

# Summary

The fascination with the night sky. A simple gaze at the wonderful landscape above us opens a window toward imagination. This fascination together with our innate human curiosity paved the passion for understanding the world we live in. From ancient cultures to nowadays, we were all concerned about the same questions: “*Where did we come from?*” and “*What is our place in the Universe?*”.

Understanding and learning more about the cosmos is not only appealing to those who dedicate their lives to science but also to the lay population in general. A quick search within the most popular forums on the internet demonstrates that all have threads up-to-date dedicated to conversations about the universe<sup>12</sup> Moreover, every time there is an exciting astronomical discovery, it spreads within a few hours, as was the case for the first-ever image captured by the Event Horizon Telescope (EHT) of the supermassive black hole Sgr A\*.

The composition of the universe, as well as its believed origin, and the main force dominating, gravity, are part of the popular culture of our civilization. But, how is (and was) the Universe studied? We rely on astronomical observations of, for instance, stars and other massive compact objects present in the Universe, to build our scientific knowledge based on the comparison of those observations with some physical assumptions, which form a cosmological model. For example, an average lay citizen is familiar with the concept of Big Bang.

Indeed, cosmologists believe that the universe was formed approximately 13.77 billion years ago from a very dense and hot state. The young Universe cooled down while it expanded, and it is still expanding today at a faster rate. Our Universe is composed of two main ingredients: an unknown matter substance that interacts gravitationally denominated dark matter and the agent responsible of the accelerated expansion of the universe called dark energy. The matter we are all made of, baryonic matter, accounts only for approximately 5% of the content of the universe. Therefore, it is not surprising that concepts such as dark energy and dark matter are surrounded by a halo of mystery, which attracts the interest of the public. Talking in scientific terms, most of the scientific knowledge about the composition of the universe was obtained from observations of the cosmic microwave background: the relic radiation

---

<sup>12</sup>See for instance <https://www.quora.com/Why-do-you-love-to-gaze-at-the-stars-of-the-night-sky>.

emitted in the early universe when the first neutral atoms were formed and photons could finally scatter away and move freely throughout the Universe. From these observations, we concluded that our Universe is flat.

As the Universe cooled down and expanded, gravity started to play a more significant role and the current structures observed in our Universe started to form: from the first stars to the first galaxies, and from the first galaxies to the first clusters of galaxies. These objects formed the so-called large scale structure of the universe, which looks like a web. However, gravity is not the only player in the formation of the large scale structure of the universe. We need the existence of an underlying Gaussian distribution of primordial density perturbations that can explain the origin of the current structure. So far, inflation, an exponential expansion phase in the primordial universe, is believed to be the mechanism of production of these primordial seeds. The simplest model of inflation is compatible with the assumption of a Gaussian distribution of the primordial density perturbations, apart from explaining why the cosmological observations indicate that the universe is causally connected and flat. The inflationary paradigm together with the assumption that our Universe is mostly composed of dark matter and dark energy are the basis of the standard cosmological model.

In the coming years, the cosmological observations available to study the Universe are expected to increase. In particular, we will exploit the information encoded in the large scale structure of the universe. For that, new missions and experiments are focusing on creating large galaxy surveys that will contain information about their positions in the sky, their shapes and also their redshifts. With the catalogues, cosmologists will study the composition of the universe using observables such as galaxy clustering (GC) and weak lensing (WL). Among these experiments, the European Space Agency medium-size *Euclid* mission, whose satellite is expected to be launched in the near future, aims to understand the physical origin of the accelerated expansion of the universe, as well as understand the nature of dark matter. *Euclid* will also study the initial conditions that seeded the early universe and were responsible for the formation of the cosmic-web structure. For that, *Euclid* will map the large scale structure of the universe by creating one of the largest galaxy catalogues ever known.

What is the methodology used to compare astrophysical observations with physical models to extract conclusions about the universe we live in? Cosmologists have used statistics, in particular, a Bayesian statistical approach, during the last 30 years to test different cosmological models against data. Bayesian statistics are based on Bayes' theorem. This theorem tells us that our degree of belief can be encoded in a probability distribution. Bayesian statistics assume that the probability distribution of some parameters of a model, given some observed data, is directly proportional to the product of two probability distributions: the distribution that gives the probability of observing the data given the model  $M$  and the parameter values, times the probability distribution that encodes some prior knowledge information known about the theory or the experiment. In the last decades, the success of Bayesian statis-

tics applied to cosmology is mostly due to the exponential increase in computational power that made massive numerical inference feasible for the first time.

Bayesian statistics can be used for a large set of scientific questions. For instance, it allows us to constrain, given a model and some data, the best fit values of the parameters of that model. On top of that, with this statistical approach we can also discern whether an alternative cosmological model is statistically favoured with respect to the standard cosmological model. Moreover, Bayesian statistics are also useful in forecasting analyses, where we aim to test how sensitive a future experiment will be to the possible detection of a new observable or even, given the experimental set-up of a future mission, if there is any chance that some extensions of the standard cosmological model can be ruled out statistically.

This thesis is dedicated to the Bayesian statistical analyses of extensions of the standard cosmological model using several astronomical data sets, and to the forecast of new observables or experiments. The use of this methodology is the common motto in all the chapters of the thesis. After **Chapter 1**, where we introduce all the main concepts of cosmology as well as the basics of Bayesian statistics, the thesis is divided into three different parts depending on which goal the Bayesian statistical methodology was used for:

- The first part focuses on data science and inflation, and it aims to constrain inflationary models using advanced inference techniques and forecasting tools. **Chapter 2** shows the first-ever results of the reconstruction of the speed of sound of the field responsible for driving inflation, using the latest cosmic microwave background (CMB) data from Planck 2018 and modern algorithms (Gaussian processes). **Chapter 3** is dedicated to the forecast of a particular class of single-field inflation models, known as  $\alpha$ -attractors, for a future CMB stage-IV experiment using a model-dependent alternative approach for the sampling of the inflationary parameters based on current constraints obtained by cosmic microwave background and large scale structure data.
- The second part of the thesis is dedicated to the novel concept in cross-correlations of gravitational-wave (GW) physics and large scale structure observables; in particular, galaxy clustering. In two projects we study how we can exploit the information contained in these new observables by forecasting their behaviour and possible detection using future experimental set-ups and Bayesian statistics. **Chapter 4** studies unresolved GW events that form the astrophysical gravitational-wave background, and how we can use the cross-correlation of the anisotropies of that background with galaxy clustering to extract both astrophysical and cosmological information. On the other hand, in **Chapter 5** we investigate how machine learning techniques can be used to reconstruct the propagation of tensor perturbations by combining the spatial correlation between resolved GW mergers and galaxies.

- The third part of this thesis is dedicated to the *Euclid* mission and the tasks of the Euclid Consortium: goals, main survey features and the primary *Euclid* observational probes. In particular, **Chapter 6** focuses on a crucial data science analysis software for the mission: the code *Cosmological Likelihood for Observables in Euclid*, also known as CLOE. The Bayesian analysis tool CLOE is designed to be able to constrain the values of the parameters of a model given the future *Euclid* data through performing Bayesian inference. In this chapter, we describe the implemented cosmological recipe of the primary probes, as well as the description of the *Euclid* likelihood. The results concerning the structure of CLOE and its performance to constrain cosmological parameters are a preview of the current state of the *Euclid* mission, for which I have dedicated a vast percentage of my time.

# Samenvatting

De fascinatie voor de nachtelijke hemel. Een eenvoudige blik op het prachtige landschap boven ons opent een venster naar de verbeelding. Uit deze fascinatie samen met onze aangeboren menselijke nieuwsgierigheid, is de passie voor het begrijpen van de wereld waarin wij leven geboren. Van oude culturen tot nu, dezelfde vragen hielden ons bezig: "Waar komen we vandaan?" en "Wat is onze plaats in het heelal?".

Het begrijpen en bestuderen van de kosmos is een onderwerp dat niet alleen wetenschappers, maar ook de bevolking in het algemeen. Een snelle zoektocht op de populairste fora op het internet toont aan dat ze allemaal actieve berichten hebben die gewijd zijn aan gesprekken over het heelal<sup>13</sup>. Bovendien wordt elke spannende astronomische ontdekking binnen een paar uur verspreid, zoals het geval was bij de allereerste foto van het superzware zwarte gat Sgr A\*, gemaakt door de Event Horizon Telescope (EHT).

De samenstelling van het heelal, de vermoedelijke oorsprong ervan en de belangrijkste dominante kracht, de zwaartekracht, maken deel uit van de popcultuur van onze beschaving. Maar hoe wordt (en werd) het heelal bestudeerd? We vertrouwen op astronomische waarnemingen, bijvoorbeeld van sterren en andere massieve, compacte objecten in het heelal, om onze wetenschappelijke kennis op te bouwen, gebaseerd op de vergelijking van die waarnemingen met voorspellingen gebaseerd op een aantal fysische veronderstellingen, die een kosmologisch model vormen. De gemiddelde leek is bijvoorbeeld bekend met het begrip "Big Bang".

Kosmologen hebben gevonden dat het heelal ongeveer 13,77 miljard jaar geleden is ontstaan uit een zeer dichte en hete toestand. Het jonge heelal koelde af terwijl het uitdijde, en het dijt nu nog steeds sneller uit. Ons heelal is opgebouwd uit twee hoofdbestanddelen: een onbekende materie die gravitationeel op elkaar inwerkt, genaamd donkere materie, en de stof die verantwoordelijk is voor de versnelde uitdijing van het heelal, donkere energie. De materie waar wij allemaal van gemaakt zijn, de baryonische materie, maakt slechts ongeveer 5% uit van de inhoud van het heelal. Het is dan ook niet verwonderlijk dat begrippen als donkere energie en donkere materie omgeven zijn door een waas van mysterie die de belangstelling van het publiek wekt. Wetenschappelijk gezien is het grootste deel van de kennis over de samenstelling van

---

<sup>13</sup>Zie bijvoorbeeld <https://www.quora.com/Why-do-you-love-to-gaze-at-the-stars-of-the-night-sky>.

het heelal verkregen door waarnemingen van de kosmische achtergrondstraling (Cosmic Microwave Background): de reststraling die in het vroege heelal is uitgezonden toen de eerste neutrale atomen werden gevormd en fotonen eindelijk konden verstrooien en zich vrij door het heelal konden bewegen. Uit deze waarnemingen hebben we ook geconcludeerd dat ons heelal geometrisch vlak is.

Naarmate het heelal afkoelde en uitdijde, begon de zwaartekracht een grotere rol te spelen en ontstonden de huidige structuren die in ons heelal waar te nemen zijn: van de eerste sterren tot de eerste sterrenstelsels, en van de eerste sterrenstelsels tot de eerste clusters van sterrenstelsels. Deze objecten vormden de zogenaamde grootschalige structuur van het heelal, die eruitziet als een web. De zwaartekracht is echter niet de enige speler bij de vorming van de grootschalige structuur van het heelal. Er moet een onderliggende Gaussische verdeling van primordiale dichtheidsverstoringen bestaan die de oorsprong van de huidige structuur kan verklaren. Tot nu toe wordt aangenomen dat inflatie, een exponentiële expansiefase in het oerheelal, het mechanisme is voor de productie van deze primordiale kiemen. Het eenvoudigste model van inflatie is verenigbaar met de veronderstelling van een Gaussische verdeling van de primordiale dichtheidsverstoringen, en verklaart tevens waarom de kosmologische waarnemingen erop wijzen dat het heelal causaal verbonden en vlak is. Het inflatieparadigma vormt samen met de aanname dat ons heelal voor het grootste deel uit donkere materie en donkere energie bestaat, de basis van het standaard kosmologisch model.

In de komende jaren zullen naar verwachting steeds meer kosmologische waarnemingen beschikbaar komen om het heelal te bestuderen. In het bijzonder zal gebruik worden gemaakt van de informatie die is gecodeerd in de grootschalige structuur van het heelal. Daartoe zijn nieuwe missies en experimenten gericht op het maken van grote verzamelingen van gegevens van sterrenstelsels, die informatie zullen bevatten over hun posities aan de hemel, hun vormen en ook hun roodverschuivingen. Met de catalogi zullen kosmologen de samenstelling van het heelal bestuderen aan de hand van waarnemingen zoals clustering van sterrenstelsels (GC) en zwakke lenseffecten (WL). Een van deze experimenten, de middelgrote *Euclid* missie van de Europese ruimtevaartorganisatie (European Space Agency), waarvan de satelliet naar verwachting binnenkort zal worden gelanceerd, is erop gericht de fysische oorsprong van de versnelde uitdijning van het heelal te begrijpen en inzicht te krijgen in de aard van donkere materie. De *Euclid* missie zal ook de begincondities bestuderen die aan de basis lagen van het ontstaan van het heelal en die verantwoordelijk waren voor de vorming van de structuur van het kosmische web. Daartoe zal *Euclid* de grootschalige structuur van het heelal in kaart brengen door een van de grootste catalogi van sterrenstelsels ooit te maken.

Wat is de methodologie die wordt gebruikt om astrofysische waarnemingen te vergelijken met fysische modellen om conclusies te trekken over het heelal waarin we leven? Kosmologen hebben in de laatste 30 jaar gebruik gemaakt van statistiek, in het bijzonder van een Bayesiaanse statistische benadering, om verschillende kos-

mologische modellen te toetsen aan data. Bayesiaanse statistiek is gebaseerd op het theorema van Bayes. Dit theorema zegt ons dat onze graad van geloof kan worden gecodeerd in een waarschijnlijkheidsverdeling. De Bayesiaanse statistiek gaat ervan uit dat de kansverdeling van een aantal parameters van een model, gegeven een aantal geobserveerde gegevens, recht evenredig is met het product van de kansverdeling die de waarschijnlijkheid geeft van het observeren van de gegevens gegeven het model  $M$  en de parameterwaarden maal de kansverdeling die een aantal *a priori* bekende informatie over de theorie of het experiment codeert. In de laatste decennia is het succes van de Bayesiaanse statistiek toegepast op de kosmologie vooral te danken aan de exponentiële toename van de rekenkracht van computers die massale numerieke inferentie voor het eerst haalbaar maakte.

Bayesiaanse statistiek kan worden gebruikt voor een groot aantal wetenschappelijke vraagstukken. Het stelt ons bijvoorbeeld in staat om, gegeven een model en enkele gegevens, de best passende waarden van de parameters van dat model te bepalen. Daarnaast kunnen wij met deze statistische benadering ook nagaan of een alternatief kosmologisch model volgens de statistiek de voorkeur geniet ten opzichte van het standaard kosmologisch model. Bovendien zijn Bayesiaanse statistieken ook nuttig in voorspellingsanalyses, waarin we willen testen hoe gevoelig een toekomstig experiment zal zijn voor de mogelijke detectie van een nieuw waarneembaar object of zelfs, gegeven de experimentele opzet van een toekomstige missie, of er een kans is dat sommige uitbreidingen van het standaard kosmologisch model statistisch kunnen worden uitgesloten.

Dit proefschrift is gewijd aan de Bayesiaanse statistische analyses van uitbreidingen van het standaard kosmologisch model met behulp van verschillende astronomische datasets, en aan de voorspelling van nieuwe waarneembare gegevens of experimenten. Het gebruik van deze methodologie is de rode draad van het proefschrift. Na hoofdstuk 1, waarin we de belangrijkste concepten van de kosmologie en de basisprincipes van de Bayesiaanse statistiek introduceren, is het proefschrift verdeeld in drie verschillende delen, afhankelijk van het doel waarvoor de Bayesiaanse statistische methodologie werd gebruikt:

- Het eerste deel richt zich op datawetenschap en inflatie, en het heeft als doel inflatiemodellen te begrenzen met behulp van geavanceerde inferentietechnieken en voorspellingsinstrumenten. **Hoofdstuk 2** toont de allereerste resultaten van de reconstructie van de geluidssnelheid van het veld dat verantwoordelijk is voor de inflatie, met behulp van de nieuwste gegevens over de kosmische achtergrondstraling (CMB) van Planck 2018 en moderne algoritmes (Gaussische processen). **Hoofdstuk 3** is gewijd aan de voorspelling van een bepaalde klasse van inflatiemodellen met één veld, bekend als -attractoren, voor een toekomstig CMB fase-IV-experiment met behulp van een model-afhankelijke, alternatieve benadering voor de verzameling van de inflatieparameters op basis van de huidige beperkingen die verkregen zijn met gegevens over de kosmische achtergrondstraling en de grote schaal structuur.

- Het tweede deel van het proefschrift is gewijd aan de nieuwe interesse in de cross-correlaties van gravitatie-golf (GW) fysica en grote schaal structuur waarnemingen; en in het bijzonder, clustering van sterrenstelsels. In deze twee projecten bestuderen we hoe we de informatie in deze nieuwe waarneemgegevens kunnen exploiteren door hun gedrag en mogelijke detectie te voorspellen met behulp van toekomstige experimenten en Bayesiaanse statistiek. In **hoofdstuk 4** bestuderen we de ruis van overlappende GW signalen die de astrofysische zwaartekrachtgolfachtergrond vormen, en hoe we de correlatie tussen de structuur van die achtergrond en de clustering van melkwegstelsels kunnen gebruiken om zowel astrofysische als kosmologische informatie te verkrijgen. Anderzijds onderzoeken we in **hoofdstuk 5** hoe machine learning technieken kunnen worden gebruikt om de propagatie van tensorverstoringen te reconstrueren door de ruimtelijke correlatie tussen individueel herkenbare GW bronnen en melkwegstelsels te combineren.
- Het derde deel van dit proefschrift is gewijd aan de *Euclid* missie en de taken van het Euclid Consortium: doelstellingen, hoofdkenmerken van de survey en de primaire *Euclid*-observatiesondes. In het bijzonder wordt in **hoofdstuk 6** aandacht besteed aan cruciale software voor de data-analyse voor de missie: de code *Cosmological Likelihood for Observables in Euclid*, ook bekend als CLOE. Dit Bayesiaanse analyse-instrument is ontworpen om de parameterwaarden van een model te kunnen afleiden uit de toekomstige metingen van de *Euclid* missie. In dit hoofdstuk beschrijven we het geïmplementeerde kosmologische recept van de primaire sondes, evenals de beschrijving van de statistische waarschijnlijkheid van *Euclid* data. De resultaten met betrekking tot de structuur van CLOE en hoe goed de code presteert bij het beperken van kosmologische parameters zijn een voorproefje van wat we hopen te leren van de *Euclid* missie, ik een groot deel van mijn tijd heb besteed.