_train_labels=(train_labels-0.24)/0.14
19 norm_val_labels=(val_labels-0.24)/0.14
20
21 ### Define the loss function
22
```

```

23 negloglik = lambda y, p_y: -p_y.log_prob(y)
24
25 ### Build the Neural Network
26
27 model = tf.keras.Sequential([
28     tf.keras.layers.Dense(1 + 1, input_shape=[30]),
29     tfp.layers.DistributionLambda(make_distribution_fn=lambda t: tfd
30         .Normal(
31             loc=t[..., :1], scale=1e-5 + tf.math.softplus(0.05 * t
32                 [..., 1:])))
33 ])
34
35 ### Compile the first phase of the training
36
37 model.compile(optimizer=tf.optimizers.Adam(learning_rate=0.002),
38               loss=negloglik, metrics=['mean_squared_error'])
39
40 ### Operate the first phase of the training
41
42 history=model.fit(train_data, norm_train_labels, epochs=750,
43                  validation_data=(val_data, norm_val_labels), verbose=0)
44
45 ### Compile the second phase of the training
46
47 model.compile(optimizer=tf.optimizers.Adam(learning_rate=0.001),
48               loss=negloglik, metrics=['mean_squared_error'])
49
50 ### Operate the second phase of the training
51
52 history=model.fit(train_data, norm_train_labels, epochs=150,
53                  validation_data=(val_data, norm_val_labels), verbose=0)
54
55 ### Compile the third phase of the training
56
57 model.compile(optimizer=tf.optimizers.Adam(learning_rate=0.0005),
58               loss=negloglik, metrics=['mean_squared_error'])
59
60 ### Operate the second phase of the training
61
62 history=model.fit(train_data, norm_train_labels, epochs=100,
63                  validation_data=(val_data, norm_val_labels), verbose=0)
64
65 ### Make predictions on the test set and the BOSS set
66
67 pred_test = model(test_data)
68 mean_test=np.squeeze(pred_test.mean())
69 std_test=np.squeeze(pred_test.stddev())
70 mean_test=(mean_test*0.14)+0.24
71 std_test=std_test*0.14

```

```
64
65 yhat_B = model(test_BOSS)
66 mean_B=np.squeeze(yhat_B.mean())
67 std_B=np.squeeze(yhat_B.stddev())
68 real_mean_B=(mean_B*0.14)+0.24
69 real_std_B=std_B*0.14
```

The full version of the code is available at the following link:

https://colab.research.google.com/drive/1eiGL287ESvU5R_g-Ns6oKgSUURvLOoQq.

Bibliography

- M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.
- U. Abbas and R. K. Sheth. The environmental dependence of galaxy clustering in the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 372(4): 1749–1754, 2006.
- P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, and et al. Planck2015 results. *Astronomy & Astrophysics*, 594:A13, Sep 2016. ISSN 1432-0746. doi: 10.1051/0004-6361/201525830. URL <http://dx.doi.org/10.1051/0004-6361/201525830>.
- N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, et al. Planck 2018 results. vi. cosmological parameters. *arXiv preprint arXiv:1807.06209*, 2018.
- S. Alam, F. D. Albareti, C. A. Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey, et al. The eleventh and twelfth data releases of the sloan digital sky survey: final data from sdss-iii. *The Astrophysical Journal Supplement Series*, 219(1):12, 2015.
- S. Alam, S. Ho, and A. Silvestri. Testing deviations from λ cdm with growth rate measurements from six large-scale structure surveys at $z=0.06-1$. *Monthly Notices of the Royal Astronomical Society*, 456(4):3743–3756, 2016.
- S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, and et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological analysis of the dr12 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 470(3):2617–2652, Mar 2017. ISSN 1365-2966. doi: 10.1093/mnras/stx721. URL <http://dx.doi.org/10.1093/mnras/stx721>.
- C. Alcock and B. Paczyński. An evolution free test for non-zero cosmological constant. *Nature*, 281(5730):358–359, 1979.

- L. Amendola, S. Appleby, A. Avgoustidis, D. Bacon, T. Baker, M. Baldi, N. Bartolo, A. Blanchard, C. Bonvin, S. Borgani, et al. Cosmology and fundamental physics with the euclid satellite. *Living reviews in relativity*, 21(1):2, 2018.
- L. Anderson, E. Aubourg, S. Bailey, F. Beutler, V. Bhardwaj, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data releases 10 and 11 galaxy samples. *Monthly Notices of the Royal Astronomical Society*, 441(1):24–62, 2014.
- F. Beutler, H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, A. S. Bolton, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, and et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the fourier space. *Monthly Notices of the Royal Astronomical Society*, 464(3):3409–3430, Sep 2016a. ISSN 1365-2966. doi: 10.1093/mnras/stw2373. URL <http://dx.doi.org/10.1093/mnras/stw2373>.
- F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, H. Gil-Marín, J. N. Grieb, N. Hand, F.-S. Kitaura, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: anisotropic galaxy clustering in fourier space. *Monthly Notices of the Royal Astronomical Society*, 466(2):2242–2260, 2016b.
- L. Blot, M. Crocce, E. Sefusatti, M. Lippich, A. G. Sánchez, M. Colavincenzo, P. Monaco, M. A. Alvarez, A. Agrawal, S. Avila, et al. Comparing approximate methods for mock catalogues and covariance matrices ii: Power spectrum multipoles. *Monthly Notices of the Royal Astronomical Society*, 485(2):2806–2824, 2019.
- L. Breuval, P. Kervella, F. Arenou, G. Bono, A. Gallenne, B. Trahin, A. Mérand, J. Storm, L. Inno, G. Pietrzynski, et al. The leavitt law of milky way cepheids from gaia dr2 static companion parallaxes. *arXiv preprint arXiv:1910.04694*, 2019.
- M. J. Brown, A. Dey, B. T. Jannuzi, T. R. Lauer, G. P. Tiede, and V. J. Mikles. Red galaxy clustering in the noao deep wide-field survey. *The Astrophysical Journal*, 597(1):225, 2003.
- F. Chollet et al. keras. github repository. <https://github.com/fchollet/keras>. Accessed on, 25:2017, 2015.
- S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, et al. The 2df galaxy redshift survey: power-spectrum analysis of the final data set and cosmological implications. *Monthly Notices of the Royal Astronomical Society*, 362(2):505–534, 2005.
- P. Coles and J. D. Barrow. Non-gaussian statistics and the microwave background radiation. *Monthly Notices of the Royal Astronomical Society*, 228(2):407–426, 1987.

- P. Coles and B. Jones. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society*, 248(1):1–13, 1991.
- M. Davis and M. J. Geller. Galaxy correlations as a function of morphological type. *The Astrophysical Journal*, 208:13–19, 1976.
- M. Davis and P. Peebles. A survey of galaxy redshifts. v—the two-point position and velocity correlations. *The Astrophysical Journal*, 267:465–482, 1983.
- M. Davis, A. Meiksin, M. A. Strauss, L. N. Da Costa, and A. Yahil. On the universality of the two-point galaxy correlation function. *The Astrophysical Journal*, 333:L9–L12, 1988.
- J. De La Calleja and O. Fuentes. Machine learning and image analysis for morphological galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 349(1):87–93, 2004.
- L. Deng, D. Yu, et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*, 7(3–4):197–387, 2014.
- A. Der Kiureghian and O. Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009.
- J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(Jul):2121–2159, 2011.
- A. Einstein. Grundgedanken der allgemeinen relativitätstheorie und anwendung dieser theorie in der astronomie. *Preussische Akademie der Wissenschaften, Sitzungsberichte*, 315:778–786, 1915.
- D. J. Eisenstein and W. Hu. Baryonic features in the matter transfer function. *The Astrophysical Journal*, 496(2):605, 1998.
- D. J. Eisenstein and W. Hu. Power spectra for cold dark matter and its variants. *The Astrophysical Journal*, 511(1):5, 1999.
- D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142(3):72, 2011.
- G. F. Ellis, R. Maartens, and M. A. MacCallum. *Relativistic cosmology*. Cambridge University Press, 2012.
- A. Friedmann. Über die krümmung des raumes. *Zeitschrift für Physik A Hadrons and Nuclei*, 10(1):377–386, 1922.

- X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- J. Gott III and E. Turner. An extension of the galaxy covariance function to small scales. *The Astrophysical Journal*, 232:L79–L81, 1979.
- J. N. Grieb, A. G. Sánchez, S. Salazar-Albornoz, R. Scoccimarro, M. Crocce, C. Dalla Vecchia, F. Montesano, H. Gil-Marín, A. J. Ross, F. Beutler, and et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: Cosmological implications of the fourier space wedges of the final sample. *Monthly Notices of the Royal Astronomical Society*, page stw3384, Jan 2017. ISSN 1365-2966. doi: 10.1093/mnras/stw3384. URL <http://dx.doi.org/10.1093/mnras/stw3384>.
- J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger, L. N. Carey, G. R. Knapp, D. G. York, W. N. Boroski, et al. The 2.5 m telescope of the sloan digital sky survey. *The Astronomical Journal*, 131(4):2332, 2006.
- A. Hamilton. Toward better ways to measure the galaxy correlation function. *The Astrophysical Journal*, 417:19, 1993.
- 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, et al. 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, 2003.
- Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555, 2017.
- B. Hoyle. Measuring photometric redshifts using galaxy images and deep neural networks. *Astronomy and Computing*, 16:34–40, 2016.
- E. Hubble. A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the national academy of sciences*, 15(3):168–173, 1929.
- M. M. Ivanov, M. Simonović, and M. Zaldarriaga. Cosmological parameters from the boss galaxy power spectrum, 2019.
- J. C. Jackson. A critique of rees’s theory of primordial gravitational radiation. *Monthly Notices of the Royal Astronomical Society*, 156(1):1P–5P, Feb 1972. ISSN 1365-2966. doi: 10.1093/mnras/156.1.1p. URL <http://dx.doi.org/10.1093/mnras/156.1.1p>.
- J. H. Jeans. I. the stability of a spherical nebula. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 199(312-320):1–53, 1902.

- N. Kaiser. On the spatial correlations of abell clusters. *The Astrophysical Journal*, 284:L9–L12, 1984.
- N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.
- E. Keihänen, H. Kurki-Suonio, V. Lindholm, A. Viitanen, A.-S. Suur-Uski, V. Allevalo, E. Branchini, F. Marulli, P. Norberg, D. Tavagnacco, et al. Estimating the galaxy two-point correlation function using a split random catalog. *arXiv preprint arXiv:1905.01133*, 2019.
- A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584, 2017.
- M. Kerscher, I. Szapudi, and A. S. Szalay. A comparison of estimators for the two-point correlation function. *The Astrophysical Journal Letters*, 535(1):L13, 2000.
- D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- F.-S. Kitaura, G. Yepes, and F. Prada. Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing. *Monthly Notices of the Royal Astronomical Society: Letters*, 439(1):L21–L25, 2013.
- S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- J. Kuruvilla and C. Porciani. On the streaming model for redshift-space distortions. *Monthly Notices of the Royal Astronomical Society*, 479(2):2256–2276, 2018.
- S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64–71, 1993.
- R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian, W. Percival, J. Rhodes, G. S. Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. L. Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. R. Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Z. Osorio, O. Boulade, J. J. Fomond, A. D. Giorgio, P. Guttridge, A. James, M. Kemp,

- J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligori, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Noddle, M. Poncet, M. Scodreggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. E. Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. D. Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid definition study report, 2011.
- A. Lewis and S. Bridle. Cosmological parameters from cmb and other data: A monte carlo approach. *Physical Review D*, 66(10):103511, 2002.
- A. Lewis and A. Challinor. Camb: Code for anisotropies in the microwave background. *Astrophysics Source Code Library*, 2011.
- C. Li, G. Kauffmann, Y. Jing, S. D. White, G. Börner, and F. Cheng. The dependence of clustering on galaxy properties. *Monthly Notices of the Royal Astronomical Society*, 368(1):21–36, 2006.
- X.-D. Li, C. Park, C. G. Sabiu, H. Park, D. H. Weinberg, D. P. Schneider, J. Kim, and S. E. Hong. Cosmological constraints from the redshift dependence of the alcock–paczynski effect: Application to the sdss-iii boss dr12 galaxies. *The Astrophysical Journal*, 832(2):103, 2016.
- E. V. Linder and A. Jenkins. Cosmic structure growth and dark energy. *Monthly Notices of the Royal Astronomical Society*, 346(2):573–583, 2003.
- S. Maddox, G. Efstathiou, W. Sutherland, and J. Loveday. Galaxy correlations on large scales. *Monthly Notices of the Royal Astronomical Society*, 242(1):43P–47P, 1990.
- F. Marulli, D. Bianchi, E. Branchini, L. Guzzo, L. Moscardini, and R. E. Angulo. Cosmology with clustering anisotropies: disentangling dynamic and geometric distortions

- in galaxy redshift surveys. *Monthly Notices of the Royal Astronomical Society*, 426(3): 2566–2580, 2012.
- F. Marulli, M. Bolzonella, E. Branchini, I. Davidzon, S. De La Torre, B. Granett, L. Guzzo, A. Iovino, L. Moscardini, A. Pollo, et al. The vimos public extragalactic redshift survey (vipers)-luminosity and stellar mass dependence of galaxy clustering at $0.5 < z < 1.1$. *Astronomy & Astrophysics*, 557:A17, 2013.
- F. Marulli, A. Veropalumbo, and M. Moresco. Cosmoboglib: C++ libraries for cosmological calculations. *Astronomy and Computing*, 14:35–42, 2016.
- H. G. Matthies. Quantifying uncertainty: modern computational representation of probability and applications. In *Extreme man-made and natural hazards in dynamics of structures*, pages 105–135. Springer, 2007.
- M. Ntampaka, H. Trac, D. J. Sutherland, N. Battaglia, B. Póczos, and J. Schneider. A machine learning approach for dynamical mass measurements of galaxy clusters. *The Astrophysical Journal*, 803(2):50, 2015.
- M. Ntampaka, D. J. Eisenstein, S. Yuan, and L. H. Garrison. A hybrid deep learning approach to cosmological constraints from galaxy redshift surveys. *arXiv preprint arXiv:1909.10527*, 2019.
- P. Ntelis, J.-C. Hamilton, J.-M. Le Goff, E. Burtin, P. Laurent, J. Rich, J. Tinker, E. Aubourg, H. D. M. Des Bourboux, J. Bautista, et al. Exploring cosmic homogeneity with the boss dr12 galaxy sample. *Journal of Cosmology and Astroparticle Physics*, 2017(06):019, 2017.
- F. Ostrovski, R. G. McMahon, A. J. Connolly, C. A. Lemon, M. W. Auger, M. Banerji, J. M. Hung, S. E. Kuposov, C. E. Lidman, S. L. Reed, et al. Vdes j2325- 5229 az= 2.7 gravitationally lensed quasar discovered using morphology-independent supervised machine learning. *Monthly Notices of the Royal Astronomical Society*, 465(4):4325–4334, 2016.
- D. Parkinson, S. Riemer-Sørensen, C. Blake, G. B. Poole, T. M. Davis, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, et al. The wigglez dark energy survey: final data release and cosmological results. *Physical Review D*, 86(10):103518, 2012.
- P. J. Peebles. The gravitational-instability picture and the nature of the distribution of galaxies. *The Astrophysical Journal*, 189:L51, 1974.
- P. J. E. Peebles. *Principles of physical cosmology*. Princeton University Press, 1993.
- S. Perlmutter, G. Aldering, S. Deustua, S. Fabbro, G. Goldhaber, D. Groom, A. Kim, M. Kim, R. Knop, P. Nugent, et al. Cosmology from type ia supernovae. *arXiv preprint astro-ph/9812473*, 1998.

- A. Pezzotta, S. de La Torre, J. Bel, B. Granett, L. Guzzo, J. Peacock, B. Garilli, M. Scodreggio, M. Bolzonella, U. Abbas, et al. The vimos public extragalactic redshift survey (vipers)-the growth of structure at $0.5 < z < 1.2$ from redshift-space distortions in the clustering of the pdr-2 final sample. *Astronomy & Astrophysics*, 604:A33, 2017.
- A. Pillepich, C. Porciani, and O. Hahn. Halo mass function and scale-dependent bias from n-body simulations with non-gaussian initial conditions. *Monthly Notices of the Royal Astronomical Society*, 402(1):191–206, 2010.
- S. Ravanbakhsh, J. Oliva, S. Fromenteau, L. C. Price, S. Ho, J. Schneider, and B. Poczos. Estimating cosmological parameters from the dark matter distribution. *arXiv preprint arXiv:1711.02033*, 2017.
- B. Reid, S. Ho, N. Padmanabhan, W. J. Percival, J. Tinker, R. Tojeiro, M. White, D. J. Eisenstein, C. Maraston, A. J. Ross, et al. Sdss-iii baryon oscillation spectroscopic survey data release 12: galaxy target selection and large-scale structure catalogues. *Monthly Notices of the Royal Astronomical Society*, 455(2):1553–1573, 2015.
- B. A. Reid, L. Samushia, M. White, W. J. Percival, M. Manera, N. Padmanabhan, A. J. Ross, A. G. Sánchez, S. Bailey, D. Bizyaev, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: measurements of the growth of structure and expansion rate at $z = 0.57$ from anisotropic clustering. *Monthly Notices of the Royal Astronomical Society*, 426(4):2719–2737, 2012.
- M. Rezaie, H.-J. Seo, A. J. Ross, and R. C. Bunesco. Improving galaxy clustering measurements with deep learning: analysis of the decals dr7 data. *arXiv preprint arXiv:1907.11355*, 2019.
- 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, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009, 1998.
- A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic. Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the hubble constant and stronger evidence for physics beyond λ cdm. *The Astrophysical Journal*, 876(1):85, 2019.
- A. J. Ross, F. Beutler, C.-H. Chuang, M. Pellejero-Ibanez, H.-J. Seo, M. Vargas-Magana, A. J. Cuesta, W. J. Percival, A. Burden, A. G. Sánchez, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: Observational systematics and baryon acoustic oscillations in the correlation function. *Monthly Notices of the Royal Astronomical Society*, 464(1):1168–1191, 2016.
- R. L. Russell and C. Reale. Multivariate uncertainty in deep learning. *arXiv preprint arXiv:1910.14215*, 2019.

- S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.
- S. Salazar-Albornoz, A. G. Sánchez, J. N. Grieb, M. Crocce, R. Scoccimarro, S. Alam, F. Beutler, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: angular clustering tomography and its cosmological implications. *Monthly Notices of the Royal Astronomical Society*, 468(3):2938–2956, 2017.
- A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM JOURNAL OF RESEARCH AND DEVELOPMENT*, pages 71–105, 1959.
- A. G. Sánchez, E. A. Kazin, F. Beutler, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, M. Manera, F. Montesano, R. C. Nichol, N. Padmanabhan, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: cosmological constraints from the full shape of the clustering wedges. *Monthly Notices of the Royal Astronomical Society*, 433(2):1202–1222, 2013.
- A. G. Sánchez, J. N. Grieb, S. Salazar-Albornoz, S. Alam, F. Beutler, A. J. Ross, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: combining correlated gaussian posterior distributions. *Monthly Notices of the Royal Astronomical Society*, 464(2):1493–1501, 2016a.
- A. G. Sánchez, R. Scoccimarro, M. Crocce, J. N. Grieb, S. Salazar-Albornoz, C. D. Vecchia, M. Lippich, F. Beutler, J. R. Brownstein, C.-H. Chuang, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: cosmological implications of the configuration-space clustering wedges. *Monthly Notices of the Royal Astronomical Society*, 464(2):1640–1658, 2016b.
- S. Satpathy, S. Alam, S. Ho, M. White, N. A. Bahcall, F. Beutler, J. R. Brownstein, C.-H. Chuang, D. J. Eisenstein, J. N. Grieb, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: On the measurement of growth rate using galaxy correlation functions. *Monthly Notices of the Royal Astronomical Society*, 469(2):1369–1382, 2017.
- N. Schanche, A. C. Cameron, G. Hébrard, L. Nielsen, A. Triaud, J. Almenara, K. Alsubai, D. Anderson, D. Armstrong, S. Barros, et al. Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. *Monthly Notices of the Royal Astronomical Society*, 483(4):5534–5547, 2018.
- Z. Slepian, D. J. Eisenstein, J. R. Brownstein, C.-H. Chuang, H. Gil-Marín, S. Ho, F.-S. Kitaura, W. J. Percival, A. J. Ross, G. Rossi, et al. Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of sdss boss dr12 cmass galaxies. *Monthly Notices of the Royal Astronomical Society*, 469(2):1738–1751, 2017.

- M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro, M. S. Vogeley, D. H. Weinberg, I. Zehavi, A. Berlind, et al. The three-dimensional power spectrum of galaxies from the sloan digital sky survey. *The Astrophysical Journal*, 606(2):702, 2004.
- T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.
- H. Totsuji and T. Kihara. The correlation function for the distribution of galaxies. *Publications of the Astronomical Society of Japan*, 21:221, 1969.
- M. Vargas-Magaña, S. Ho, A. J. Cuesta, R. O’Connell, A. J. Ross, D. J. Eisenstein, W. J. Percival, J. N. Grieb, A. G. Sánchez, J. L. Tinker, et al. The clustering of galaxies in the completed sdss-iii baryon oscillation spectroscopic survey: theoretical systematics and baryon acoustic oscillations in the galaxy correlation function. *Monthly Notices of the Royal Astronomical Society*, 477(1):1153–1188, 2018.
- A. Veropalumbo, F. Marulli, L. Moscardini, M. Moresco, and A. Cimatti. Measuring the distance–redshift relation with the baryon acoustic oscillations of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 458(2):1909–1920, Feb 2016. ISSN 1365-2966. doi: 10.1093/mnras/stw306. URL <http://dx.doi.org/10.1093/mnras/stw306>.
- N. Vittorio. *Cosmology*. Astronomy and Astrophysics. CRC Press, Taylor & Francis Group, 2017. ISBN 9781498731324. URL <https://books.google.it/books?id=WvNbMQAACAAJ>.
- M. White, J. L. Tinker, and C. K. McBride. Mock galaxy catalogues using the quick particle mesh method. *Monthly Notices of the Royal Astronomical Society*, 437(3):2594–2606, 2013.
- C. N. Willmer, L. N. da Costa, and P. S. Pellegrini. Southern sky redshift survey: Clustering of local galaxies. *The Astronomical Journal*, 115(3):869, 1998.
- H. S. Xavier, F. B. Abdalla, and B. Joachimi. Improving lognormal models for cosmological fields. *Monthly Notices of the Royal Astronomical Society*, 459(4):3693–3710, 2016.
- I. Zehavi, D. H. Weinberg, Z. Zheng, A. A. Berlind, J. A. Frieman, R. Scoccimarro, R. K. Sheth, M. R. Blanton, M. Tegmark, H. J. Mo, et al. On departures from a power law in the galaxy correlation function. *The Astrophysical Journal*, 608(1):16, 2004.
- M. D. Zeiler. Adadelat: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

- Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh, S. Cole, R. Dave, C. S. Frenk, N. Katz, and C. G. Lacey. Theoretical models of the halo occupation distribution: Separating central and satellite galaxies. *The Astrophysical Journal*, 633(2):791–809, Nov 2005. ISSN 1538-4357. doi: 10.1086/466510. URL <http://dx.doi.org/10.1086/466510>.
- F. Zwicky. On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, 86:217, 1937.

Ringraziamenti

Dedico questo mio lavoro a tutte le persone che mi hanno accompagnato in questo stupendo viaggio durato più di due anni e senza le quali questo stesso lavoro non avrebbe mai potuto vedere la luce.

Il primo grazie va a Federico, il mio relatore, che con estrema pazienza e competenza ha saputo spronarmi e spingermi a dare il meglio di me stesso, facendo sempre il proverbiale "miglio in più" per non farmi mai mancare nulla. Grazie per tutte le ore di tempo che mi hai dedicato, sono state tanto apprezzate quanto provvidenziali.

Grazie anche ai miei correlatori: Alfonso e il Prof. Moscardini, i quali hanno contribuito in maniera tutt'altro che secondaria alla buona riuscita del progetto, senza mai farmi mancare l'appoggio di cui ho avuto bisogno in questi ultimi mesi.

Come potrei mai dimenticarmi di ringraziare i miei compagni di corso, o meglio, i miei compagni d'arme. Siete stati il meglio che potessi chiedere, sempre pronti a soffrire insieme... anche a gioire e ridere eh... ma soprattutto soffrire, non nascondiamocelo. Grazie, gruppo Dark Matter.

Ovviamente un caldo ringraziamento va alla mia mamma Alessandra, al mio papà Roberto, alla mia sorella Ottavia e alle mie nonne Alfonsina e Maria. Siete sempre stati per me un esempio di impegno, sacrificio e dedizione, tutte cose che, se sono riuscito a mettere in pratica in questo lavoro di tesi, è solo grazie a voi.

Infine grazie Nicole, il tuo sostegno e il tuo amore sono i motori che mi spingono a non accontentarmi mai, a cercare sempre di raggiungere la vetta più alta, perchè so che su qualsiasi vetta io raggiunga, non importa la fatica fatta per arrivarci, ci sarai tu, sempre al mio fianco.