Cosmic Dissonance

Addressing tensions in modern cosmology

Dissertation submitted for the degree of

Philosophiæ Doctor

to the PhD School of the Faculty of Science,
University of Copenhagen

Nikki Arendse

Examiners
Dr. Eleonora di Valentino
Dr. Philip James Marshall
Assoc. Prof. Steen Harle Hansen

Supervisors
Prof. Jens Hjorth
Dr. Radosław Jan Wojtak

PUBLICATIONS RELATED TO THIS THESIS

This thesis has used material from the following first- and co-authored publications:

- *"Low-redshift measurement of the sound horizon through gravitational time-delays,"*
  Nikki Arendse, Adriano Agnello & Radosław J. Wojtak.
  A&A, 632, A91 (2019). ArXiv: 1905.12000.

- *"Cosmic dissonance: are new physics or systematics behind a short sound horizon?,"*
  Nikki Arendse, Radosław J. Wojtak, Adriano Agnello, Geoff C.-F. Chen, Christopher D. Fassnacht, Dominique Sluse, Stefan Hilbert, Martin Millon, Vivien Bonvin, Kenneth C. Wong, Frédéric Courbin, Sherry H. Suyu, Simon Birrer, Tommaso Treu & Leon V.E. Koopman.
  A&A, 639, A57 (2020). ArXiv: 1909.07986.

- *"AI-driven spatio-temporal engine for finding gravitationally lensed supernovae,"*
  Doogesh Kodi Ramanah, **Nikki Arendse** & Radosław J. Wojtak.
  Under review at Nature Astronomy. ArXiv: 2107.12399.

- *"Inferring the Hubble constant from lensed supernovae in LSST with spatio-temporal neural networks,"*
  Nikki Arendse, Doogesh Kodi Ramanah & Radosław J. Wojtak.
  Work in progress.

- *"Simulation-based inference of dynamical galaxy cluster masses with 3D convolutional neural networks,"*
  Doogesh Kodi Ramanah, Radosław J. Wojtak & **Nikki Arendse**
  MNRAS, 476, 2825 (2021). ArXiv: 2009.03340.

Abstract

In the early Universe, primordial density perturbations left their imprints on the first visible light, the Cosmic Microwave Background (CMB) radiation. While the Universe expanded and cooled down, the density fluctuations evolved under the influence of gravity into stars, galaxies and large-scale cosmic structures, some of which can be employed today as distance indicators. When these distance indicators and the CMB are used independently to infer the cosmic expansion rate, they yield conflicting results. This discrepancy, often referred to as the 'Hubble tension', constitutes one of the key mysteries in present-day cosmology.

The aim of this thesis is to investigate the aforementioned tension by means of existing data sets and forecasts for future observations. In order to do so, the presented doctoral research lies at the intersection of observations and theory, while drawing from the latest advances in the field of machine learning (ML). The thesis is organised in three main parts.

In the first part, we employ observations of gravitationally lensed quasars, type Ia supernovae, Baryon Acoustic Oscillations, and the CMB to study the tension in a broader framework, in terms of both the Hubble constant and the sound horizon. Moreover, we investigate whether new cosmological models can resolve the combined tension and find that none of the modifications to our standard model manage to do so. These findings highlight the importance of novel independent measurements of the cosmic expansion rate.

The second part of the thesis explores one such new avenue to infer the Hubble constant: the use of gravitationally lensed supernovae for distance measurements. For this purpose, we develop a deep learning framework to identify lensed supernovae from optical transient surveys by means of both their spatial and time-variable features. We demonstrate the improvement in classification accuracy when using time-series images instead of single-epoch observations. Additionally, we present a proof of concept of a similar spatio-temporal ML pipeline to infer the Hubble constant from simulated time-series images of lensed supernovae as expected from next-generation surveys.

Finally, the third part of the thesis shifts the focus to galaxy clusters, the most massive gravitationally bound cosmic structures, and their potential for cosmological inference. We design a deep learning mass estimator that exploits the full 3D projected phase-space distribution of galaxy clusters. As a result, the mass estimator presently yields the most precise cluster mass estimates among the recent ML-based methods. For the first time, such an ML-based mass estimator is employed to construct a cluster mass function from real galaxy cluster observations, yielding results that are consistent with predictions from the standard cosmological model.

Dansk resumé

I det tidlige univers efterlod de oprindelige densitetspertubationer deres aftryk på det første synlige lys – den kosmiske mikrobølgebaggrundsstråling (CMB). Som Universet udvidede sig og kølede ned, udviklede densitetsfluktuationerne sig under indflydelse fra tyngdekraften til stjerner, galakser og storskalastrukturer, hvoraf nogle af disse i dag kan anvendes til afstandsmål. Når disse afstandsmålere og CMB'en bruges uafhængigt til at bestemme den kosmiske ekspantionsrate, giver de modstridende resultater. Denne uoverenstemmelse som ofte kaldes "Hubble tension", udgør en af hovedmysterierne i nutidens kosmologi.

Målet med denne afhandling er at udforske den førnævnte uoverenstemmelse, ved hjælp af eksisterende datasæt og forudsigelser af fremtidige observationer. For at gøre dette placerer denne forskning sig i intersektionen mellem observationer og teori, mens den trækker på de nyeste udviklinger inden for machine learning (ML). Denne afhandling består af 3 hoveddele.

I den første del anvender vi observationer af gravitationelt linsede kvasarer til at studere uoverenstemmelsen i en større sammenhæng – både i forhold til Hubblekonstanten og lydhorisonten. Derudover undersøger vi om de nye kosmologiske modeller kan løse uoverenstemmelsen, og vi finder frem til at ingen af modifikationerne til vores standardmodel er en løsning. Disse resultater tydeliggør vigtigheden af nye, uafhængige målinger af den kosmiske ekspantionsrate.

Anden del af afhandlingen udforsker én ny vej til bestemme Hubblekonstanten: Brugen af gravitationelt linsede supernovaer til afstandsmålinger. For at støtte os i disse bestræbelser, udvikler vi et deep learning framework til at identificere linsede supernovaer fra optiske observationer af transienter ved at udnytte både deres rumlige og tidsligt varierende egenskaber. Vi demonsterer forbedringen af klassifikationsnøjagtighed ved anvendelse af tidsseriebilleder i stedet for enkeltepokeobservationer. I tillæg til dette viser vi en lignende rumlig-tidsmæssig ML pipeline til at bestemme Hubblekonstanten fra simulerede tidsseriebilleder af linsede supernovaer som forventet fra næste generation af observationer.

Til slut skifter den tredje del af afhandlingen fokus til galaksehobe, de mest massive gravitationelt bundne kosmiske strukturer, og deres potentielle værdi for kosmologien. Vi designer en deep learning galaksehobsmasseestimator som udnytter den fulde 3D-projekterede faserumsfordeling af galaksehobe. Som resultat giver masseestimator i øjeblikket de mest præcise skøn af massen blandt de nyligt foreslåede ML-baserede metoder. For første gang anvendes en sådan ML-baseret masseestimator til at konstruere en galaksehobsmassefunktion fra ægte galaksehobsobservationer. Dette giver resultater, der er konsistente med forudsigelser fra den kosmologiske standardmodel.

Acknowledgements

It has been a great experience and adventure to come to Copenhagen to pursue a PhD. In this part of the thesis, I would like to express my gratitude to those that helped me throughout my journey.

First and foremost, I would like to thank Radek for being such a fantastic supervisor. For always being available whenever I needed your help, and for encouraging me to follow my own research interests. I have been extremely lucky to have you as a supervisor, which has allowed me to enjoy doing research without any type of pressure. With your broad foundation in cosmology and your genuine passion for research, you are a true inspiration to me. If all scientists were like you, the Universe would be a better place.

Secondly, I would like to express my deepest gratitude towards Doogesh, for growing into the best mentor I could possibly have wished for. I cannot thank you enough for taking an interest in me and for your unwavering support with research projects, job applications, and with shaping this thesis into what it is today. You have always kept me going and made me believe in myself. I am constantly striving to follow in your footsteps in terms of machine learning knowledge, writing style, figure layout, work ethics, planning, life organisation, and so much more. I sincerely hope we will always continue to do research together.

I am very grateful to the members of my PhD assessment committee; Phil Marshall, Eleonora di Valentino, and Steen Hansen, for taking time out of their busy schedule to review my thesis. I feel honoured to have the opportunity to discuss my work with such experts in the fields of gravitational lensing, the upcoming LSST survey, non-standard cosmological models, and the Hubble tension.

There are several more people that contributed greatly to my scientific development. Adriano, thank you for transforming the first year of my PhD into a true success. I am intensely happy that I have been able to work with you on my first two projects and I have learned so much from you. Jens, your honesty and sharp questions shaped me into a better scientist. Additionally, I am grateful for having been supported by a VILLUM FONDEN Investigator grant to Jens (project number 16599). Charlotte, thank you for being so helpful, supportive and patient while teaching me everything about observations and data reduction. Marianne, I truly appreciate your support with the administrative side of finishing a PhD. Michelle and

Evi, thank you for making DARK a wonderful place and for keeping the plants alive! Troels, I feel honoured to have been part of the epic experience that is your Applied Statistics course. Benedikt, our meeting at the Alpine Cosmology Workshop was brief, but you have made a lasting impression on me in terms of research and presentation techniques, and life philosophy. Sofie, Cecilie and Jo, thank you for taking on the burdensome task of translating the Abstract of this thesis into Danish.

I am also deeply appreciative of the supportive atmosphere and healthy work-life balance at DARK, which allowed me to enjoy my PhD from start to finish. To all my colleagues at DARK and DAWN, thank you for the many nice journal clubs, cake talks, Christmas parties, good conversations, and evenings out. In particular, I would like to thank Cecilie for warmly welcoming me to the Danish student life and for being my personal stylist.

I am intensely grateful to have encountered such amazing friends in Copenhagen, who have made my time here a truly wonderful experience. To Albert, Anika, Iary, Cameron, Lisa, Markku, Martin, Mathias, Nicolas, Randi, Reggie, and Vera, I thoroughly enjoyed all our climbing sessions, parties, dinner evenings, and bike rides. A special thanks to Vera, for providing me with the *hyggeligste* workplace during the last week of writing. To Rasa, Michaela and Magnus, for being a great legacy of the pedagogy course and for always inspiring me to make more out of life. To Andrew and Dominika, you can climb and party like no one I have ever met! To Karoline and Max, for the best random encounter at winter bathing. And especially to Alex, I am so lucky to have found such a good friend as you and I hope that we will always stay in touch. I would also like to express my appreciation to the climbing halls *Beta Boulders* and *Nørrebro Klatreklub* for being the perfect places to recharge after a day of work.

A big thanks to my friends in the Netherlands for always supporting me and being understanding when I could not visit you as often as I wanted. You have been a great practice audience for my outreach presentations. And a shout-out to everyone that visited me here in Copenhagen (especially the ones that came during rainy February!).

Callum, thank you for your continuous belief in me and for being the visionary behind many of my presentations and works. For encouraging me, while at the same time having the courage to criticise me when it is necessary.

Last but not least, I would like to express my biggest gratitude to the best parents and sister in the world. Jonna, bedankt voor de wekelijkse 'teken'-sessies die altijd gevuld zijn met goede gesprekken en inzichten, en voor de geweldige tijden samen in Nederland en Kopenhagen. Ik ben mijn ouders heel erg dankbaar voor de eindeloze steun en veilige thuishaven die ze me altijd bieden, tot op de dag van vandaag. Pap, bedankt dat je me geleerd hebt om georganiseerd en productief te zijn, wat onmisbare vaardigheden zijn in het afronden van een PhD. Mam, bedankt voor je grenzeloze liefde, steun en toeverlaat.

Contents

Introduction

motivation

The Universe came into existence around 13.8 billion years ago in a hot, dense state, as postulated by the Big Bang model. Primordial density perturbations grew under the influence of gravity to form structures and a myriad of astrophysical objects that we observe around us today. The early Universe was dominated by radiation, followed by a phase of matter domination, both in the form of visible and dark matter. Today, we are dominated by dark energy, an unidentified substance that drives the accelerated expansion of the Universe.

Although our current knowledge about the history of the Universe is remarkable, open questions remain pertaining to its inner-workings, composition and evolution. What is the nature of dark matter and dark energy, and how does dark energy evolve over time? What is the current expansion rate of the Universe, i.e. the Hubble constant? Can the Universe be accurately described by our standard ΛCDM cosmological model or do we need new physics? What is the exact expansion history and geometry of the Universe?

Next-generation galaxy surveys, such as the Legacy Survey of Space and Time (LSST) to be conducted at the Vera C. Rubin Observatory, as well as upcoming missions at the Nancy Grace Roman Space Telescope, the Dark Energy Spectroscopic Instrument (DESI), and *Euclid* will collect an unprecedented amount of high-resolution data. Consequently, there will be many opportunities to make progress in deciphering the above cosmological puzzles. However, the sheer volume of data will also pose new challenges. Machine learning techniques have recently emerged as promising tools to optimally extract information from these large data sets.

The main focus of this thesis is to address the above questions by combining observational data with simulations and theoretical models, whilst using machine learning methods to link these three aspects. The underlying philosophy is that challenges in cosmology are most likely to be solved at the intersection of observations and theory. More specifically, the work presented in the thesis follows two distinct avenues. The first strategy infers the cosmic expansion history by connecting distances and redshifts from astrophysical objects, thereby employing existing data sets as well as forecasts for future observations. The second approach computes the mass distribution of galaxy clusters, the most massive gravitationally bound objects in the Universe, with the ultimate goal of constraining the matter density and clustering amplitude.

THESIS OUTLINE

The thesis is structured as follows. Chapter 1 provides an introduction to our present view of the formation and evolution of the Universe. It gives an overview of our standard ΛCDM model and describes how initial density perturbations are imprinted in the Cosmic Microwave Background (CMB) radiation and have evolved into stars, galaxy clusters and other astrophysical objects, some of which can be employed today as distance indicators. The remainder of the chapter reviews the tension that arises between different measurements of the cosmic expansion rate, and the prospects of resolving the tension via a new cosmological model.

Chapter 2 outlines the statistical inference and machine learning methods adopted in this thesis. It begins with a description of Bayesian statistics and Markov Chain Monte Carlo techniques, and proceeds with the presentation of two types of neural networks. Finally, it describes two methods that can provide reliable uncertainties for the neural network predictions.

Part I: The Hubble constant & sound horizon tension

In Chapter 3, we present a measurement of the Hubble constant and sound horizon from type Ia supernovae and Baryon Acoustic Oscillations, calibrated by gravitationally lensed quasars. The analysis is done in a cosmographic framework without adopting any assumptions about the underlying cosmology. As a consequence, our results are completely independent of a choice of cosmological model, CMB observations, and the Cepheid calibration. We find a weak ($\sim 2\sigma$) tension with predictions from CMB measurements and the standard ΛCDM model. Additionally, we investigate the effects of including a sample of quasars with standardisable ultraviolet and X-ray luminosity distances, which produce a slightly higher tension. This study also demonstrates the potential of constraining the cosmic curvature solely through low-redshift observations, yielding a result consistent with a flat Universe. The corresponding paper is published in *Astronomy & Astrophysics* (Arendse, Agnello, & Wojtak, 2019).

Chapter 4 revisits the tension, with additional distance measurements from two gravitationally lensed quasars and calibrations from Cepheids and stars at the Tip of the Red Giant Branch. The resulting measurements of the Hubble constant and the sound horizon are in strong (up to 5σ) tension with CMB observations. In order to address this discrepancy, we investigate whether modifications of the standard ΛCDM model can reconcile the tension. We show that *early-time* extensions, which alter the physics before recombination, slightly decrease the tension, but do not manage to dissipate it completely. Models that change the physics after recombination, i.e. *late-time* modifications, are often put forward as promising solutions to the Hubble tension. However, we demonstrate that they fail to address the combined tension because they are incapable of changing the value of the sound horizon. These findings, which are published in *Astronomy & Astrophysics* (Arendse, Wojtak, Agnello, Chen, Fassnacht, Sluse, Hilbert, Millon, Bonvin, Wong, Courbin, Suyu, Birrer, Treu, & Koopmans, 2020), are crucial to be taken into account when devising new models to resolve the tension, and may tentatively point in the direction of systematics as an explanation for the prevailing tension. For this reason, it is essential to explore independent avenues to measure the cosmic expansion rate, such as the one described in Part II of the thesis.

Part II: Gravitationally lensed supernovae

Strong gravitationally lensed supernovae constitute a promising alternative method to constrain cosmological parameters, such as the Hubble constant, especially in light of the expected colossal volumes of data from future transient surveys. However, distinguishing between lensed and unlensed (i.e. normal) supernovae in large quantities of survey data is a daunting task, particularly considering the lack of lensed supernovae observations that could serve as training data for a machine learning model. To aid in coping with this challenge, Chapter 5 presents a deep learning pipeline for the identification of gravitationally lensed supernovae in transient surveys such as the Young Supernova Experiment (YSE) and the Legacy Survey of Space and Time (LSST). Our framework uses recurrent convolutional layers to exploit both the spatial and time-variable features of multi-epoch observations, while drawing from recent advances in variational inference to quantify approximate Bayesian uncertainties via a confidence score. We test our pipeline on simulated YSE observations and report an improvement in classification accuracy of nearly 20 per cent when time-series images are used compared to single-epoch observations. Another important result is that our machinery is able to discriminate between lensed and unlensed supernovae with ~ 99 per cent accuracy for mock LSST observations. The corresponding paper is currently under review at *Nature Astronomy* (Kodi Ramanah, Arendse, & Wojtak, 2021).

Chapter 6 focuses on the use of gravitationally lensed supernovae for cosmological inference, specifically to measure the expansion rate of the Universe. We simulate time-series images to investigate the constraining power of pure LSST data, without any follow-up observations. A spatio-temporal convolutional neural network is employed to convert the input data into estimates of the lens parameters and time delays, which are subsequently used to calculate the time-delay distance and the Hubble constant. In order to quantify reliable uncertainties for each neural network prediction, we adopt a simulation-based inference framework. By including realistic predictions for the lensed type Ia supernovae rates, we forecast to find 400 objects during the 10 year duration of LSST that are suitable for cosmological inference, yielding a joint 1.2% unbiased estimate of the Hubble constant. We find that doubly imaged supernovae account for the majority of the constraining power, whereby the dominant source of uncertainty is the source position and time delays between the lensed images. This proof of concept encourages the use of pure LSST data in a joint population analysis, and provides a framework to quantify the dominant sources of uncertainty on the cosmic expansion rate from lensed supernovae.

Part III: Galaxy cluster masses

Chapter 7 shifts the focus to dynamical measurements of galaxy cluster masses, and their potential for cosmological inference. The cluster mass function is a powerful probe of the matter density and amplitude of mass fluctuations, although obtaining accurate dynamical cluster mass estimates is a challenging task, largely due to the presence of interloper (non-member) galaxies in galaxy clusters. We address this problem by employing the full 3D phase-space distribution of the projected galaxy positions and their line-of-sight velocities, thereby providing a better separation between cluster members and interlopers. A 3D convolutional neural network is used to optimally exploit the input data, combined with a simulation-based inference framework to derive cluster mass uncertainties. We generate a realistic mock catalogue that closely emulates the Sloan Digital Sky Survey (SDSS) Legacy observations and we illustrate explicitly the challenges posed by interloper galaxies. Moreover, we apply our framework to a set of SDSS clusters, which constitutes the first time that a machine learning-based mass estimator is applied to such an extensive set of real galaxy cluster observations. The resulting cluster mass function is fully consistent with mass estimates from the literature and with the theoretical halo mass function as predicted for the standard cosmological model. The corresponding paper is published in *Monthly Notices of the Royal Astronomical Society* (Kodi Ramanah, Wojtak, & Arendse, 2021).

Finally, Chapter 8 summarises the main findings of the work presented in the thesis and proposes some avenues for future studies.

CHAPTER 1

Astrophysical background

"The history of astronomy is a history of receding horizons."
- Edwin Powell Hubble

As time progresses, so does humanity's ability to look further into the Universe and further back in time. This has allowed us to establish a compelling picture of the Universe and the astrophysical objects within it. The elements of this picture that are relevant to the doctoral work are outlined in this chapter. After an overview of cosmology (Section 1.1), we proceed with a description of the Cosmic Microwave Background radiation (Section 1.2) and the formation of galaxy clusters (Section 1.3). Finally, several low-redshift distance indicators are discussed (Section 1.4), as well as the tension between different data sets in terms of the cosmic expansion rate (Section 1.5).

1.1 COSMOLOGY

The essence of cosmology, as a field of scientific research, entails pushing the boundaries of our understanding of the large-scale structure and evolution of the Universe. In order to do so, we require knowledge of the constituents that make up the Universe and a description of how they evolve as the Universe expands. When combined with a theory of gravity, this yields our current cosmological standard model. This section presents the underlying principles of this model, while drawing inspiration from Arendse (2018), Fumagalli (2018), Gariazzo (2016), Harmark (2021), Kodi Ramanah (2019), and Lee (2017).

1.1.1 Einstein's theory of general relativity

The discipline of cosmology can be traced back to the beginning of the 20th century, when Albert Einstein presented a completely new interpretation of space, time and gravity.

Einstein was inspired by the work of some of the great physicists before him, like the Scottish scientist James Maxwell, whose equations elegantly described how electric and magnetic fields are generated from charges (Maxwell, 1865). From the fact that Maxwell's equations are invariant with respect to the choice of inertial reference frame, it followed that the speed of electromagnetic waves is constant for all observers. Einstein decided to embrace the notion of a constant speed of light to see where this line of reasoning would take him. The second important postulate that he adopted for his work was the *principle of relativity*: the laws of physics are the same in every inertial reference frame. From these two assumptions, Einstein's theory of special relativity was born in 1905 (Einstein, 1905b). As his theory introduces the speed of light as an absolute quantity that all observers agree upon, the concept of time shifted from its previously absolute interpretation to something relative. Consequently, moving clocks tick more slowly than stationary ones. Another outcome of the theory was that space and time could not be seen as different entities anymore; they formed the universal fabric called *spacetime*. Similarly, he came up with the mass-energy equivalence, stating that mass can be converted into energy as expressed by his famous formula $E = mc^2$ (Einstein, 1905a).

In the ten years that followed, Einstein worked on unifying his theory of special relativity with a theory of gravity. Gravity works instantly in Newtonian physics, which is in contradiction with Einstein's universal speed limit. The first step towards this unification was the formulation of the *equivalence principle*, which states that gravitational and inertial masses are indistinguishable. His resulting theory of general relativity explains gravity as an emergent property of spacetime when it is curved by mass, energy or momentum. *"Spacetime tells matter how to move; matter tells spacetime how to curve."* (Wheeler & Ford, 1998). Not only is matter affected by gravity; light rays are deflected and clocks run more slowly in a stronger gravitational field.

The geometrical properties of spacetime can be described via the *metric tensor* $g_{\mu\nu}$, which allows us to calculate the distance $\mathrm{d}s$ between two points separated by $\mathrm{d}x_\mu$:

$$\mathrm{d}s^2 \equiv g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{1.1}$$

Here, repeated indices are summed over following the Einstein summation convention. The metric reduces to $\eta_{\mu\nu} = \mathrm{diag}(-1,+1,+1,+1)$ in flat Minkowski space. The degree to which the metric tensor deviates from flatness is characterised by the *Ricci tensor* $R_{\mu\nu}$:

$$R_{\mu\nu} \equiv \partial_\alpha \Gamma^\alpha{}_{\mu\nu} - \partial_\nu \Gamma^\alpha{}_{\mu\alpha} + \Gamma^\alpha{}_{\beta\alpha}\Gamma^\beta{}_{\mu\nu} - \Gamma^\alpha{}_{\beta\nu}\Gamma^\beta{}_{\mu\alpha}, \tag{1.2}$$

$$\Gamma^\rho{}_{\mu\nu} \equiv \tfrac{1}{2} g^{\rho\sigma}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}\right), \tag{1.3}$$

where $\Gamma^\rho{}_{\mu\nu}$ is called the *Christoffel symbol*. The metric tensor with upper indices $g^{\mu\nu}$ corresponds to the inverse of $g_{\mu\nu}$ and $\partial_\mu = \partial/\partial x^\mu$ denotes the partial derivative. The trace of the Ricci tensor is the *Ricci scalar*: $\mathcal{R} \equiv R^\mu_\mu = g^{\mu\nu}R_{\mu\nu}$, a scalar quantity that represents the curvature of spacetime and is invariant under any coordinate transformations.

The theory of general relativity proposes that the curvature of spacetime, as described by $g_{\mu\nu}$, $R_{\mu\nu}$ and $\mathcal{R}$, depends on the matter and energy content of the Universe. The latter is given by the *energy-momentum tensor* $T_{\mu\nu}$, which is a physical quantity that characterises a continuous configuration of matter and energy, thus generalising the mass density in Newtonian physics. Einstein constructed a set of ten equations, known as the Einstein field equations, that quantify the curvature of spacetime due to the matter-energy content of the Universe (Einstein, 1915), as follows:

$$R_{\mu\nu} - \tfrac{1}{2}\mathcal{R}\, g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}\,, \tag{1.4}$$

with G the gravitational constant and c the speed of light in vacuum. Λ is the *cosmological constant*, introduced initially by Einstein to balance the effects of the matter-energy density and create a static universe. Nowadays, it is used to describe dark energy, as discussed in more detail in Section 1.1.4. From the definition of the Riemann curvature tensor, which contains a quadratic term of $g_{\mu\nu}$, one can see that the Einstein equations are non-linear. This is one of the crucial differences with Newtonian gravity and it can be interpreted as a self-interaction of the gravitational field. Einstein's new formalism of gravity can explain previously inexplicable phenomena, such as the precession of Mercury, and has correctly predicted the existence of gravitational lensing, gravitational redshift and gravitational waves.

1.1.2 The expanding Universe

Einstein, as many others at the time, believed the Universe to be static and eternal. This assumption was entirely shattered in 1929, when Edwin Hubble published his paper entitled *"A relation between distance and radial velocity among extra-galactic nebulae"* (Hubble, 1929), an idea that two years earlier was also proposed by George Lemaître (Lemaître, 1931). Their work considered several nearby galaxies and measured their distances using Cepheid variable stars and their velocities via the Doppler shift of spectral lines. Some nearby galaxies, like the Andromeda galaxy, were moving towards us, but the vast majority of galaxies were heading away from us. Figure 1.1 shows Hubble's famous plot of the recession velocity as a function of the distance to the galaxy, in which it is evident that galaxies located further away from us are moving away from us at higher velocities. The slope of the graph gives the proportionality constant between the velocity and the distance, also known as the *Hubble constant* H_0. The Hubble-Lemaître law is then given by

$$v = H_0 D, \tag{1.5}$$

where v is the recession velocity in km/s, D is the distance to the galaxy in Mpc and H_0 is the Hubble constant, or present-day expansion rate of the Universe. The discovery of the Hubble-Lemaître law led to a true paradigm shift, since these observations could be explained perfectly by an expanding Universe. Furthermore, it led to the notion that the Universe is not eternal but must have had a beginning; the *Big Bang*.

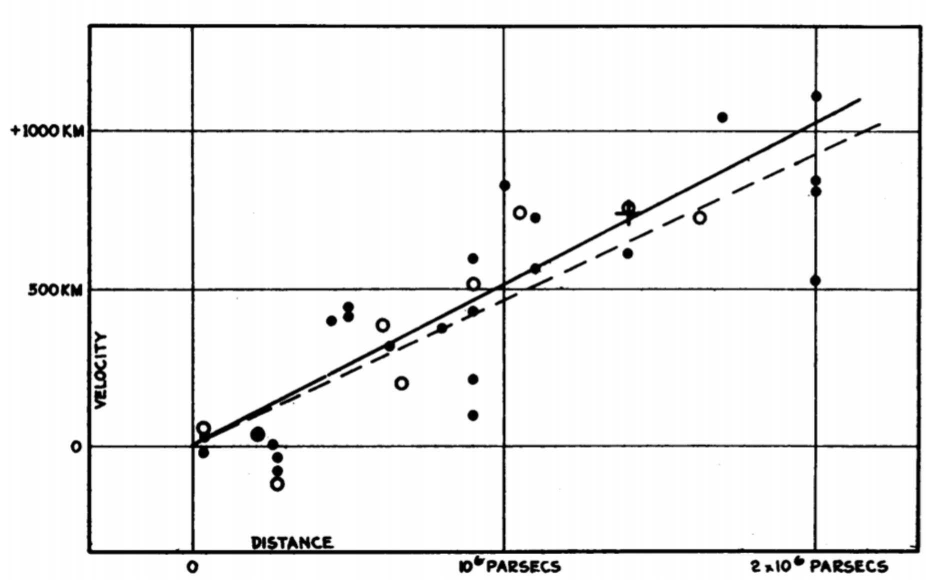


Figure 1.1 – The original plot of the velocity-distance relation among extra-galactic nebulae (galaxies). The black dots and the black solid line correspond to individual galaxies and their average slope, the Hubble constant. The open circles and the dashed line correspond to grouped galaxies and their corresponding Hubble constant. Figure reprinted from (Hubble, 1929).

In order to describe such an expanding Universe, it is convenient to introduce a coordinate system that moves along with the expansion; *comoving coordinates*. The scale factor $a(t)$ describes the relative size of the Universe and relates the comoving coordinates $\boldsymbol{x}$ to the physical coordinates $\boldsymbol{r}$ according to

$$\boldsymbol{r} = \boldsymbol{x}\, a(t)\,. \tag{1.6}$$

By definition, the scale factor is normalised to unity at present time, $a_0 = a(t_0) = 1$. The time evolution of the scale factor can be used to describe the cosmic expansion history, characterised by the *Hubble parameter* $H(t)$:

$$H(t) = \frac{\dot{a}(t)}{a(t)}\,, \tag{1.7}$$

where $\dot{a} = \mathrm{d}a/\mathrm{d}t$ is the time derivative of the scale factor. The Hubble parameter evaluated at the present time corresponds to the Hubble constant, $H_0 = H(t_0)$, which is often expressed in a dimensionless form as $h = H_0$ / (100 km s^{-1}Mpc^{-1}). The exact value of H_0 is still an open question in cosmology, as discussed further in Section 1.5.

1.1.3 The Friedmann equations

A fundamental assumption that allows us to solve Einstein's equations for the entire Universe is the *cosmological principle*; which states that there is nothing special about our place in the

Universe and that the matter distribution at sufficiently large scales is spatially homogeneous (invariant under translations) and isotropic (invariant under rotations). Since we observe every part of the Universe around us to be approximately the same (isotropic) and because we assume that we are not privileged observers, this implies isotropy around every point in the Universe; therefore, the Universe must also be homogeneous (Peacock, 1998). Strong evidence for the cosmological principle comes from the Cosmic Microwave Background radiation, which is discussed in Section 1.2.

The scientists Alexander Friedmann, Georges Lemaître, Howard P. Robertson and Arthur Geoffrey Walker independently applied the cosmological principle to the geometry of spacetime and constructed the maximally symmetric FLRW metric (Friedmann, 1922, 1924; Lemaître, 1927, 1931, 1933; Robertson, 1935, 1936a, 1936b; Walker, 1937), which in spherical comoving coordinates (t, r, θ, ϕ) is given by:

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2\theta \, d\phi^2) \right], \tag{1.8}$$

where k refers to the global spatial geometry or *curvature* of the Universe and can assume three values:

- For $k = 1$, the spatial part of the FLRW-metric describes a three-dimensional (3D) sphere $\mathbb{S}^3$ with radius a. This corresponds to a finite, *closed* Universe, where the sum of the angles of a triangle exceeds 180 degrees and parallel lines eventually converge. In a closed Universe, the energy-matter content is large enough to overcome the expansion in a finite time, which will lead the Universe to recollapse.
- For $k = 0$, the spatial part of the metric reduces to 3D Euclidean space $\mathbb{R}^3$ and the geometry of the Universe is *flat*. In such a Universe, the energy-matter content is *exactly* enough to stop the expansion of the universe, but only after an infinite amount of time. The rate of expansion will be slowing down and will asymptotically approach zero. Observational data, such as the Cosmic Microwave Background radiation and the results from gravitationally lensed quasars and type Ia supernovae as presented in Chapters 3 and 4, indicate that we live in a flat Universe.
- For $k = -1$, the spatial part of the metric defines a 3D hyperboloid $\mathbb{H}^3$, corresponding to an *open* Universe in which the angles of a triangle add up to less than 180 degrees and parallel lines diverge. In this case, there is not sufficient energy-matter content to halt the initial expansion.

Friedmann subsequently used the FLRW-metric (eq. (1.8)) to find a solution to Einstein's equations of general relativity (eq. (1.4)). He assumed that at larger scales, the Universe can be approximated as a perfect, isotropic and homogeneous fluid described by energy-density ρ and pressure P, thereby reducing the energy-momentum tensor to $T_{\mu\nu} = \mathrm{diag}(\rho, P, P, P)$. With these conditions, there are two independent solutions to Einstein's field equations, known as the Friedmann equations:

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{1.9}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}\,. \tag{1.10}$$

By combining the two Friedmann equations, or by adopting conservation of the energy-momentum tensor $\nabla^\mu T_{\mu\nu} = 0$, the *continuity* equation can be derived:

$$\dot{\rho} + 3H(\rho + P) = 0\,. \tag{1.11}$$

This equation tells us that the Universe expands adiabatically, i.e. with conservation of energy. Most of the fluids in the Universe can be described by a simple *equation of state* of the form $P = w\rho$. Adopting this relation, the continuity equation can be solved to yield the the cosmological evolution of the fluid:

$$\rho \propto a^{-3(1+w)}. \tag{1.12}$$

The three main substances that make up our Universe are radiation, matter and dark energy, whose properties and cosmic evolution are described in the following subsection.

1.1.4 Constituents of the Universe

- **Matter**. Stars, planets and nearly all matter that we experience in everyday life is composed of baryons and is therefore called *baryonic matter*. However, the majority of matter is non-baryonic *dark matter*, which provides an extra mass density that accounts for the rotation curve of galaxies, the growth of cosmic structures and various other astrophysical phenomena. Both types of matter have a negligible pressure compared to their energy density and thus have equation of state parameter $w = 0$, such that the matter density evolves with a^{-3}. Intuitively, this can be understood as the dilution of matter due to space expanding in three dimensions.
- **Radiation**. For radiation, which consists of relativistic massless particles, the equation of state parameter is equal to $\frac{1}{3}$, resulting in a density evolution of $\rho \propto a^{-4}$. Radiation decays by a factor a faster than matter, because the expansion of space does not only cause light to dilute away in three dimensions, it also stretches its wavelength, thereby making it less energetic. Although radiation was the dominant component in the first $\sim 50,000$ years after the Big Bang, the current radiation contribution to the total energy budget is negligible due to its rapid decrease in energy density.
- **Dark Energy**. When Hubble and Lemaître discovered the expansion of the Universe, initially it was assumed that the gravitational effects of matter must be slowing down the expansion. It would take another 70 years to disprove this assumption. In 1998, two research teams used type Ia supernovae to show that the expansion of the Universe is accelerating (Perlmutter et al., 1998; Riess et al., 1998; Schmidt et al., 1998). This strange discovery led to the term *dark energy* being coined; an unidentified component with

negative pressure that drives the accelerated expansion of the Universe. At the present time, dark energy is estimated to constitute 69% of the Universe and its energy density appears to remain constant over time (Planck Collaboration, Aghanim, et al., 2018), in which case it is referred to as a *cosmological constant*. The cosmological constant used in Einstein's equations, initially introduced to create a static Universe, was assigned the new function to account for the accelerating cosmic expansion. Dark energy has equation of state parameter $w < -\frac{1}{3}$ and if it corresponds to a cosmological constant, $w = -1$. Since the radiation and matter densities are decreasing, dark energy is currently the dominant component in the Universe and its relative contribution will continue to increase.

For each of the above species, we can express their density in terms of the *critical density* ρ_c, which is the average density the Universe should have in order to be flat, and so defines the critical point between an expanding and contracting Universe. The dimensionless density parameter Ω_i for each different species i is defined as the ratio between the absolute energy density ρ and the critical density ρ_c:

$$\Omega_i \equiv \frac{\rho_i}{\rho_\mathrm{c}} = \frac{8\pi G \rho_i}{3H^2}, \tag{1.13}$$

where $i = \mathrm{m}, \mathrm{r}, \Lambda, \mathrm{k}$ for matter, radiation, cosmological constant and curvature, respectively. The density parameters satisfy $\Omega_\mathrm{r} + \Omega_\mathrm{m} + \Omega_\Lambda + \Omega_\mathrm{k} = 1$. In terms of these parameters, we can rephrase the Friedmann equation (eq. (1.9)) in the form

$$H^2(a) = H_0^2 \left(\frac{\Omega_{\mathrm{r},0}}{a^4} + \frac{\Omega_{\mathrm{m},0}}{a^3} + \frac{\Omega_{\mathrm{k},0}}{a^2} + \Omega_{\Lambda,0} \right), \tag{1.14}$$

where $\Omega_{i,0}$ denotes the density parameter evaluated at the present time. The concepts reviewed so far form the basis of the current standard model in cosmology, the Λ*CDM model*. It assumes a flat Universe with general relativity as a description of gravity, the cosmological constant as dark energy to explain the accelerating expansion of the Universe, and cold (i.e. non-relativistic) dark matter to account for the growth of cosmic structures. Within the ΛCDM model, the expansion history of the Universe is elegantly and simply described by eq. (1.14) with $\Omega_{\mathrm{k},0} = 0$, which has been remarkably successful at explaining a series of astrophysical and cosmological observations.

1.1.5 Cosmological distances and redshift measurements

Another remarkable consequence of the expansion of the Universe is *cosmological redshift*; the stretching of light as it travels through expanding space. Consequently, photons emitted at a wavelength λ_e at time t_e will be observed with redshifted wavelength λ_o at time t_o. The fractional change in wavelength is related to the scale factor and cosmological redshift z as:

$$\frac{\lambda_\mathrm{o}}{\lambda_\mathrm{e}} = \frac{a(t_\mathrm{o})}{a(t_\mathrm{e})} = 1 + z. \tag{1.15}$$

Usually, the observer is located at Earth at the present time and therefore, the scale factor $a(t_\mathrm{o})$ equals 1, providing the convenient relation of $1 + z = 1/a$ between the redshift and scale factor at emission. An additional effect that can induce a Doppler shift in the observed wavelengths of light is due to relative velocities between the observer and the source. By observing the shift in the spectral lines of distant astronomical objects, we can compute their redshifts, which include both the effects of relative velocities and the cosmic expansion.

The cosmological redshift of an object provides an indication of its distance from us. Several useful distance measures are defined below, all under the assumption of the cosmological principle leading to the FLRW metric.

- The **comoving distance d_C** can be calculated from the redshift in the following way:

$$d_C(z) = c \int_0^z \frac{dz'}{H(z')}. \tag{1.16}$$

  Within the ΛCDM framework, $H(z)$ is given by eq. (1.14).

- Correcting for the curvature of the Universe yields the **transverse comoving distance d_M**:

$$d_M(z) = \begin{cases} \frac{d_H}{\sqrt{\Omega_k}} \sinh\left(\frac{\sqrt{\Omega_k} d_C(z)}{d_H}\right) & \Omega_k > 0 \\ d_C(z) & \Omega_k = 0 \\ \frac{d_H}{\sqrt{|\Omega_k|}} \sin\left(\frac{\sqrt{|\Omega_k|} d_C(z)}{d_H}\right) & \Omega_k < 0 \end{cases} \tag{1.17}$$

$$d_H = \frac{c}{H_0}, \tag{1.18}$$

  where d_H is the *Hubble horizon*, beyond which objects move away from us faster than the speed of light. The trigonometric function 'sinh' accounts for hyperbolic curvature of space in an open Universe ($\Omega_k > 0$) and 'sin' for spherical curvature in a closed Universe ($\Omega_k < 0$). A flat Universe corresponds to $\Omega_k = 0$.

- The **angular diameter distance d_A** stems from the notion that objects appear smaller when they are further away, and can be obtained by comparing the angular size $\delta\theta$ of an object to its physical size R:

$$d_A = \frac{R}{\delta\theta}. \tag{1.19}$$

  The angular diameter distance is related to the transverse comoving distance by:

$$d_A(z) = \frac{d_M(z)}{1+z}. \tag{1.20}$$

  In astrophysical distance measurements, the angular diameter distance can be determined when the physical size of an object is known, such as for the Baryon Acoustic Oscillations as discussed in Section 1.4.2.

- Finally, the **luminosity distance d_L** is based on the fact that objects appear fainter as they are further away. The observed flux F of a source is inversely proportional to the square of its distance:

$$F = \frac{L}{4\pi d_{\mathrm{L}}^2}, \tag{1.21}$$

where L is the absolute luminosity of the source and d_{L} is the luminosity distance, which is related to the transverse comoving distance by:

$$d_{\mathrm{L}}(z) = (1+z)\, d_{\mathrm{M}}(z). \tag{1.22}$$

Measuring the luminosity distance to an object requires knowledge of its absolute luminosity. A particular class of astrophysical sources, called *standard candles*, have a known intrinsic brightness and therefore, allow for such a measurement. Examples of this class of astrophysical sources, including Cepheid variable stars and type Ia supernovae, are reviewed in Section 1.4.

1.1.6 Cosmography

Although the ΛCDM model is highly successful at describing observations, controversies remain regarding the exact value of the Hubble constant and the nature of dark energy and dark matter. In order to address these problems, several extensions to the standard ΛCDM model have been put forward, as discussed in more detail in Section 1.5. An alternative to assuming a certain underlying cosmological model is *cosmography*; a mathematical framework that describes the Universe in a purely observationally driven way. Introduced by Weinberg (1972) and extended by Visser (2004), cosmography abandons all assumptions about general relativity and the constituents of the Universe and relies only on the cosmological principle. By doing so, it should be able to encompass ΛCDM and all its extensions.

Cosmography aims to reconstruct the expansion history of the Universe, while only having access to the current value of the scale factor and its derivatives. We have already encountered the first derivative of the scale factor in eq. (1.7), which describes the cosmic expansion rate and is expressed as the Hubble parameter. In mechanics, the first four time derivatives of position are often referred to as velocity, acceleration, jerk, and snap. Analogously, we can define the first four time derivatives of the scale factor as the Hubble $H(t)$, deceleration $q(t)$, jerk $j(t)$, and snap $s(t)$ parameters:

$$H(t) = +\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t} \tag{1.23}$$

$$q(t) = -\frac{1}{a}\frac{\mathrm{d}^2 a}{\mathrm{d}t^2}\left[\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right]^{-2} \tag{1.24}$$

$$j(t) = +\frac{1}{a}\frac{\mathrm{d}^3 a}{\mathrm{d}t^3}\left[\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right]^{-3} \tag{1.25}$$

$$s(t) = -\frac{1}{a}\frac{\mathrm{d}^4 a}{\mathrm{d}t^4}\left[\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}t}\right]^{-4}, \tag{1.26}$$

where H_0, q_0, j_0 and s_0 denote the cosmographic parameters evaluated at the current time. Since the expansion of the Universe is accelerating, the present-day deceleration parameter q_0 actually takes on a negative value. Within the ΛCDM model, $q_0 = -0.55$. There are many different approaches of how cosmography can be carried out. One commonly used strategy is to combine the above derivatives of the scale factor with a Taylor expansion around $t = t_0$ to construct the evolution of the scale factor:

$$a(t) = a_0 + \left.\frac{\mathrm{d}a}{\mathrm{d}t}\right|_0 (t-t_0) + \frac{1}{2}\left.\frac{\mathrm{d}^2 a}{\mathrm{d}t^2}\right|_0 (t-t_0)^2 + \frac{1}{3!}\left.\frac{\mathrm{d}^3 a}{\mathrm{d}t^3}\right|_0 (t-t_0)^3 + \mathcal{O}(t-t_0)^4 \tag{1.27}$$

$$= a_0 \left\{ 1 + H_0(t-t_0) - \frac{1}{2} q_0 H_0^2 (t-t_0)^2 + \frac{1}{3!} j_0 H_0^3 (t-t_0)^3 + \mathcal{O}(t-t_0)^4 \right\}. \tag{1.28}$$

Similarly, the Hubble parameter as function of redshift can be rewritten in terms of the cosmographic parameters (Chiba & Nakamura, 1998):

$$H(z) = H_0 + \left.\frac{\mathrm{d}H}{\mathrm{d}z}\right|_0 z + \frac{1}{2}\left.\frac{\mathrm{d}^2 H}{\mathrm{d}z^2}\right|_0 z^2 + \frac{1}{3!}\left.\frac{\mathrm{d}^3 H}{\mathrm{d}z^3}\right|_0 z^3 + \mathcal{O}(z^4) \tag{1.29}$$

$$= H_0 \left\{ 1 + (1+q_0)z + (j_0 - q_0^2)\frac{z^2}{2} + \left[3q_0^3 + 3q_0^2 - j_0(3+4q_0) - s_0\right]\frac{z^3}{6} + \mathcal{O}(z^4) \right\}. \tag{1.30}$$

After several Taylor expansions and reversions of the power series, the luminosity distance can be written as a function of redshift in the following way (Visser, 2004; Weinberg, 1972):

$$d_L(z) = \frac{cz}{H_0} \left\{ 1 + [1-q_0]\frac{z}{2} - \left[1 - q_0 - 3q_0^2 + j_0 + \frac{kc^2}{H_0^2 a_0^2}\right]\frac{z^2}{6} + \left[2 - 2q_0 \right.\right.$$
$$\left.\left. -15q_0^2 - 15q_0^3 + 5j_0 + 10q_0 j_0 + s_0 + \frac{2kc^2(1+3q_0)}{H_0^2 a_0^2}\right]\frac{z^3}{24} + \mathcal{O}(z^4) \right\}. \tag{1.31}$$

Using eqs. (1.22) and (1.20), this can be converted to angular diameter distances and transverse comoving distances.

There are several challenges surrounding the use of cosmography as described above, i.e. through polynomial parametrisations. Firstly, the order of truncation of the Taylor expansions affects the resulting cosmographic parameters. Higher orders of expansions can better approximate the full extent of the data, but also introduce more free parameters and therefore larger uncertainties. Furthermore, the fact that the expansion is carried out around $z = 0$ means that it is ill-suited to describe high-redshift data ($z \gtrapprox 1$), as they are far removed from the interval of convergence of the Taylor expansion (Busti et al., 2015; Capozziello et al., 2020). The introduction of auxiliary variables, which expand the convergence radius of the Taylor expansion, can partly alleviate this problem. The latter approach is further described in Chapters 3 and 4, where cosmographic expansions are used to constrain cosmological parameters using lensed quasars, type Ia supernovae and Baryon Acoustic Oscillations in a cosmology-independent manner. Convergence tests, as performed in Appendix A.2, are required to find the optimal polynomial truncation order and to ensure that the chosen parametrisations do not introduce any biases.

1.2 THE COSMIC MICROWAVE BACKGROUND RADIATION

After the previous section's overview of the science of cosmology, this section continues by describing the physics of the early Universe and the information we can extract from light relics of the Big Bang. It draws inspiration from Hu (2001), Kodi Ramanah (2019), Lee (2017), Tojeiro (2006), and Wallisch (2018).

1.2.1 The early Universe

Although the cosmological principle provides us with essential tools to describe our Universe, we know that it cannot hold true perfectly because we observe cosmic structures around us. In order for gravity to be able to create these, there must have been some density perturbations in the early Universe. In our current best picture of the beginning of the Universe, it started out in an initial state of high density and temperature, known as the Hot Big Bang. Almost immediately after the Big Bang, the Universe underwent a period of rapid expansion known as *cosmic inflation*, during which it grew by a factor of 10^{26} in around 10^{-32} seconds (Albrecht & Steinhardt, 1982; Guth, 1981; Linde, 1982). In the first fraction of a second, the four fundamental forces, namely the electromagnetic force, the strong nuclear force, the weak nuclear force, and the gravitational force, were unified as one and started to decouple as the Universe expanded and cooled down. The period of exponential growth, as driven by cosmic inflation, is necessary to explain why the Universe is so flat, since any initial curvature was stretched to near flatness, and homogeneous, since the Universe originated from a small region in thermal equilibrium. It also provides a natural mechanism for creating primordial density perturbations that formed the seeds of cosmic structures. Microscopic quantum fluctuations, as predicted by Heisenberg's uncertainty principle, were magnified into cosmological perturbations.

The early Universe consisted mostly of hydrogen, some helium and small traces of lithium, which were all ionised due to the high temperatures and densities. Thomson scattering of the photons with free electrons caused the Universe to be opaque, trapping the photons and baryons in a primordial plasma. Any overdensity or gravitational potential well would first initiate gravitational collapse, until radiation pressure would build up and drive the baryons outwards again. These counteracting forces of gravity and pressure produced oscillations, similar to sound waves moving at a speed c_s of

$$c_s = \frac{c}{\sqrt{3(1+R)}}, \tag{1.32}$$

where $R = 3\rho_b/4\rho_\gamma$ is the ratio of the baryon density ρ_b to photon density ρ_γ. As illustrated in Figure 1.2, the sound waves oscillated between two states:

- compression in potential wells and rarefaction in the surrounding potential peaks;
- rarefaction in potential wells and compression in potential peaks.

The resulting fluctuations in the baryon density are known as the *Baryon Acoustic Oscillations*.

About 370,000 years after the Big Bang, at the *epoch of recombination*, the Universe had cooled down enough for the protons and electrons to combine into neutral hydrogen. At the *drag epoch* ($z = z_d$), the baryons and photons decoupled and the radiation pressure fell away.

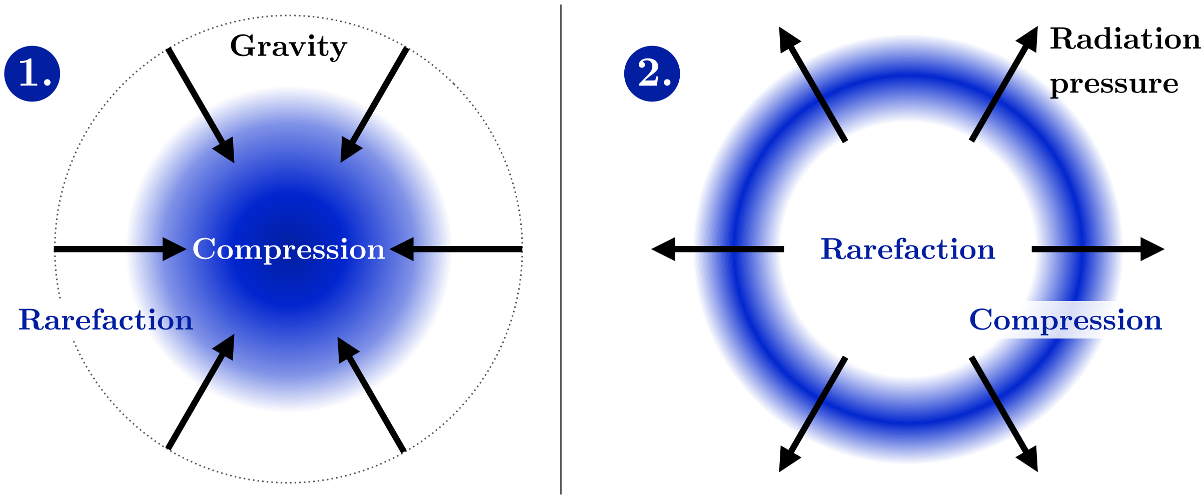


FIGURE 1.2 – A visualisation of the oscillations in the primordial photon-baryon plasma. The potential well in the centre of the figure oscillates between two states: *1.)* Gravity compresses the baryons in the centre, leading to a rarefaction in the surrounding potential peaks; *2.)* Radiation pressure drives the baryons outwards, creating a rarefaction in the central region and a compression in the surrounding potential peaks.

Oscillatory modes that reached extrema at this point in time (such as the ones visualised in Figure 1.2) contained a high density contrast and, therefore, enhanced temperature fluctuations. The longest mode, for which the oscillations had exactly enough time for one compression between the Big Bang and recombination, defines the characteristic distance scale r_d:

$$r_\mathrm{d} \equiv r_\mathrm{s}(z_\mathrm{d}) = \int_{z_\mathrm{d}}^{\infty} \frac{c_\mathrm{s} \mathrm{d}z}{H(z)} \,, \tag{1.33}$$

where r_d is the *sound horizon*; the maximum distance the sound waves could have travelled in the primordial plasma. During the oscillations of the photon-baryon plasma, dark matter remained in the potential well at the centre of the oscillations, since it only reacts gravitationally and is not affected by radiation pressure. Therefore, the structures that were created consisted of a large perturbation at the centre with a small perturbation in a spherical shell around it. After recombination, the gravitational interaction between the dark matter and the baryons caused them to redistribute, with some of the baryons returning to the centre of the overdensity and some of the dark matter ending up at the overdense shells. In this way, overdensities were formed that would eventually collapse to form stars and galaxies, with the sound horizon scale as a preferred separation between structures. Nowadays, we can observe the Baryon Acoustic Oscillations in the clustering of galaxies, as reviewed in Section 1.4.2.

1.2.2 Light relics from the Big Bang

Concurrently, the photons were completely decoupled from the baryons after recombination and could propagate predominantly freely through space. They have been travelling ever since, while the expansion of space redshifted their wavelengths to the microwave region. Today, we can observe these photons, constituting the afterglow of the Big Bang, as the *Cosmic Microwave Background radiation* (CMB). Detected in 1965 by Penzias and Wilson (1965a), the CMB provides the most compelling evidence for the Big Bang theory and cosmic inflation.

Alpher and Herman reasoned in 1948 that if the beginning of the Universe followed the Big Bang scenario, the leftover radiation would look like a black body spectrum of approximately five Kelvin (Alpher & Herman, 1948). In fact, the CMB spectrum can be almost perfectly described by a black body with temperature of $\bar{T} = 2.7255 \pm 0.0006$ K (Fixsen, 2009), and small perturbations of order 10^{-5}. Cosmic inflation theory predicts that these perturbations are adiabatic, i.e. that the fractional number density of each type of particle is the same everywhere. Observations of the CMB confirm this adiabatic nature of the perturbations. Additionally, since the fluctuations are so small in size, it also shows us that the early Universe was extremely homogeneous and isotropic.

The primordial density perturbations are imprinted on both the temperature and polarisation of the CMB radiation, but in this thesis, we only focus on the temperature fluctuations. These were first observed by the COsmic microwave Background Explorer (COBE) (Smoot et al., 1992) and later with more precision by the Wilkinson Microwave Anisotropy Probe (WMAP)

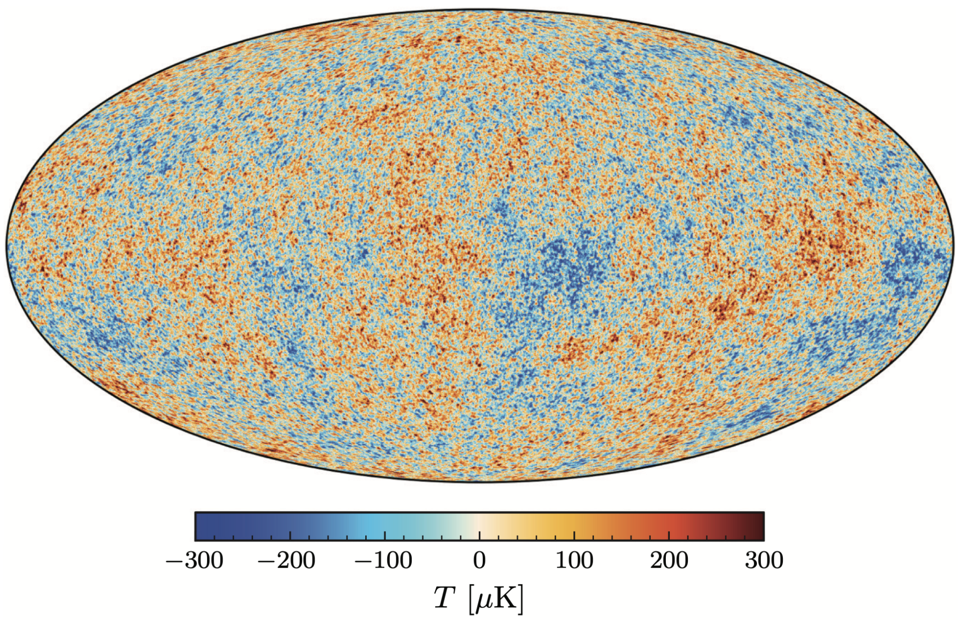


FIGURE 1.3 – The Cosmic Microwave Background radiation intensity map at 5 arcmin resolution as observed by the Planck satellite (Planck Collaboration, Adam, et al., 2016). A strip of the galactic plane is masked and filled in using the same statistical properties as the rest of the sky. The small temperature fluctuations correspond to over and under densities that are the seeds of cosmic structures.

(e.g. Dunkley et al., 2009) and Planck satellites (e.g. Planck Collaboration, Akrami, et al., 2018), as depicted in Figure 1.3.

1.2.3 CMB power spectrum

Besides confirming the Big Bang theory, cosmic inflation and the cosmological principle, the CMB contains another wealth of information when considering the angular size distribution of the temperature fluctuations. Light that we receive from the CMB originates from a celestial sphere at the edge of our observable Universe corresponding to the last photons to be scattered, and is, therefore, referred to as the *surface of last scattering*. In order to study the sizes of the fluctuations, it is convenient to use the spherical analogue of Fourier analysis, i.e. spherical harmonics. The temperature fluctuations ΔT can be expanded as a function of direction on the sky $\hat{\boldsymbol{n}}$ in terms of spherical harmonics $Y_{\ell m}(\hat{\boldsymbol{n}})$ as:

$$\Delta T(\hat{\boldsymbol{n}}) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\hat{\boldsymbol{n}}), \tag{1.34}$$

where m describes the orientation of the nodes on the sphere and ℓ is the multipole, which represents a given angular scale in the sky, with higher values of ℓ corresponding to smaller angular sizes. Under the assumption that the initial perturbations are statistically isotropic, such that there is no dependence on m, the variance of the harmonic coefficients $a_{\ell m}$ can be written as:

$$\langle a_{\ell m} a_{\ell' m'} \rangle = \delta_{\ell\ell'} \delta_{mm'} C_\ell, \tag{1.35}$$

where C_ℓ corresponds to the angular *power spectrum* of the temperature fluctuations, as shown in Figure 1.4 for the 2018 Planck data release. The most defining feature of the power spectrum are the acoustic peaks formed by the gravitational compression and radiation pressure felt by the primordial plasma. These peaks reveal an abundance of information about our Universe and can be perfectly described by the ΛCDM model with 6 free parameters: the baryon (Ω_b) and dark matter (Ω_dm) densities, the Hubble constant, the scalar spectral index n_s, the curvature fluctuation amplitude A_s, and the reionization optical depth τ.

The amplitude of the acoustic peaks scales with the physical density of dark matter. As shown in Figure 1.4, the second peak is much lower than the first. This can be explained by the presence of baryons, which enhance the compression in the potential wells (left panel of Figure 1.2) but rebounce to the same position in the potential peaks (right panel of Figure 1.2). The odd numbered (first, third, fifth, ...) peaks in the power spectrum are associated with the compression of the waves, so they are intensified by the presence of baryons. The even numbered (second, fourth, sixth, ...) peaks correspond to the rarefactions of the plasma, or in other words, how far it rebounds. Since the maximum rarefaction does not depend on mass, these peaks are relatively suppressed by a higher baryon content. As a result, the ratio of the first to second peak amplitude provides information about the baryon fraction of the Universe.

At smaller angular scales (higher ℓ), the acoustic peaks are exponentially damped by photons that travelled from dense to underdense regions and dragged protons and electrons along. This

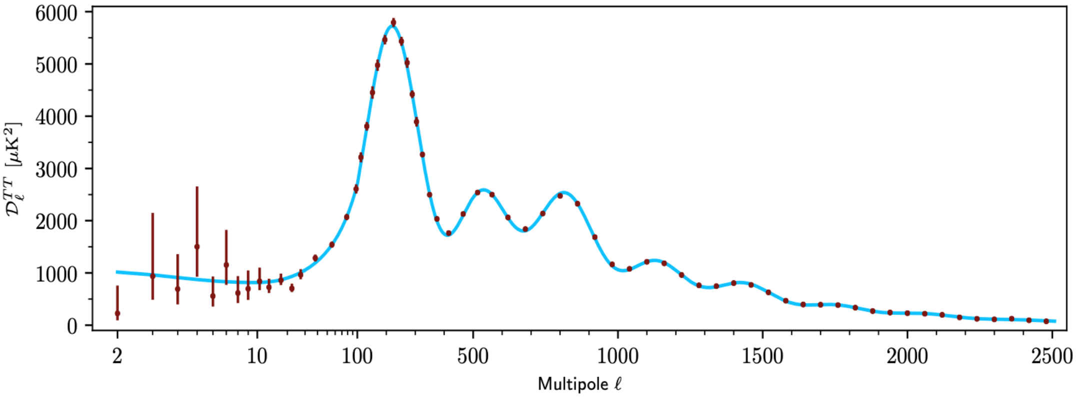


FIGURE 1.4 – Angular power spectrum of the CMB temperature fluctuations from the 2018 Planck mission (Planck Collaboration, Akrami, et al., 2018). The blue line is the best-fit theoretical ΛCDM spectrum, which is used to infer cosmological parameters.

process is called *photon diffusion damping* and makes the Universe more isotropic (Silk, 1968; Weinberg, 1972).

Because the first peak of the power spectrum corresponds to the first compression of the sound waves, it is a direct probe of the angular size of the sound horizon, θ_d. The physical size of the sound horizon is known from eq. (1.33) and hence, the angular diameter distance to the CMB can be determined using the identity $d_{CMB} = r_d/\theta_d$. Since the distance to the CMB depends on how light rays converge or diverge, it provides primarily a measure of the cosmic curvature. The resulting value of θ_d measured from the CMB is consistent with a flat Universe (Pierpaoli et al., 2000; D. D. Reid et al., 2002) and for that reason, a flat ΛCDM model is generally adopted for the CMB analysis. Within that choice of cosmology, the angular diameter distance is related to the Hubble constant according to eqs. (1.14) and (1.17) and, therefore, can be used to constrain its value. Recent analyses from Planck 2018 data find $r_d = 147.2 \pm 0.3$ Mpc and $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} (Planck Collaboration, Aghanim, et al., 2018) for the physical size of the sound horizon and the Hubble constant, respectively. As described in more detail in Section 1.5, these values are in tension with some of the low-redshift measurements.

1.3 THE FORMATION OF STRUCTURES AND GALAXY CLUSTERS

As the evolution of the Universe proceeded, the primordial density perturbations grew under the influence of gravity and formed structures and galaxy clusters. The abundance of these structures as a function of mass provides invaluable information about the amplitude of density fluctuations and the matter content of the Universe. This section covers the linear growth of perturbations into structures and galaxy clusters, and their potential for cosmological inference. It is inspired by material drawn from Bocquet (2015), Borgani (2008), Dayal and Ferrara (2018), Kodi Ramanah (2019), Maltoni and Maccio (2010), Sridhar (2016), Takey (2014), van de Weygaert (2009), and Wu (2011).

1.3.1 Linear perturbation growth

A primordial density perturbation,

$$\delta \equiv \frac{\rho - \bar{\rho}}{\bar{\rho}}, \tag{1.36}$$

where $\bar{\rho}$ denotes the average density of the Universe, is amplified by gravity according to linear perturbation theory when $|\delta(\boldsymbol{x})| << 1$. In this regime, which is a good approximation for the earlier stages of structure formation, the growth of density perturbations follows the three linearised fluid equations in Newtonian physics. The *continuity equation*, which ensures mass conservation, is given by:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot \boldsymbol{v} = 0, \tag{1.37}$$

where $\boldsymbol{v}$ denotes the peculiar velocity of a fluid element in comoving coordinates. As high density regions contract, a pressure gradient builds up that competes with the inward force of gravity. The *Euler equation* describes how the acceleration of the medium depends on the forces of gravity and pressure, as follows:

$$\frac{\partial \boldsymbol{v}}{\partial t} + \frac{\dot{a}}{a}\boldsymbol{v} = -\frac{\nabla \phi}{a} - \frac{c_s^2}{a}\nabla \delta, \tag{1.38}$$

with $c_s^2 = \partial P/\partial \rho$ equal to the fluid sound speed. The gravitational potential ϕ is defined by the *Poisson equation*:

$$\nabla^2 \phi = 4\pi G \bar{\rho} a^2 \delta. \tag{1.39}$$

Together, these three linearised fluid equations describe the evolution of density fluctuations with time, as a result of self-gravity. Combining them yields the time-evolution of a density contrast δ:

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta + \frac{c_s^2}{a^2}\nabla^2 \delta. \tag{1.40}$$

This equation describes how the gravitationally-driven perturbation growth (first term on the RHS) is opposed by the internal pressure gradient of the fluid (second term on the RHS), as well as the expansion of the Universe (second term on the LHS). The regime in which the pressure exactly balances gravity is called the *Jeans scale* λ_J, which defines a critical length that a cloud must exceed to undergo gravitational collapse (Jeans, 1902):

$$\lambda_J \equiv c_s \sqrt{\frac{\pi}{G\bar{\rho}}}. \tag{1.41}$$

Density fluctuations with length smaller than the Jeans scale have enough pressure support to keep them from collapsing, while structures whose size is larger than the Jeans scale are unstable and will collapse to form stars, galaxies and galaxy clusters. The Jeans scale is proportional to the Hubble horizon (eq. (1.18)) and increases over time, since the average density of the Universe decreases. As a result, the growth of the small perturbations that enter the Jeans scale during the radiation-dominated epoch is suppressed, which plays an important role in shaping the matter power spectrum, as discussed in more detail in Section 1.3.2

Eq. (1.40) is a second-order differential equation for the density perturbation δ, which can be similarly written in terms of a growing and a decaying mode acting on δ:

$$\delta(t) = D_{-}(t)\Delta_1 + D_{+}(t)\Delta_2, \tag{1.42}$$

where Δ_1 and Δ_2 represent primordial density contrasts, and D_{-} and D_{+} correspond to the decaying and growing modes, respectively. In a matter-dominated Universe, density fluctuations grow proportionally to the scale factor; $D_{+} \propto a$. When dark energy dominates the Universe, the growth of matter fluctuations is suppressed. Often, the derivative of the linear growing mode is taken with respect to the scale factor to yield the dimensionless logarithmic growth rate f:

$$f = \frac{\mathrm{d}\ln D_{+}}{\mathrm{d}\ln a}, \tag{1.43}$$

which is an important cosmological parameter that characterises the growth of structures and can function as a probe of dark energy. Additionally, it can aid in distinguishing between models of modified gravity, where f is scale-dependent, and general relativity, where f only depends on the matter density as $f = \Omega_{\mathrm{m}}^{0.55}$ (Linder, 2005). Redshift space distortions, as discussed in Section 1.4.2, can be employed to constrain the logarithmic growth rate.

1.3.2 The matter power spectrum

As we move from a single density perturbation to a density field $\delta(\boldsymbol{x})$, which provides the density for each point in space, we can establish a formalism for the statistical properties of large-scale cosmic structures. According to our current understanding of the beginning of the Universe, the primordial density fluctuations comprise a Gaussian random field, whose characteristics are completely captured by the two-point correlation function:

$$\xi(|\boldsymbol{x_1} - \boldsymbol{x_2}|) = \langle \delta(\boldsymbol{x_1})\delta(\boldsymbol{x_2})\rangle, \tag{1.44}$$

which provides the excess probability of finding two structures separated by a distance $|\boldsymbol{x_1} - \boldsymbol{x_2}|$. The Fourier transform of the density field describes the density contrast as a function of scale:

$$\hat{\delta}(\boldsymbol{k}) = \frac{1}{(2\pi)^{3/2}} \int \delta(\boldsymbol{x}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \mathrm{d}\boldsymbol{x}, \tag{1.45}$$

where $\boldsymbol{k}$ corresponds to the 3D wavevector with amplitude k. Consequently, the density perturbations in Fourier space are fully described by the Fourier transform of the two-point correlation function, the so-called linear *matter power spectrum* $P(k)$:

$$P(k) = \langle |\delta_{\boldsymbol{k}}|^2 \rangle. \tag{1.46}$$

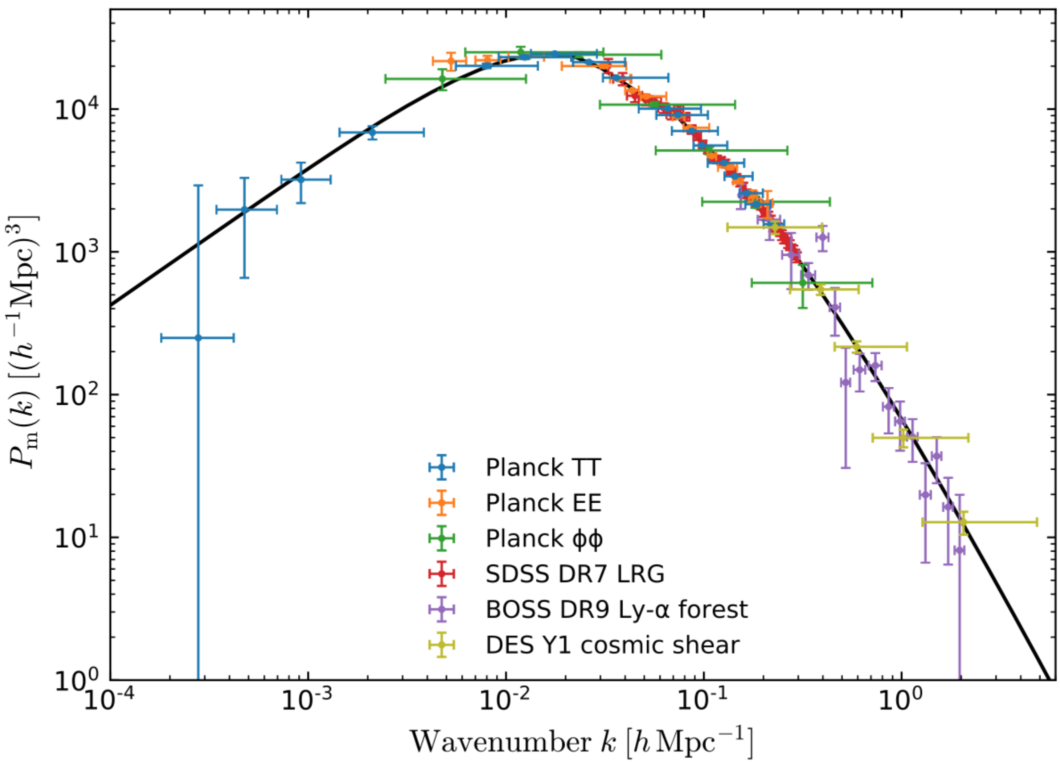


FIGURE 1.5 – The linear matter power spectrum (at $z = 0$) computed from different cosmological probes, with the black line corresponding to the best-fit ΛCDM model. Figure reprinted from Planck Collaboration, Akrami, et al. (2018).

Figure 1.5 displays the linear matter power spectrum inferred from different cosmological probes. The features in the power spectrum reflect the evolution of density perturbations in the early Universe (Norman, 2010). Due to the rapid expansion caused by inflation, primordial density fluctuations were driven beyond the Hubble horizon (eq. (1.18)), where they only experienced the effects of gravity and not of radiation pressure. In this state, the perturbations grew continuously and were characterised by a scale-invariant power spectrum, i.e. there was no preferred mass scale for the fluctuations. However, as the expansion rate of the Universe slowed down after inflation, the Hubble horizon expanded to encompass increasingly larger perturbation modes. Modes that crossed the Hubble horizon during the radiation-dominated epoch ($z < 3600$) predominantly experienced acoustic oscillations driven by the radiation pressure that halted their growth. Only when matter started to dominate the cosmic energy budget did these modes continue to increase in amplitude. As a result, the power of small-scale perturbations (corresponding to larger k) is suppressed in the matter power spectrum, leading to a slope of $P(k) \propto k^{-3}$. In contrast, the larger scale modes (corresponding to smaller k) crossed the Hubble horizon in the matter-dominated epoch and, therefore, did not experience any interruption in their development. Consequently, these modes follow an almost scale-invariant power spectrum with a slope of $P(k) \propto k^{n_\mathrm{s}}$, where $n_\mathrm{s} = 0.965 \pm 0.004$ corresponds to the scalar spectral index (Planck Collaboration, Aghanim, et al., 2018). The peak that transitions between the two regimes corresponds to the modes that entered the Hubble horizon during matter-radiation equality and hence, its position is sensitive to the ratio of photon to matter densities, $\Omega_\gamma/\Omega_\mathrm{m}$.

1.3.3 Halo mass function

The next step is to consider what happens when an overdensity of mass M and radius R enters the stage of non-linear collapse. In practice, it is convenient to make the approximation of spherical symmetry and focus on the mass within a finite volume defined by a "window function". The smoothed density field $\delta_\mathrm{R}(\boldsymbol{x})$ is then obtained by the convolution of the density field $\delta(\boldsymbol{x})$ with a top-hat window function $W_\mathrm{R}(\boldsymbol{x})$:

$$\delta_\mathrm{R}(\boldsymbol{x}) = \int \delta(\boldsymbol{y}) W_\mathrm{R}(|\boldsymbol{x}-\boldsymbol{y}|)\mathrm{d}\boldsymbol{y}. \tag{1.47}$$

Essentially, this smooths out perturbations of sizes $< R$. We can then compute the variance of the fluctuations at scale R as:

$$\sigma^2(R) = \langle \delta_R^2 \rangle = \frac{1}{2\pi^2} \int P(k) \left|\hat{W}_\mathrm{R}(k)\right|^2 k^2 \mathrm{d}k, \tag{1.48}$$

with $\hat{W}_\mathrm{R}(k)$ the Fourier transform of the top-hat window function; $\hat{W}_\mathrm{R}(k) = [3/(kR)^3][\sin(kR) - (kR)\cos(kR)]$. The variance of the density field evaluated at comoving radius $R = 8$ Mpc/h is defined as σ_8, an important cosmological parameter that relates to the clustering of matter and provides the normalisation for the matter power spectrum. As it turns out, a sphere of size 8 Mpc/h contains the typical mass to form a moderately rich galaxy cluster, thereby making the galaxy cluster mass function a convenient way to measure σ_8.

An analytical description of the halo mass function (HMF) that evolves Gaussian initial conditions in a spherical collapse model was obtained by Press and Schechter (1974). Their formalism states that the fraction of mass contained in halos with mass greater than M is equal to the fraction of the density field (smoothed on a scale σ_M) that exceeds a given critical density threshold δ_c. The corresponding Press-Schechter mass function describes the halo abundance per unit mass per unit comoving volume:

$$\frac{\mathrm{d}n}{\mathrm{d}M} = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M^2} \frac{\delta_\mathrm{c}}{\sigma_\mathrm{M}(z)} \left|\frac{\mathrm{d}\log\sigma_\mathrm{M}(z)}{\mathrm{d}\log M}\right| \exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_\mathrm{M}(z)^2}\right), \tag{1.49}$$

where $\bar{\rho} = \Omega_\mathrm{m}\rho_\mathrm{crit}$ is the mean density of the Universe. Initially, the Press-Schechter formalism underestimated the total amount of halos by a factor of 2, since it ignored underdense regions within larger collapsed objects. After accounting for this missing factor of 2, results from the Press-Schechter mass function are in reasonable agreement with numerical simulations, a remarkable fact given the underlying simplified assumptions. However, the mass function requires to be calibrated against cosmological simulations and is not able to describe the process of halo formation fully accurately.

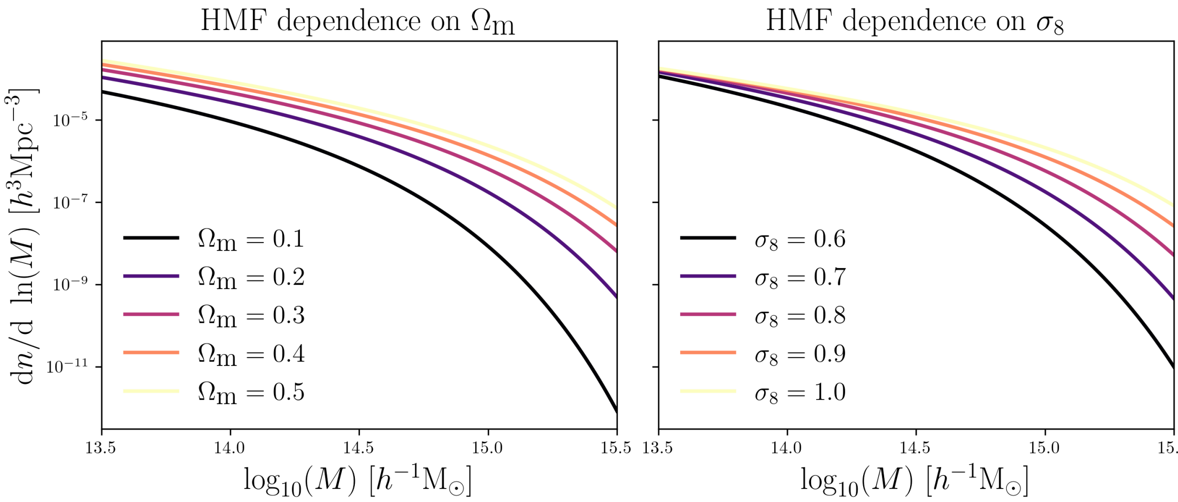


FIGURE 1.6 – Halo mass function (HMF) at $z = 0$ for different values of Ω_m and σ_8, demonstrating the sensitivity of the HMF to these cosmological parameters. Figure constructed using the matter power spectrum from CAMB[1] (Code for Anisotropies in the Microwave Background), with cosmological parameters from Planck Collaboration, Akrami, et al. (2018).

An improvement over the Press-Schechter mass function was obtained by Tinker et al. (2008), who introduced additional fitting parameters to make the HMF more in line with predictions from cosmological simulations, resulting in the following form:

$$\frac{dn}{d\ln M} = \frac{\bar{\rho}}{M}\left|\frac{d\ln\sigma}{d\ln M}\right| f(\sigma). \tag{1.50}$$

The function $f(\sigma)$ is presumed to be universal to changes in cosmology and is given by:

$$f(\sigma) = A\left[\frac{\sigma}{b}^{-a} + 1\right]e^{-c/\sigma^2}, \tag{1.51}$$

where the constants A, a, b and c are determined from simulations. The parameter A corresponds to the overall amplitude of the mass function, a and b set the slope and amplitude of the low-mass power spectrum, respectively, while c denotes the cut-off scale at which the abundance of halos decreases exponentially.

The resulting shape of the HMF using eq. 1.50 is displayed in Figure 1.6, illustrating its sensitivity to the parameters Ω_m and σ_8. A higher matter content or a larger amplitude of density perturbations both result in more high-mass structures, although changes in σ_8 leave the HMF at the lower mass end unaltered. Consequently, the HMF can be compared to an observational measurement of the cluster mass function, as discussed in the following subsection, to yield constraints on Ω_m and σ_8.

[1] https://camb.readthedocs.io/en/latest/

1.3.4 Galaxy clusters

The highest peaks in the initial density field evolved to become galaxy clusters; the most massive gravitationally bound structures in the Universe. Clusters dominate the high-mass tail of the HMF and are, therefore, an important probe of cosmological structure formation (Borgani, 2008). The cluster mass function can be computed by counting the number of galaxy clusters in a logarithmic mass interval Δ per unit comoving volume:

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M} \approx \frac{1}{\Delta}\sum_i \frac{1}{V}\frac{1}{\mathcal{S}(M_i, d_i)}, \tag{1.52}$$

for all i that satisfy $|\ln M_i - \ln M| \leq \frac{\Delta}{2}$. V is the comoving volume of the survey and $\mathcal{S}$ is the selection function, which takes into account the survey's level of incompleteness.

Measuring the mass of galaxy clusters is a challenging endeavour, for which several different observational techniques can be employed. Cluster mass is not a direct observable; instead, it is usually derived from observational properties that correlate with mass. Several assumptions are regularly made to simplify the process of mass estimation, such as spherical symmetry and hydrostatic or dynamical equilibrium. However, clusters are highly complex and dynamical objects, often surrounded by infalling structures (Gunn & Gott, 1972) and showing signs of recent mergers. Additionally, mass estimates are often plagued by projection effects, which arise from uncertainties in distance measurements and can incorrectly assign non-member galaxies to a cluster. The most commonly used methods to obtain galaxy cluster mass estimates are described below.

- **X-ray observations** can map the hot, ionised gas in the intracluster medium, which behaves as a plasma and emits thermal bremsstrahlung. Since the radiation emitted is proportional to the square of the gas density, it traces the inner parts of the cluster and is consequently not as susceptible to projection effects. The X-ray luminosity shows a tight correlation with cluster mass and the X-ray cluster mass function can provide competitive cosmological constraints (Mantz et al., 2014; Vikhlinin et al., 2009). Currently, eROSITA is in the process of detecting $\sim 100,000$ galaxy clusters up to redshift $z \sim 1.3$ (Merloni et al., 2012).
- **Gravitational lensing** is able to constrain cluster masses due to their gravitational influence on light rays. The deep gravitational potential well of a galaxy cluster can cause distortions in light (weak lensing), as well as produce multiple images of the same object (strong lensing). The effects are proportional to the cluster's mass and are free of assumptions about the dynamical state of the matter (see, for e.g., Dahle, 2006; Hoekstra et al., 2013; Umetsu, 2020).
- The **Sunyaev-Zel'dovich (SZ) effect** maps hot electrons in the intracluster medium by means of inverse compton scattering with CMB photons (Sunyaev & Zeldovich, 1972). This process boosts the energies of low-energy CMB photons, thereby distorting the blackbody spectrum. Since the SZ effect is independent of the cluster distance, this method is very promising for high-redshift clusters. Both the South Pole Telescope (N. Huang et al., 2020) and the Atacama Cosmology Telescope (Hilton et al., 2021) are currently producing large catalogues of galaxy clusters via the SZ effect.

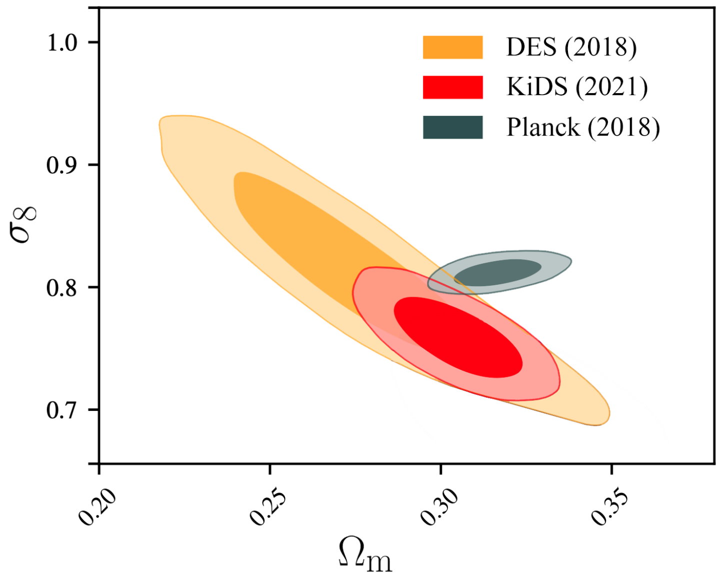


FIGURE 1.7 – Posterior contours in the $\sigma_8 - \Omega_m$ plane for large-scale structure measurements (LSS) from DES and KiDS, and CMB measurements from Planck. The figure illustrates the weak tension between LSS and CMB measurements in terms of these parameters. Figure adapted from Heymans et al. (2021).

- **Optical and near IR emission** of galaxy clusters is mainly radiated as starlight, which traces the galaxy distribution of clusters. The first optical galaxy cluster catalogue was compiled by Abell (1958) as far back as 1958. Mass estimates are usually obtained from the number of member galaxies in the clusters (the *richness*), or from the cluster velocity dispersion. The latter is related to the dynamical mass by the Jeans equation (e.g. Łokas & Mamon, 2003) or the virial theorem (e.g. Abdullah, Wilson, Klypin, et al., 2020).

In Chapter 7, a new method for dynamical mass estimation is presented that does not rely on any scaling relations or assumptions regarding spherical symmetry or dynamical equilibrium. Instead, it employs a machine learning approach to optimally exploit the 3D phase-space information of the clusters and more adequately account for projection effects. The resulting cluster mass function from clusters in the Sloan Digital Sky Survey is computed and is shown to be consistent with the HMF from the Planck cosmological model.

1.3.5 The growth tension

The matter content of the Universe and the amplitude of density fluctuations are tightly constrained by the CMB as $\Omega_m = 0.315 \pm 0.007$ and $\sigma_8 = 0.811 \pm 0.006$ (Planck Collaboration, Aghanim, et al., 2018). Alternatively, they can be inferred from weak lensing constraints from large-scale structure (LSS) measurements, as performed by the Kilo-Degree Survey (KiDS; de Jong et al., 2013) and the Dark Energy Survey (DES; The Dark Energy Survey Collaboration, 2005). Figure 1.7 illustrates the weak tension that currently exists between LSS and CMB

FIGURE 1.8 – G299.2-2.9: a type Ia supernova remnant. Credits: NASA & 2MASS.

observations, known as the *growth tension*. With the latest measurements, the tension is at the level of $\sim 2\sigma$ (Amon et al., 2021; Heymans et al., 2021). Independent determinations of σ_8 and Ω_m are important to assess whether this tension is due to a statistical fluctuation or has an underlying physical cause.

Galaxy clusters can provide such an alternative measurement. Thus far, cosmological constraints obtained via the SZ effect (Salvati et al., 2018), X-ray mass estimates (Mantz et al., 2014), abundance and weak-lensing mass measurements (Rozo et al., 2010), and dynamical mass estimates (Abdullah, Klypin, et al., 2020) are in agreement with the CMB measurements.

1.4 LOW-REDSHIFT DISTANCE MEASUREMENTS

In the low-redshift Universe, there are a myriad of astrophysical objects that are utilised for distance determinations to ultimately constrain our cosmological model. This section gives an overview of distance indicators that have been used throughout the work in the thesis. It draws inspiration from Alam et al. (2017), Andersen (2018), Philcox et al. (2020), Sánchez et al. (2017), Satpathy et al. (2017), and Yuan (2017).

1.4.1 Type Ia supernovae

Type Ia supernovae are bright thermonuclear explosions that occur in binary systems where at least one of the stars is a white dwarf. They are classified as 'type I' on the basis that their spectra do not contain any hydrogen emission lines. By virtue of their consistent peak luminosity, they are excellent standard candles, such that they can be employed to measure distances to their host galaxies and to constrain cosmological parameters (Riess et al., 1998). The remnant of a type Ia supernova can be seen in Figure 1.8.

Currently, two main progenitor scenarios are proposed to explain type Ia supernovae; the single degenerate and the double degenerate systems. In the single degenerate scenario, a white dwarf accretes matter from a companion star until it reaches the Chandrasekhar mass limit of $\sim 1.4\,M_\odot$ (Chandrasekhar, 1931) and explodes (Nomoto, 1982; Whelan & Iben, 1973). The double degenerate scenario is caused by the merger of two white dwarfs that also results in a final

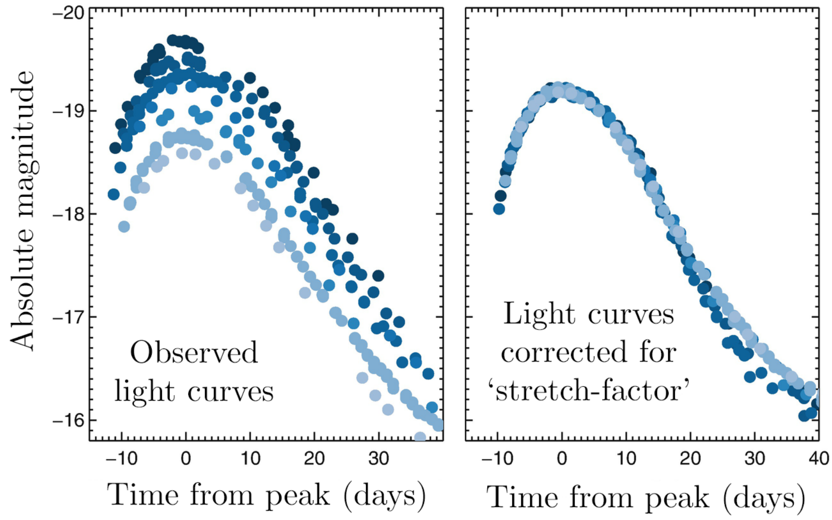


FIGURE 1.9 – *Left panel:* Light curves of type Ia supernovae observed in the *B*-band by Hicken et al. (2009) and Stritzinger et al. (2011), clearly demonstrating the correlation between light curve width ('stretch-factor') and peak *B*-band magnitude. *Right panel:* Correcting the light curves for the stretch-factor greatly reduces the scatter. Figure adapted from Maguire (2017).

mass close to the Chandrasekhar mass (Iben & Tutukov, 1984; Webbink, 1984). Surprisingly, these two evolutionary channels seem to result in thermonuclear explosions with a high degree of homogeneity, although the uncertainty in progenitor scenario may introduce some unknown systematic errors in the standardisation of type Ia supernovae.

Observed supernova brightness is generally provided in terms of the *apparent magnitude* m in a given filter x, which is related to the flux in that filter F_x by:

$$m_x \equiv -2.5\log_{10}\left(\frac{F_x}{F_{x,0}}\right), \tag{1.53}$$

where F_x is defined in eq. (1.21) and $F_{x,0}$ is the reference flux (or zero-point) for the filter x. The apparent magnitude is a measure of how bright we observe an object to be from Earth, i.e. it depends on the distance to the object and any extinction along the line of sight. The *absolute magnitude* M corresponds to the intrinsic luminosity emitted by the source, defined as the apparent magnitude the object would have if it were exactly 10 parsecs away. If both the apparent and absolute magnitudes are known, the luminosity distance d_L in parsecs to an object can be determined via the following relationship:

$$M = m + 5 - 5\log_{10}(d_L). \tag{1.54}$$

Often, distances are expressed on a logarithmic scale by means of the *distance modulus* μ, which is defined as:

$$\mu \equiv m - M. \tag{1.55}$$

In 1993, Phillips (1993) greatly enhanced the standard candle potential of type Ia supernovae by discovering that their peak brightness correlates tightly with the decline rate of their light curve. Broader light curves are generally brighter, as illustrated in Figure 1.9. A few years later, an additional correlation was found between the peak luminosity and optical colour (Riess et al., 1996; Tripp, 1998); bluer supernovae are brighter. These findings are summarised in the *Tripp formula*, which yields the absolute B-band peak magnitude M_B that a type Ia supernova, based on its light curve width and colour, is expected to have:

$$M_\mathrm{B} = -\alpha x_1 + \beta c + M_0, \tag{1.56}$$

with c being the colour parameter and x_1 denoting the stretch parameter, that characterises the width of the light curve. M_0 is the expected absolute magnitude of a supernova with $x_1 = c = 0$, which takes on a value of $M_0 \approx -19.3$, depending on the specific calibration. Although the colour and stretch corrections greatly reduce the scatter in observed peak magnitudes, some intrinsic scatter of ~ 0.12 magnitude remains. It is likely that these small variations can be accredited to the properties of the host galaxy, dust along the line of sight, position in the host galaxy, and, potentially, the progenitor scenario leading to the formation of the supernova.

Type Ia supernovae have played a crucial role in our current understanding of the Universe. The Nobel prize for physics of 2011 was awarded to Saul Perlmutter, Brian P. Schmidt and Adam G. Riess for *"the discovery of the accelerating expansion of the Universe through observations of distant supernovae"* (Perlmutter et al., 1998; Riess et al., 1998; Schmidt et al., 1998). Their findings can be seen in Figure 1.10. As both teams demonstrated, type Ia supernovae are an invaluable tool for mapping out the cosmic expansion history and confirming the existence of dark energy. The supernovae do, however, only allow for relative distance measurements, and need to be calibrated first in order to provide constraints on the Hubble constant.

1.4.2 Baryon Acoustic Oscillations

The Baryon Acoustic Oscillations (BAO), as discussed in Section 1.2, are imprints of the oscillations in the photon-baryon plasma that form the seeds of structure formation. The galaxy distribution follows these patterns, such that there is a correlation between galaxies in the centre of the primordial overdensities and those in the shells around them. This preferred scale of clustering acts as a *standard ruler*; an object whose physical size is known and, therefore, can be employed to determine distances by measuring its angular size in the sky (Bassett & Hlozek, 2010). Since the Universe consists of many multi-scale structures that are overlaid on each other, the BAO feature cannot be seen by eye but can only be extracted using statistical methods. The signal shows up as a peak in the two-point correlation function, indicating an excess of clustering at that scale (left panel of Figure 1.11). Because the power spectrum is the Fourier transform of the correlation function, the characteristic signal emerges as a series of oscillations in Fourier space (right panel of Figure 1.11). The BAO feature was first detected in 2005 in the two-point correlation function of the Sloan Digital Sky Survey (D. J. Eisenstein et al., 2005) and in the power spectrum of the 2dF Galaxy Redshift Survey (Cole et al., 2005).

In addition to the clustering of galaxies, the BAO signal has also been observed in the correlation function of the Lyman-α forest absorption (de Sainte Agathe et al., 2019) and in its

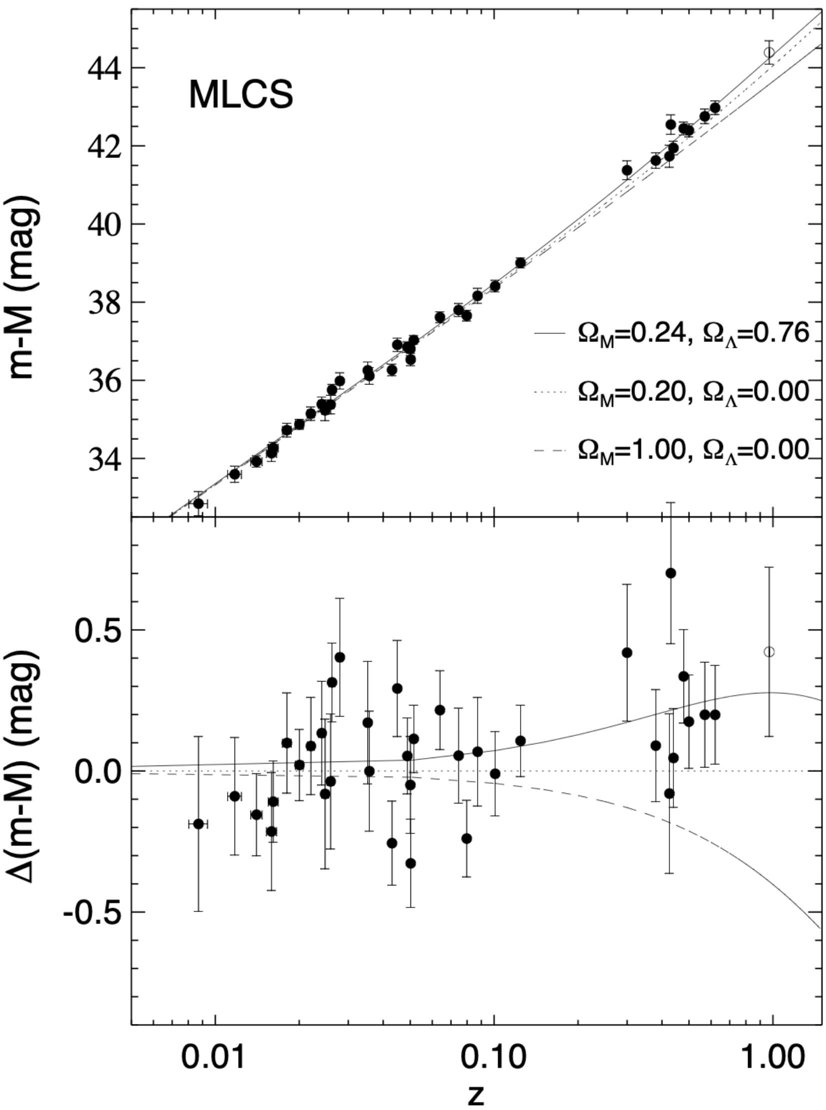


Figure 1.10 – Figure from the Nobel prize winning paper by Riess et al. (1998) that employs type Ia supernovae to demonstrate the accelerating expansion of the Universe and the existence of dark energy. The supernova data points are consistent with $\Omega_m = 0.24$ and $\Omega_\Lambda = 0.76$.

cross-correlation with quasars (Blomqvist et al., 2019). A future spectroscopic galaxy survey by the Dark Energy Spectroscopic Instrument (DESI) will cover 14000 square degrees and measure the BAO signal out to a redshift of 3.5 through a combination of Lyman-α forest absorption and galaxy and quasar clustering (DESI Collaboration et al., 2016).

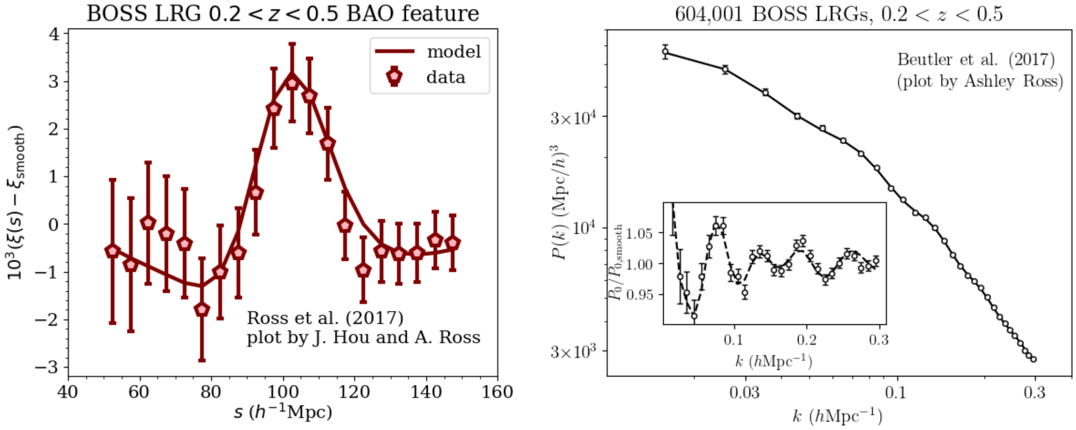


FIGURE 1.11 – BAO signal in the two-point correlation function (*left panel*) and the power spectrum (*right panel*) of BOSS data (Alam et al., 2017).

1.4.2.1 *Radial and tangential features*

There are two distinct ways in which BAO measurements can provide cosmological constraints, which are illustrated in Figure 1.12. Firstly, the tangential BAO length $s_\perp$ constrains the angular diameter distance through the geometric relation:

$$d_{\rm A}(z) = \frac{s_\perp}{\Delta\theta(1+z)}, \tag{1.57}$$

where $\Delta\theta$ is the angle under which we observe the BAO signal. Secondly, the BAO scale along the line of sight $s_\|$ is sensitive to the expansion of the Universe, thereby allowing a measurement of the Hubble parameter:

$$H(z) = \frac{c\Delta z}{s_\|(z)}, \tag{1.58}$$

where Δz is the redshift difference between the front and the back of the BAO feature. Isotropic BAO measurements do not have sufficient statistical power to separate the radial and tangential BAO features and instead combine them in a volume-averaged distance:

$$d_{\rm V} = \left[c\, z\, d_{\rm M}^2(z)\, H^{-1}(z)\right]^{1/3}. \tag{1.59}$$

Additional information can be obtained through the Alcock-Paczynski test (Alcock & Paczynski, 1979; Montanari & Durrer, 2012), which uses the fact that the BAO shells are statistically spherical structures, implying that $s_\|$ is equal to $s_\perp$. This provides constraints on the product $H(z) \times d_{\rm A}$ that are complementary to the separate $H(z)$ and $d_{\rm A}$ measurements. Although the BAO shells are spherical in real space, distances obtained in redshift space contain contributions from peculiar velocities of the galaxies that deviate from the pure Hubble flow. Therefore, the reconstructed distances suffer from distortions along the radial direction, called *redshift space*

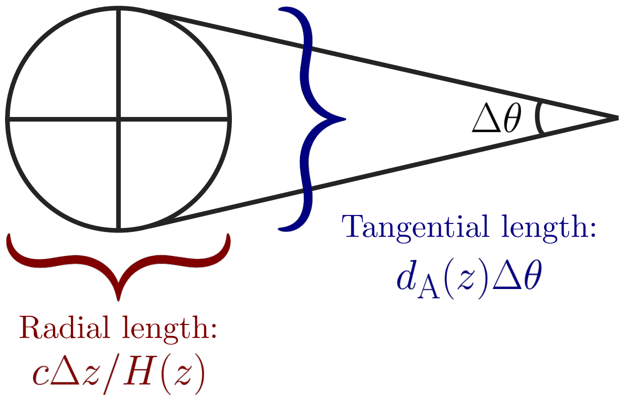


FIGURE 1.12 – BAO measurements constrain the cosmic expansion rate in the radial direction and the angular diameter distance tangentially. By assuming that the BAO features are spherical, the Alcock-Paczynski test provides additional constraints on $H(z) \times d_A$. Figure inspired by Bassett and Hlozek (2010).

distortions (RSD; Kaiser, 1987). On small scales, the spatial distribution of galaxies appears to be elongated due to their velocity dispersion along the line of sight, a phenomenon known as the *'Fingers of God'* effect (Jackson, 1972). On larger scales, RSD are caused by infall velocities of galaxies that are gravitationally attracted to larger clusters, resulting in an apparent flattening of structures in the radial direction. These two opposing effects combine to form complex distortion patterns in redshift maps. Since the infall velocities are sensitive to the logarithmic growth rate f of structures (eq. (1.43)), which is usually reported in terms of $f\sigma_8$, it follows that BAO measurements are capable of constraining $f\sigma_8$ in addition to $H(z)$ and d_A.

1.4.2.2 *BAO-only and full-shape analyses*

The analysis of BAO data can be carried out in two different ways. The first approach is the **BAO-only** method, which focuses solely on identifying the BAO features in the correlation function and power spectrum. This analysis employs a reconstruction technique to reduce the effects of RSD and to displace galaxies closer to their initial positions, where they were located before large-scale bulk flows weakened the BAO signal. This procedure sharpens the peak considerably. However, in order to do so, it assumes a cosmological model and a value of the growth rate of structures. Both an analytical calculation of the bias introduced by the reconstruction technique (Sherwin & White, 2019) and an investigation of the bias using simulations for an underlying wCDM cosmology (Carter et al., 2020) find that the introduced shift of the BAO peak location is negligible for current data sets, although this might become relevant for future high-precision observations.

In practice, the BAO-only analysis is carried out by constraining the Alcock-Paczynski parameters $\boldsymbol{\alpha} = \{\alpha_\parallel, \alpha_\perp\}$, which represent deviations of the radial and tangential BAO features from an assumed fiducial cosmology. The fiducial cosmology is adopted to convert redshifts into distances, as well as to supply a template for the correlation function and power spectrum.

The shift between the observed BAO feature and the one in the fiducial template depends on the cosmic expansion rate, angular diameter distance and sound horizon at the redshift of decoupling in the following way:

$$\alpha_{\parallel} = \frac{H^{\rm fid}(z)\, r_{\rm d}^{\rm fid}}{H(z)\, r_{\rm d}}, \qquad \alpha_{\perp} = \frac{d_{\rm A}(z)\, r_{\rm d}^{\rm fid}}{d_{\rm A}^{\rm fid}(z)\, r_{\rm d}}, \tag{1.60}$$

where z corresponds to the effective redshift of the sample and the superscript 'fid' refers to the assumed fiducial cosmology. In this way, BAO measurements constrain the combination $H(z) r_{\rm d}$ or $d_{\rm A}/r_{\rm d}$. Only by assuming a value for the sound horizon, e.g. from CMB physics or Big Bang Nucleosysnthesis (BBN), do they provide absolute measurements of distances or the cosmic expansion rate.

The second method is the **full-shape** analysis, which models the full correlation function and power spectrum and does not include any reconstruction techniques. This approach also models the RSD effects with $f\sigma_8$ as a free parameter. Therefore, the full-shape analysis is the preferred one when it comes to inference with minimal cosmological assumptions. However, the perturbation theory model which is used to construct the correlation function and power spectrum still adopts a ΛCDM model. Additionally, for both the full-shape and the BAO-only methods, a cosmological model is used to construct the covariance matrices from mock galaxy catalogues. Two distinct sets of mock simulations with a different underlying cosmology were investigated for this purpose, yielding consistent results (Alam et al., 2017).

Recently, a new analysis that combines the BAO-only and full-shape methods, while adopting a prior on the baryon density from BBN, has obtained a 1.6% measurement of the Hubble constant. The resulting value of $H_0 = 68.6 \pm 1.1$ km s^{-1} Mpc^{-1} is mostly independent of, and completely consistent with, CMB constraints (Philcox et al., 2020).

1.4.3 Going to high-redshift with quasars

Quasars can be employed as relative distance indicators, with the advantage of being visible up to redshifts above 7. Their usefulness lies in the non-linear relation between ultraviolet emission from their accretion disk and X-ray emission from the corona in the surrounding region (Avni & Tananbaum, 1986; Tananbaum et al., 1979; Zamorani et al., 1981), which is generally parametrised as:

$$\log_{10}(L_{\rm X}) = \gamma \log_{10}(L_{\rm UV}) + \beta, \tag{1.61}$$

where $L_{\rm X}$ denotes the X-ray luminosity, $L_{\rm UV}$ the ultraviolet luminosity, while $\gamma \sim 0.6$ and $\beta \sim 9$ are fitting parameters. Risaliti and Lusso (2018) have constructed a Hubble diagram from high-redshift quasars, where they employ a cosmographic parametrisation to describe the cosmic expansion history. They find that the high-redshift quasars are in agreement with type Ia supernovae and the concordance model up to $z \sim 1.4$, but deviate from ΛCDM at higher redshifts. However, as high-redshift quasars exhibit a large amount of intrinsic scatter, their reliability as standard candles is still a matter of debate (Velten & Gomes, 2020). Additionally,

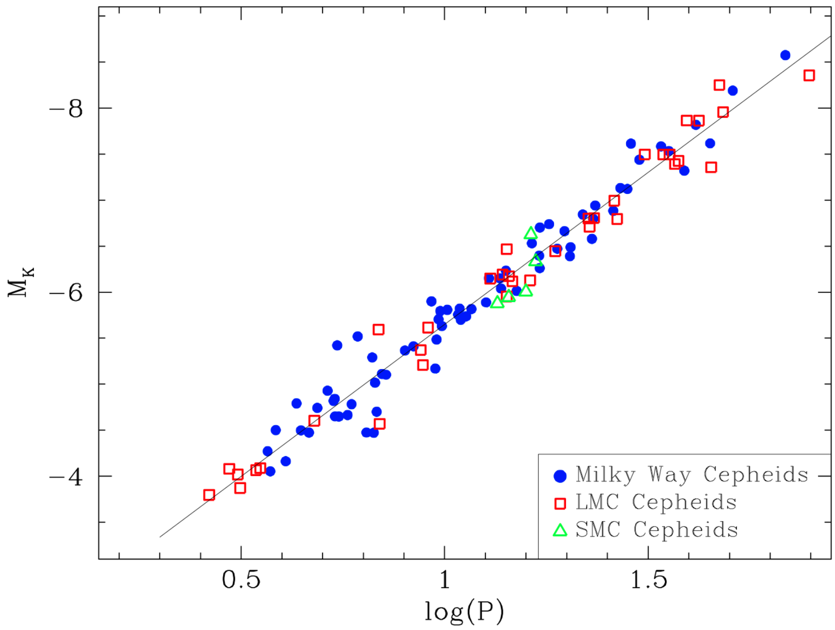


FIGURE 1.13 – The period-luminosity relation in the K-band for 111 Cepheids from the Milky Way, Large Magellanic Clouds (LMC) and Small Magellanic Clouds (SMC). Figure reprinted from Storm et al. (2011).

it has been pointed out that the use of cosmographic expansions at high redshift can introduce artificial tensions (Yang et al., 2019).

In Chapter 3, high-redshift quasars are used as an optional standard candle in addition to type Ia supernovae and BAO.

1.4.4 Cepheids

Cepheids are massive stars that are transitioning from the main sequence in the colour-magnitude diagram to the giant branch and, hence, display regular pulsations in their radius and temperature that are tightly correlated with their pulsation period (Leavitt, 1908). The pulsation mechanism is thought to stem from the opposing forces of gravity and radiation pressure in an ionised helium layer in the star's envelope. The period-luminosity relation, as displayed in Figure 1.13, takes on the following form:

$$M = a\log(P) + f(T), \tag{1.62}$$

where M denotes the absolute magnitude, P is the pulsation period, $f(T)$ represents the temperature dependence and the coefficient a is equal to $-10/3$. The correlation between the period and luminosity makes Cepheids great standard candles that are highly suitable for absolute distance measurements. To this end, the period-luminosity relation is first calibrated with geometric measurements of nearby distances, such as trigonometric parallaxes (Breuval et al., 2020; Riess, Casertano, Yuan, Macri, Anderson, et al., 2018), eclipsing binary systems (Pietrzyński et al., 2019) and water masers (M. J. Reid et al., 2019). Subsequently, the distances inferred for the Cepheids are used to calibrate type Ia supernovae in galaxies that contain both objects. This step-wise approach of building up larger and larger distances is known as the *cosmic distance ladder*, and when used in combination with Cepheids, it provides the tightest low-redshift constraints on the Hubble constant. The analysis of the *Supernovae and H_0 for the Equation of State of dark energy* project (SH0ES; Riess et al., 2021) yields a measurement of $H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1}.

1.4.5 Tip of the Red Giant Branch

As low-mass stars reach the end of the red giant branch evolutionary track in the colour-magnitude diagram, the helium at their core will undergo nuclear fusion in a process called a *helium flash*. The helium flash leads to a sharp discontinuity in the colour-magnitude diagram which occurs at a predictable luminosity, making the Tip of the Red Giant Branch (TRGB) an excellent standard candle. Similarly to Cepheids, the TRGB can be calibrated with nearby geometric measurements and used as a rung in the distance ladder to determine absolute distances to supernovae.

The TRGB method holds several advantages over Cepheid distance measurements. While Cepheid variables are mostly found in the galactic disk, TRGB stars are located in the halo of the galaxy. Therefore, they are less affected by dust extinction and the overlapping of point-spread functions of surrounding stars (‘*crowding*’). Additionally, they can be found in galaxies of all morphological types, while Cepheid variables only occur in spiral galaxies, which might introduce biases. The effects of different metallicities of TRGB stars are well-understood and can be corrected for. Finally, while Cepheids require multi-epoch observations to accurately map out their pulsation period, the TRGB method only requires a single-epoch observation in two filters.

The Carnegie Supernova Project has employed the TRGB method in combination with type Ia supernovae (Freedman, 2021) to yield a Hubble constant of $H_0 = 69.8 \pm 0.6$ (stat) ± 1.6 (sys) km s^{-1} Mpc^{-1}.

1.4.6 Gravitational lensing

Gravitational lensing is the bending of light due to a gravitational field, as predicted by Einstein’s theory of general relativity. In the Universe, objects such as galaxies and clusters can act as gravitational lenses that magnify and distort light from distant background sources. These effects allow us to see further into the Universe, map out the distribution of matter in the lens, determine distances to the lens system and ultimately infer the cosmic expansion rate (Narayan & Bartelmann, 1996).

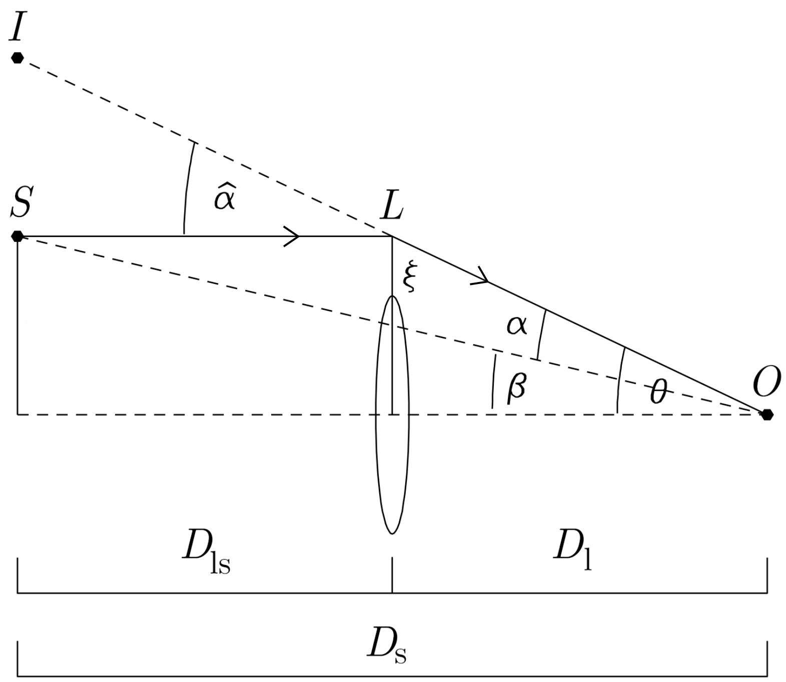


FIGURE 1.14 – Schematic of a gravitational lens system. Light travelling from a source S is deflected by a lens L under an angle $\hat{\alpha}$, such that it looks for an observer O as though it originates from an image I. The optical axis makes an angle β with the source and θ with the image. Angular diameter distances are given between the observer and the lens ($D_{\rm l}$), the observer and the source ($D_{\rm s}$) and the lens and the source ($D_{\rm ls}$). Figure adapted from Narayan and Bartelmann (1996).

1.4.6.1 *Lensing formalism*

When light passes by an object of mass M at a distance ξ, it is bent by a deflection angle $\hat{\alpha}$:

$$\hat{\alpha} = \frac{4GM}{c^2 \xi}. \tag{1.63}$$

Figure 1.14 shows the geometry of a typical lens system. A light ray emitted by a source S is deflected under an angle $\hat{\alpha}$ by the lens L and arrives at the observer O. The original angle between the observer and the source is β, while the observed angle of the image I is θ. The angular diameter distances between the observer and the lens, observer and the source, and lens and the source are $D_{\rm l}$, $D_{\rm s}$, and $D_{\rm ls}$, respectively. Note that in general, the relation $D_{\rm ls} = D_{\rm s} - D_{\rm l}$ does not hold. Angular diameter distances are defined such that: *observed angle* × *distance* = *separation*, and applying this relation to Figure 1.14 yields the following identity:

$$\boldsymbol{\beta} D_{\rm s} = \boldsymbol{\theta} D_{\rm s} - \hat{\boldsymbol{\alpha}} D_{\rm ls}, \tag{1.64}$$

where, in the case of a general non-symmetric mass distribution, the angles are vectors because they have a magnitude as well as a direction. It is convenient to define the *reduced deflection*

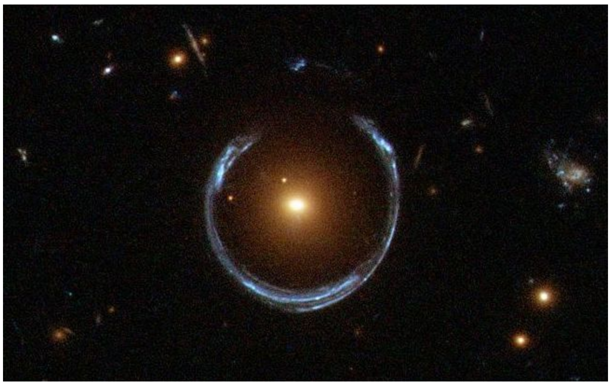


Figure 1.15 – The horseshoe Einstein ring (LRG 3-757). Image Credit: ESA/Hubble & NASA.

angle $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \frac{D_{\mathrm{ls}}}{D_{\mathrm{s}}} \hat{\boldsymbol{\alpha}}. \tag{1.65}$$

Combining eqs. (1.64) and (1.65) results in the *lens equation*:

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}. \tag{1.66}$$

The lens equation also holds in curved spacetimes, because angular diameter distances are always defined as angle × distance = separation. Solutions to the lens equation can produce multiple image positions for one source, which means that we observe the same object multiple times; a phenomenon called *strong gravitational lensing*. When a source is located right behind a nearly spherical lens galaxy, it will be warped into an *Einstein ring* around the lens galaxy, as depicted in Figure 1.15. The radius of the ring, the *Einstein radius* θ_{E}, is given by:

$$\theta_{\mathrm{E}} = \sqrt{\frac{4GM}{c^2} \frac{D_{\mathrm{ls}}}{D_{\mathrm{l}} D_{\mathrm{s}}}}. \tag{1.67}$$

The Einstein radius plays a central role in lens studies, as it relates to the mass of the lens galaxy and, therefore, to the strength of the gravitational lensing effect.

Typically, the thickness of the lens galaxy is negligible compared to the distance the light has to travel between the source and the observer. Therefore, the 3D mass distribution of the lens can be projected along the line of sight and approximated by a 2D mass sheet. This is known as the *thin lens approximation*. The projected surface mass density $\Sigma(\boldsymbol{\xi})$ in the z-direction can be calculated as:

$$\Sigma(\boldsymbol{\xi}) = \int \rho(\boldsymbol{\xi}, z) \mathrm{d}z, \tag{1.68}$$

where ρ is the 3D mass density and $\boldsymbol{\xi}$ is a 2D vector in the lens plane which is related to the image angle and lens distance as $\boldsymbol{\xi} = D_l \boldsymbol{\theta}$. The surface mass density is often represented as a dimensionless quantity, known as the convergence $\kappa(\boldsymbol{\theta})$:

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(D_\mathrm{l}\boldsymbol{\theta})}{\Sigma_\mathrm{crit}}, \tag{1.69}$$

where the critical surface mass density Σ_crit is given by:

$$\Sigma_\mathrm{crit} = \frac{c^2}{4\pi G}\frac{D_\mathrm{s}}{D_\mathrm{l}D_\mathrm{ls}}. \tag{1.70}$$

Similarly, the 3D lens potential $\Phi(\boldsymbol{\theta}, z)$ can be projected along the line of sight to yield the effective lens potential $\psi(\boldsymbol{\theta})$:

$$\psi(\boldsymbol{\theta}) = \frac{D_\mathrm{ls}}{D_\mathrm{l}D_\mathrm{s}}\frac{2}{c^2}\int \Phi(D_\mathrm{l}\boldsymbol{\theta}, z)\,\mathrm{d}z. \tag{1.71}$$

The derivatives of $\psi(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ can in turn be related to the reduced deflection angle and the convergence:

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla\psi(\boldsymbol{\theta}), \qquad \kappa(\boldsymbol{\theta}) = \frac{1}{2}\nabla^2\psi(\boldsymbol{\theta}), \tag{1.72}$$

where the Poisson equation was used to obtain the latter identity.

Gravitational lensing can magnify objects, which gives us the opportunity to observe astrophysical objects that would normally be too faint to detect. Since the lensing process conserves surface brightness, the magnification is simply defined as: *image area / source area*. The shape and size of the image are dictated by a combination of isotropic magnification and distortions. The convergence κ acts to cause an isotropic magnification, which maps the source into an image with a different size but the same shape. The shear γ distorts and tangentially stretches the image. In terms of the convergence and shear, the magnification μ can be written as:

$$\mu = \frac{1}{(1-\kappa)^2 - \gamma^2}. \tag{1.73}$$

The regions in the image plane where μ is infinite are called *critical curves*, and the corresponding lines in the source plane are *caustics*. The location of the source within the caustics determines the image multiplicity; a source in the inner, middle and outer caustic region produces five, three and one image, respectively. Typically, in the case of strong lensing we only observe four or two images, because the central image is highly demagnified. When the source crosses from an inner to an outer caustic, two images merge and disappear. For an elliptical lens, the inner caustic takes on a diamond shape that consists of *fold caustics* (the smooth regions) and *cusp caustics* (the tips). Figure 1.16 illustrates the image configurations resulting from a source moving across a fold and a cusp caustic.

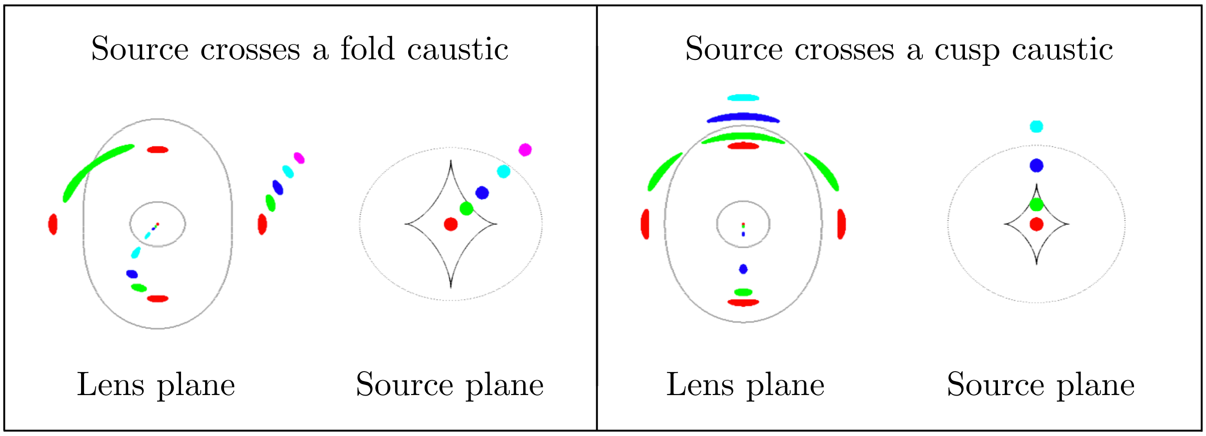


FIGURE 1.16 – A source behind an elliptical lens moving from the centre toward the outer regions. As it progresses, it crosses a fold caustic (left panel) or a cusp caustic (right panel) in the source plane, which leads to different image configurations in the lens plane. Figure adapted from Narayan and Bartelmann (1996).

1.4.6.2 *Time-delay cosmography*

If the object behind the lens galaxy is a variable source, such as a quasar or a supernova, the images will vary in brightness. These variations will arrive at the image positions at different moments in time, due to a geometric time delay and a gravitational (or Shapiro) time delay (Shapiro, 1964). For a single lens plane, the excess time delay of an image relative to an unperturbed path is:

$$t(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{D_{\Delta t}}{c} \cdot \phi(\boldsymbol{\theta}, \boldsymbol{\beta}), \tag{1.74}$$

$$\phi(\boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{(\boldsymbol{\theta} - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}), \tag{1.75}$$

where $\phi(\boldsymbol{\theta}, \boldsymbol{\beta})$ is called the *Fermat potential*, which consists of a geometrical term $(\boldsymbol{\theta} - \boldsymbol{\beta})^2/2$ and a gravitational term $\psi(\boldsymbol{\theta})$. $D_{\Delta t}$ is the time-delay distance (Refsdal, 1964; Schneider et al., 1992; Suyu et al., 2010), which is defined as:

$$D_{\Delta t} \equiv (1 + z_{\mathrm{l}}) \frac{D_{\mathrm{l}} D_{\mathrm{s}}}{D_{\mathrm{ls}}}, \tag{1.76}$$

where z_{l} is the lens redshift. For a given source, the images are formed at the stationary points of the time-delay surface $t(\boldsymbol{\theta})$. This follows from *Fermat's principle*, which states that the path taken by a light ray between two points is the one that can be travelled in a time that is stationary with respect to neighbouring trajectories. Figure 1.17 illustrates that the geometric and gravitational effects are opposed: in terms of geometry, light would arrive first at the centre of the lens, whereas time there is strongly slowed down due to gravitational time dilation. The result of these opposing effects is that brightness fluctuations generally arrive first at the image position furthest away from the lens.

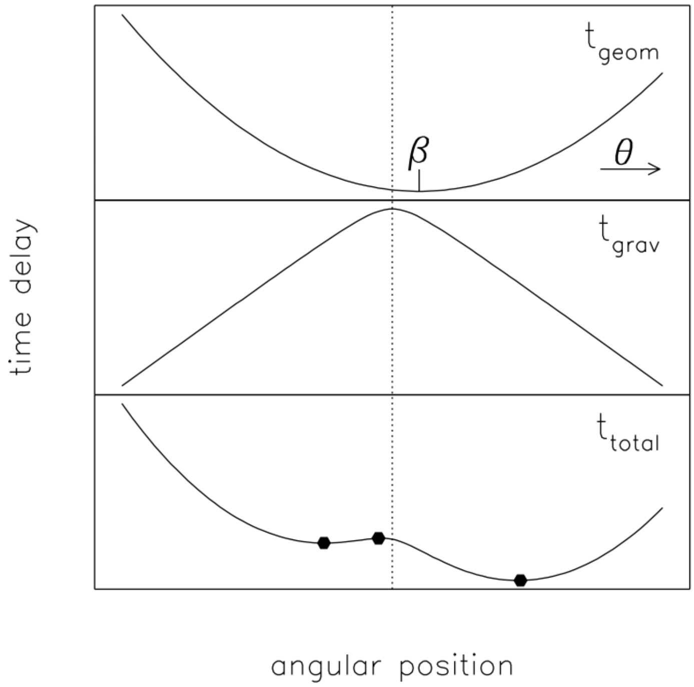


FIGURE 1.17 – Geometric (t_{geom}), gravitational (t_{grav}) and total (t_{total}) time-delay surfaces for a spherical lens and a source slightly offset behind it. The dotted vertical line marks the location of the lens and β signifies the source position. The black dots correspond to the images, which are formed at the stationary points of the total time-delay surface. Figure reprinted from Narayan and Bartelmann (1996).

A major challenge in time-delay cosmography is the mass-sheet degeneracy (E. E. Falco et al., 1985; Gorenstein et al., 1988; Saha, 2000; Schneider & Sluse, 2013), which is a transformation of the mass distribution by an arbitrary constant λ of the following form:

$$\kappa \rightarrow \kappa = \lambda\kappa' + (1 - \lambda), \tag{1.77}$$

together with a scaling of the source plane coordinates $\beta \rightarrow \lambda\beta$. This transformation, which is equivalent to adding a sheet of constant surface mass density $1 - \lambda$, leaves all imaging observables the same, while changing the time delays. This additional mass sheet can be physically interpreted in two different ways. Firstly, as an external convergence κ_{ext} due to lensing contributions from substructures along the line of sight. The external convergence can be estimated using galaxy number counts (Rusu et al., 2017) or weak lensing estimates (Tihhonova et al., 2018) in combination with cosmological simulations for each line of sight. The second interpretation is a transformation of the internal density profile of the main deflector, which can be constrained by information regarding the lensing potential, such as stellar kinematics of the lens galaxy (Birrer et al., 2020; Birrer et al., 2016; Koopmans, 2004; Shajib et al., 2021). Besides contributing to breaking the mass-sheet degeneracy, stellar kinematics can also be employed to determine the angular diameter distance to the lens galaxy (Jee et al., 2019; Paraficz & Hjorth, 2009; Shajib et al., 2018). For a singular isothermal sphere lens profile (see eq. 1.79) with velocity dispersion σ, the distance to the lens galaxy can be obtained via the following identity (Paraficz & Hjorth, 2009):

$$D_{\mathrm{l}} = \frac{c^3 \Delta t}{4\pi\sigma^2 (1+z_{\mathrm{l}})(\theta_2 - \theta_1)}, \tag{1.78}$$

where θ_1 and θ_2 denote the image positions and Δt is the corresponding time delay between the images.

In order to determine the time-delay distance, a measurement of the time delay and a realistic model for the lens potential are required (see eq. (1.74)). Distances are primarily sensitive to the Hubble constant, since a higher (lower) value of H_0 corresponds to shorter (longer) distances, respectively. The time-delay distance also depends weakly on several other cosmological parameters, such as the matter density, cosmic curvature and equation of state of dark energy, which renders time-delay cosmography a promising method for cosmological inference (Treu & Marshall, 2016). However, computing an accurate lens model from limited observations is a challenging endeavour, and for this reason, several approximate lens models are often used in lens analyses. In the next subsection, three parametric lens profiles of increasing complexity are presented.

1.4.6.3 *Lens mass profiles*

The simplest parametrisation of a lens galaxy is the singular isothermal sphere (SIS), whose convergence is given by:

$$\kappa(\boldsymbol{\theta}) = \frac{\theta_{\mathrm{E}}}{2|\boldsymbol{\theta}|}. \tag{1.79}$$

An extension to this model that includes ellipticity is the singular isothermal ellipsoid (SIE; Kormann et al., 1994):

$$\kappa(\theta_1, \theta_2) = \frac{\theta_{\mathrm{E}}}{2\sqrt{\theta_1^2 + q_{\mathrm{lens}}^2 \theta_2^2}}, \tag{1.80}$$

where q_{lens} is the projected axis ratio of the lens. A higher level of flexibility is obtained by allowing the power-law mass slope γ_{lens} to vary, instead of being fixed to 2 as in the SIE model. This is the case for the power-law elliptical mass distribution (PEMD) (Barkana, 1998; Kormann et al., 1994):

$$\kappa(\theta_1, \theta_2) = \frac{3-\gamma_{\mathrm{lens}}}{2} \left(\frac{\theta_E}{2\sqrt{\theta_1^2 + q_{\mathrm{lens}}^2 \theta_2^2}} \right)^{\gamma_{\mathrm{lens}}-1}. \tag{1.81}$$

In Chapters 5 and 6, we use the PEMD model, in combination with external shear, to simulate gravitationally lensed supernovae. Although the PEMD model has been shown to provide a decent fit to stellar kinematics (Koopmans et al., 2009) and X-ray data (Humphrey & Buote, 2010) in the local Universe, it is partly responsible for breaking the mass-sheet degeneracy, which can lead to a biased cosmological inference if the true underlying mass profile deviates from a power-law model (Birrer et al., 2020; Kochanek, 2020; Sonnenfeld, 2018).

1.4.6.4 *Lensed quasars*

Quasars are ideal candidates for time-delay cosmography, because they vary in brightness and can be seen up to high redshifts. The first strongly lensed quasar was discovered in 1979 (Walsh et al., 1979) and at present, we know of ~ 500 multiply-imaged quasars (Ducourant et al., 2018). Since quasars typically vary in brightness on long timescales and in a stochastic manner, extensive monitoring campaigns of the order of several years are required to obtain accurate time-delay measurements. Furthermore, high-resolution imaging is needed to resolve the extended emission of the quasar host galaxy, which provides additional constraints on the lens mass profile that can break the mass-sheet degeneracy.

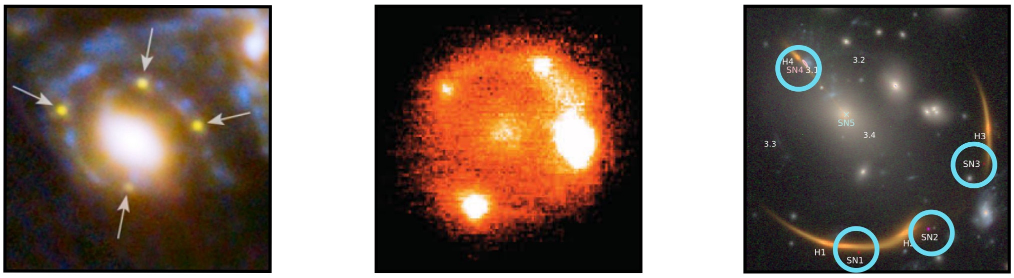
The H_0 *Lenses in COSMOGRAIL's Wellspring* collaboration (HoLiCOW, Suyu et al., 2017) has provided few per cent-level precision constraints on the Hubble constant from six gravitationally lensed quasars. The velocity dispersion measurements of the lens galaxies in the sample are not powerful enough to break the mass-sheet degeneracy completely. Assumptions about the shape of the lens mass profile (a PEMD model and a composite model consisting of a baryonic and a dark matter halo component) break the remaining part of the degeneracy. The H_0 measurement resulting from the HoLiCOW anaylis is H_0 of $73.3^{+1.7}_{-1.8}$ km s^{-1} Mpc^{-1} (Wong et al., 2020), which is completely consistent with the local distance ladder value from SHoES and in 3.1σ tension with CMB observations from *Planck*.

A recent analysis of seven lensed quasars (of which six overlap with the HoLiCOW sample) by TDCOSMO (Millon et al., 2019) relaxes the assumptions on the lens mass models, and instead includes an external data set of 33 strong gravitational lenses from the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006). Since these are lensed galaxies, they do not allow for time-delay measurements, but they do provide imaging and kinematics constraints. Combining the aforementioned data sets relies on the assumption that the lensed quasars and lensed galaxies are drawn from the same parent population. This analysis yields a Hubble constant of $67.4^{+4.1}_{-3.2}$ km s^{-1} Mpc^{-1} (Birrer et al., 2020), which is completely consistent with the CMB results, although the uncertainties are much larger.

In Chapter 4, the six lensed quasars as analysed by the HoLiCOW collaboration are used to provide constraints on the sound horizon, as well as the Hubble constant, using cosmographic methods (as discussed in Section 1.1.6) that are agnostic to the underlying cosmology.

1.4.6.5 *Lensed supernovae*

When Sjur Refsdal predicted in 1964 how the phenomena of strong gravitational lensing could be used to calculate the cosmic expansion rate, he initially suggested lensed supernovae for this purpose (Refsdal, 1964). Only recently has his vision become a reality. The discovery of the first multiply-imaged supernova ("SN Refsdal") lensed by a galaxy cluster (Kelly et al., 2015)



SN Refsdal   iPTF16geu   SN Requiem

FIGURE 1.18 – The three lensed supernovae that have been detected so far: SN Refsdal (Kelly et al., 2015), iPTF16geu (Goobar et al., 2017) and SN Requiem (S. Rodney et al., 2021).

enabled a 6 per cent precision measurement of the Hubble constant based on gravitational time delays determined from observations of light curves in lens images of the same supernova (Grillo et al., 2018; Vega-Ferrero et al., 2018). Strongly magnified supernova iPTF16geu lensed by a single galaxy was observed as a transient source in the intermediate Palomar Transient Factory (Goobar et al., 2017), although the corresponding arrival time delays were too short to obtain a meaningful estimate of H_0. The third multiply-imaged supernova AT2016jka ("SN Requiem") was discovered in Hubble Space Telescope data, and is predicted to host a fourth image in two decades time, which should allow for a sub-per cent precision measurement of the time delay (S. Rodney et al., 2021). The three detected lensed supernovae are depicted in Figure 1.18.

Compared to lensed quasars, the advantages of lensed supernovae lie in their well-defined light curves and the possibility for follow-up observations to constrain the lens galaxy properties after the supernova has faded away. In the case of lensed type Ia supernovae, absolute magnifications provide another promising way to break the mass-sheet degeneracy (Birrer et al., 2021; Foxley-Marrable et al., 2018; Kolatt & Bartelmann, 1998; Oguri & Kawano, 2003). Challenges associated with lensed supernovae are that they typically have much shorter time delays and smaller image separations than quasars, which make it difficult to precisely constrain their time delays and renders them more susceptible to microlensing effects from substructures in the lens galaxy (Dobler & Keeton, 2006; Goldstein et al., 2018; Huber et al., 2019).

In Chapter 5, we present an AI-driven pipeline to distinguish between lensed and unlensed supernovae in survey images. In order to employ the confirmed lensed supernovae for cosmological inference, Chapter 6 presents a spatio-temporal neural network that infers the cosmic expansion rate from an ensemble of time-series images of simulated lensed supernovae.

1.5 COSMOLOGICAL TENSIONS

THE PREVIOUS sections described how CMB-based measurements by *Planck* infer a low value of the Hubble constant ($H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1}; Planck Collaboration, Aghanim, et al., 2018), while low-redshift observations of type Ia supernovae calibrated by Cepheids obtain a high Hubble constant ($H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1}; Riess et al., 2021). The low-redshift calibration by lensed quasars as carried out by the HoLiCOW collaboration supports this conclusion ($H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1} Mpc^{-1}; Wong et al., 2020). The difference in H_0 values inferred from CMB-based methods and low-redshift observations is one of the most outstanding problems in present-day cosmology, and is often referred to as the *Hubble tension*. It might be explained by residual systematics in either of the approaches, or it could hint at new physics beyond the standard ΛCDM model. The CMB-based measurements rely on ΛCDM to propagate the parameters at recombination forward to redshift zero, or in other words, to convert distance measurements into the Hubble constant. Hence, modifying the underlying cosmological model used in the CMB inference might return a higher H_0 value that is consistent with SHoES and HoLiCOW.

Nevertheless, the tension is lowered considerably by several recent low-redshift measurements, such as distance calibrations from the Tip of the Red Giant Branch (TRGB), as measured by the Carnegie-Chicago Hubble Project ($H_0 = 69.8 \pm 0.6$ (stat) ± 1.6 (sys) km s^{-1} Mpc^{-1}; Freedman, 2021) and the new analysis of seven gravitationally lensed quasars by the TDCOSMO collaboration ($67.4^{+4.1}_{-3.2}$ km s^{-1} Mpc^{-1}; Birrer et al., 2020). An overview of several recent H_0 measurements is provided in Figure 1.19.

1.5.1 Alternative interpretations of the Hubble tension

Although the tension is generally discussed in terms of the Hubble constant, it can be more broadly regarded as a disagreement in distance calibrations, which can be recast as a tension in a number of cosmological parameters. Firstly, the discrepancy in distance measurements also influences the predictions of the sound horizon. Since BAO measurements constrain the product of $H_0 \times r_\mathrm{d}$ (see Section 1.4.2), a higher value of the Hubble constant corresponds automatically to a lower sound horizon, and vice versa. In our work in Chapters 3 and 4, we focus on the combined tension in the $H_0 - r_\mathrm{d}$ plane, and show that this yields valuable information about the validity of some proposed solutions to the tension.

Since the age of the Universe t_0 is inversely proportional to the Hubble constant;

$$t_0 = \int_0^\infty \frac{\mathrm{d}z}{(1+z)H(z)}, \tag{1.82}$$

the problem can also be regarded as a t_0 tension (Bernal et al., 2021). H_0 measurements by SHoES predict a shorter age of the Universe than observations by *Planck*. Recent constraints from the age of the oldest globular clusters (Valcin et al., 2020) in combination with low-redshift Ω_m measurements within a ΛCDM cosmology yield an estimate of $H_0 = 71 \pm 2.8$ km s^{-1} Mpc^{-1} (Jimenez et al., 2019).

Additionally, there is a partial degeneracy between the Hubble constant and the CMB monopole temperature T_0. When combining data from *Planck* and SHoES while allowing T_0 to vary within a ΛCDM framework, the Hubble tension can be resolved with a corresponding $T_0 = 2.56 \pm 0.05$ (Ivanov et al., 2020). However, this measurement is in tension with temperature measurements by FIRAS (Fixsen, 2009) and temperature estimates from the Sunyaev–Zel'dovich effect (Luzzi et al., 2015).

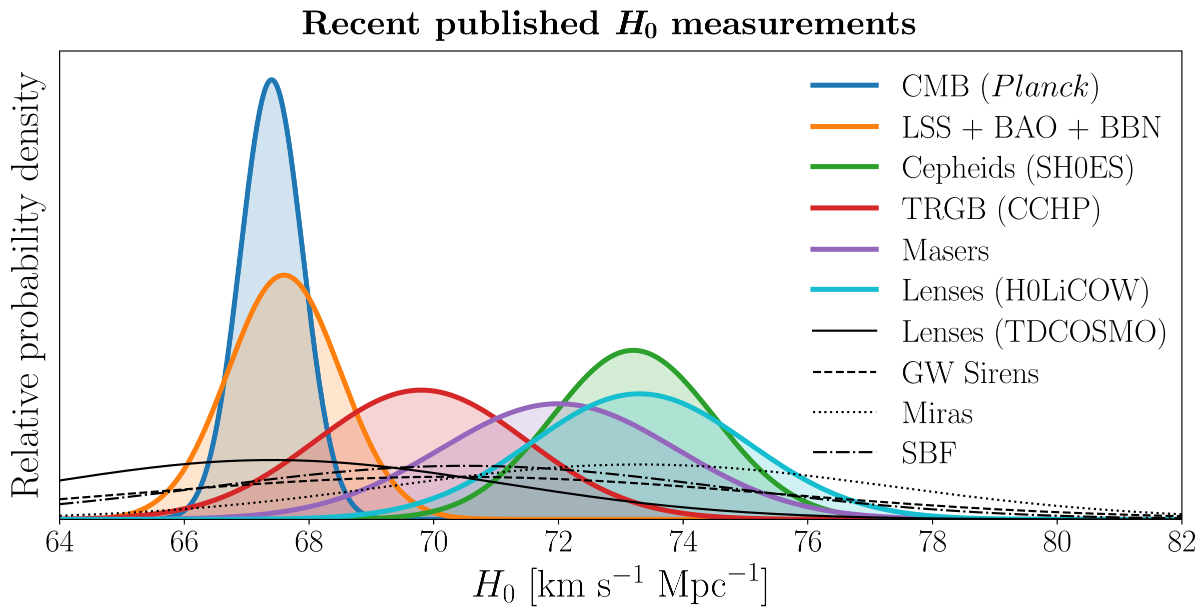


Figure 1.19 – An overview of several recent published H_0 probability distributions, demonstrating the discrepancy between different measurement methods. High-precision constraints on the Hubble constant are provided by observations of the CMB (Planck Collaboration, Aghanim, et al., 2018), LSS + BAO + BBN (DES Collaboration et al., 2021), Cepheids (Riess et al., 2021), TRGB (Freedman, 2021), water masers (M. J. Reid et al., 2019), and lensed quasars (Wong et al., 2020). Additionally, several lower-precision measurements have been carried out using lensed quasars (Birrer et al., 2020), gravitational wave (GW) sirens (Hotokezaka et al., 2019), miras (C. D. Huang et al., 2020), and surface brightness fluctuations (SBF; Khetan et al., 2021).

Furthermore, in the case of the SHoES analysis, the H_0 tension should actually be regarded as a discrepancy in the type Ia supernovae absolute peak magnitude M_B (Benevento et al., 2020; Camarena & Marra, 2021; Efstathiou, 2021). Since SHoES uses a cosmographic expansion to extrapolate the expansion rate from supernovae in the redshift range $0.023 \leq z \leq 0.15$ to redshift zero, their analysis does not directly measure H_0. The true measured quantity is $M_B = -19.244 \pm 0.037$ magnitudes (Camarena & Marra, 2021), which is in tension with the inferred absolute magnitude from CMB observations, $M_B = -19.401 \pm 0.027$ (Camarena & Marra, 2020a).

Finally, the tension is sometimes described as a disagreement between the early (CMB) and late (low-redshift) Universe. This view hints at a possible solution to the tension by means of an alternative cosmological model to connect the early-time measurements to the late-time observations. An extensive series of different models have been proposed for this purpose (see e.g. Di Valentino et al., 2021; Knox & Millea, 2020; Shah et al., 2021). In our work, we investigate the effects of modifications to the standard ΛCDM model. Solutions in the form of extensions to ΛCDM can follow two distinct approaches: they can either change the early physics (pre-recombination) or they can modify the late physics (post-recombination). Both classes of models are discussed below.

1.5.2 Early-time modifications of ΛCDM

Early-time extensions change the physics before recombination by introducing an additional component that contributes to the energy density. Some commonly used examples are additional relativistic particles or early dark energy, as will be discussed in more detail in Chapter 4. Effectively, this increases the expansion rate in the early Universe, thereby shortening the time between the Big Bang and recombination. As a consequence, the maximum distance which the sound waves can travel before decoupling is shorter and, therefore, the physical size of the sound horizon is reduced.

The angular size of the sound horizon is imprinted in the position of the first acoustic peak of the CMB power spectrum, as discussed in Section 1.2.3. In order to keep this observed angular scale unchanged, the decrease in the physical size of r_{d} automatically implies a shorter angular diameter distance to the CMB. Shorter distances in turn correspond to a higher value of the Hubble constant. The core idea behind changing the early-time physics is illustrated in Figure 1.20.

Early-time modifications simultaneously entail an increase in H_0 and a decrease in r_{d}, thereby preserving a good fit to BAO observations. However, the increased expansion rate in the early Universe suppresses the growth of perturbations, which needs to be compensated with a higher physical dark matter content. This leads to an increased amplitude of density fluctuations; a higher value of σ_8. These changes exacerbate the weak growth tension (cf. Section 1.3.5) between CMB and LSS measurements (Hill et al., 2020). In fact, it has been pointed out by Jedamzik et al. (2021) that any solution which only reduces the sound horizon can never fully resolve the Hubble tension while still maintaining agreement with other cosmological data sets, such as BAO or LSS.

1.5.3 Late-time modifications of ΛCDM

The second class of models leaves the physics before recombination unaltered. Instead, the cosmic expansion history is modified at later times, with a decreased value at intermediate redshifts and a higher value at low redshifts, as illustrated in Figure 1.20. Generally, this is achieved by allowing the dark energy density to increase over time. In the standard ΛCDM scenario, dark energy is a property of the vacuum with equation of state parameter $w = -1$, i.e. its density remains constant as the Universe expands. Modifying the equation of state parameter to be less than -1 will cause the dark energy density to increase, eventually leading to the “Big Rip” scenario in which cosmic expansion will tear everything in the Universe apart[2]. The advantage of this scenario is that it solves the H_0 tension by increasing the present-day expansion rate. However, as we show in Chapter 4, since late-time extensions do not change the physics in the early Universe, they are unable to alter the value of the sound horizon, thereby inducing a tension with BAO and type Ia supernova observations (Arendse et al., 2020; Knox &

[2]We do not have to worry about this scenario just yet. In light of the current *Planck* 2018 data, the Big Rip is not likely to happen for another 200 billion years (Mack, 2020).

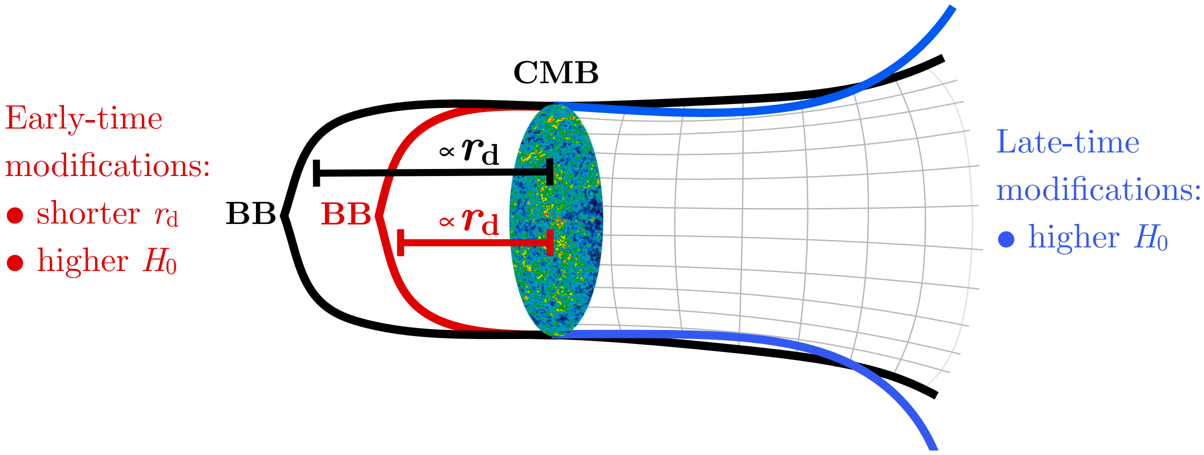


FIGURE 1.20 – Expansion history of the Universe as predicted by the standard ΛCDM model (black), early-time modifications (red) and late-time modifications (blue). The horizontal axis represents the time from the Big Bang to the present, from left to right. The length scale indicated in the figure is proportional to the sound horizon and corresponds to r_d/c_s, the time during which the sound waves have propagated. Early-time modifications result in a shorter r_d and higher H_0, while late-time modifications only produce a higher H_0.

Millea, 2020). Additionally, it has recently been demonstrated by Alestas and Perivolaropoulos (2021) that late-time modifications also worsen the growth tension.

If we take into account constraints from all cosmological data sets (e.g. type Ia supernovae, BAO, CMB, LSS, globular clusters), we seem to be in an over-constrained system, where fixing one problem inevitably creates another one. Anchordoqui et al. (2021) show that even when combining the most promising early and late-time modifications, a simultaneous solution to the Hubble and the growth tensions appears out of reach. The quest to devise a cosmological model that outperforms ΛCDM and manages to resolve the tensions in all cosmological parameters of interest remains open.

CHAPTER 2

Statistical inference and Machine Learning

"All models are wrong, but some are useful"
- George E. P. Box

In this chapter, we provide details about the statistical inference and machine learning methods that are used throughout the thesis. Section 2.1 presents the Bayesian interpretation of probability and Bayesian inference as a tool for cosmological data analysis. Sections 2.2 and 2.3 describe how machine learning models, more specifically neural networks, can provide estimates of cosmological parameters along with reliable uncertainties.

2.1 Bayesian statistics

The Bayesian interpretation of statistics expresses probability as a measure of the degree of belief in an event, which can be influenced by prior knowledge about the event (Bayes & Price, 1763). This differs from the frequentist interpretation, where probability is measured as the relative frequency of occurrence in the limit of infinite trials, independent of any prior knowledge (Heavens, 2009). Therefore, Bayesian methods constitute a more principled approach to work with incomplete information (Jaynes, 1957).

2.1.1 Bayes' theorem and its applications

Bayes' identity can be derived from the definition of conditional probability. If A and B represent two events, the conditional probability $\mathcal{P}(A \mid B)$ that A occurs given the occurrence of B, is expressed as:

$$\mathcal{P}(A \mid B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(B)}, \tag{2.1}$$

where $\mathcal{P}(A \cap B)$ refers to the intersection of A and B, i.e. the probability that *both* A and B occur. Analogously, the conditional probability $\mathcal{P}(B \mid A)$ that B occurs given A can be written in a similar way. Combining and rearranging the equations for $\mathcal{P}(A \mid B)$ and $\mathcal{P}(B \mid A)$ result in Bayes' theorem (Bayes & Price, 1763):

$$\mathcal{P}(A \mid B) = \frac{\mathcal{P}(B \mid A)\mathcal{P}(A)}{\mathcal{P}(B)}. \tag{2.2}$$

In statistical inference, A generally corresponds to a theory and B to the data, in which case each term of eq. 2.2 carries a distinct meaning:

- $\mathcal{P}(A \mid B)$ is the *posterior* probability distribution. The posterior characterises how probable the theory is in light of the data, which is usually the quantity of interest.
- $\mathcal{P}(B \mid A)$ constitutes the *likelihood*; the probability of the data under the assumption of the theory.
- $\mathcal{P}(A)$ is the *prior* probability distribution, the probability of the theory in absence of the data, which represents our prior knowledge of the theory.
- $\mathcal{P}(B)$ is the *evidence*; the probability of the data integrated over the full parameter space. Since the evidence is independent of the theory, it acts as a normalisation constant for the posterior distribution.

Bayes' theorem provides a means to solve inverse problems, such that if we know how the data arises from a theory, we can use the data to constrain the theory. Additionally, it tells us how to update our knowledge in light of new data.

In practice, if we want to constrain a physical model $\mathcal{M}$ with parameter vector Θ using a data set d, each parameter should have a corresponding prior $\mathcal{P}(\Theta \mid \mathcal{M})$ and likelihood $\mathcal{P}(d \mid \Theta, \mathcal{M})$. Since observations are often accompanied with Gaussian noise, the likelihood is generally taken to be a normal distribution. The choice of a suitable prior can be challenging, which is an occasional point of critique in Bayesian analysis. However, specifying a prior beforehand ensures that all assumptions are on the table, whereas assumptions in a frequentist analysis may not be as apparent. The resulting posterior distribution can be obtained from the likelihood and prior via Bayes' theorem (eq. (2.2)) as follows:

$$\mathcal{P}(\Theta \mid d, \mathcal{M}) \propto \mathcal{P}(d \mid \Theta, \mathcal{M})\, \mathcal{P}(\Theta \mid \mathcal{M}). \tag{2.3}$$

In the above formula, the evidence $\mathcal{P}(d|\mathcal{M})$ is omitted since it only functions as a normalisation constant for parameter inference problems. Nevertheless, the evidence is important for model comparison. The posterior odds ratio of model $\mathcal{M}_1$ compared to model $\mathcal{M}_2$ can be calculated from a ratio of the models' priors and evidences:

$$\frac{\mathcal{P}(\mathcal{M}_1 \mid d)}{\mathcal{P}(\mathcal{M}_2 \mid d)} = \frac{\mathcal{P}(\mathcal{M}_1)}{\mathcal{P}(\mathcal{M}_2)} \frac{\mathcal{P}(d \mid \mathcal{M}_1)}{\mathcal{P}(d \mid \mathcal{M}_2)}. \tag{2.4}$$

The computation of the evidence can be highly computationally expensive. For that reason, alternative model selection methods are often used that rely on approximations of the data distribution. One such method is the Bayesian Information Criterion (BIC), which is defined as:

$$\text{BIC} \;=\; k \ln(N) - 2\ln(\mathcal{L}_{m.a.p.}), \tag{2.5}$$

where N corresponds to the number of data points, k represents the number of free parameters and $\mathcal{L}_{m.a.p.}$ denotes the maximum a posteriori likelihood (i.e. evaluated where the posterior is maximised). The BIC score expresses how well a model describes the data, with a lower score corresponding to a better agreement. It also introduces a penalty term for added complexity in a model, thereby favouring models with fewer parameters that can describe the data sufficiently well. This philosophy is known as *Occam's razor*, and it states that when presented with competing theories that produce the same predictions, one should select the one with the fewest assumptions.

The ensemble of parameters Θ generally consist of several physically interesting parameters ζ and a number of nuisance parameters ξ. The posterior calculated in eq. (2.3) is the *joint* posterior of $\Theta = \{\zeta, \xi\}$. If we are instead interested in the *marginal* posterior for the parameters of interest, we can integrate out the nuisance parameters as follows:

$$\mathcal{P}(\zeta|d, \mathcal{M}) \propto \int \mathrm{d}\xi \; \mathcal{P}(d|\zeta, \xi, \mathcal{M})\mathcal{P}(\zeta, \xi|\mathcal{M}). \tag{2.6}$$

In summary, the posterior distribution $\mathcal{P}(\Theta \mid d, \mathcal{M})$ provides a measure of how likely a model $\mathcal{M}$ with parameter set Θ is in light of the data d. A powerful way to sample the posterior density is by means of Markov Chain Monte Carlo techniques, which are discussed in the following section.

2.1.2 Markov Chain Monte Carlo techniques

One of the challenges associated with statistical analysis is to sufficiently explore the full posterior distribution, especially in a high-dimensional parameter space. Often, the posterior is not analytically tractable and, instead, has to be approximated by numerical techniques (Leclercq et al., 2014). Monte Carlo (MC) methods provide such an approximation of the true posterior distribution by drawing a set of random samples from it. In this way, the true posterior $\mathcal{P}(\Theta \mid d, \mathcal{M})$ is estimated as a sum of N Dirac delta distributions $\mathcal{P}_N(\Theta \mid d)$:

$$\mathcal{P}(\Theta \mid d) \approx \mathcal{P}_N(\Theta \mid d) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\mathrm{D}}(\Theta - \Theta_i), \tag{2.7}$$

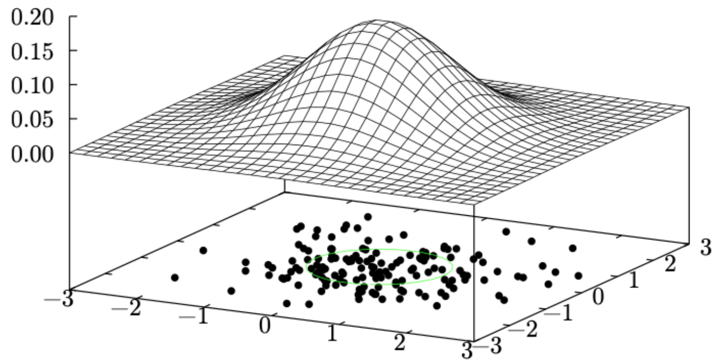


Figure 2.1 – Visualisation of a Monte Carlo method in two dimensions, which draws random samples (represented by the black dots) to approximate the true posterior distribution. The local density of the samples is proportional to the posterior distribution. Figure reprinted from Leclercq et al. (2014).

with $\delta_{\rm D}$ corresponding to the Dirac delta function. The local density of the resulting samples is proportional to the true posterior distribution, as visualised in Figure 2.1.

A highly efficient method of sampling the posterior distribution is the Markov Chain Monte Carlo (MCMC) technique, in which a sequence of points $\{X_1, X_2, .., X_n, ..\}$, known as a *Markov Chain*, is constructed in parameter space. The chain consists of random values, whereby the probability of element X_{n+1} depends only on element X_n. This correlation between subsequent samples allows the chain to navigate the probability landscape and find its way to the high-density region of the target distribution (see Figure 2.2). The initial phase corresponding to the low-density regions is known as the *burn-in*, which can be discarded after the chain has converged to the region of interest.

Several well-known MCMC algorithms include the Metropolis–Hastings algorithm (Hastings, 1970; Metropolis et al., 1953), Gibbs sampling (Smith & Roberts, 1993), and Hamiltonian Monte Carlo (Hanson, 2001). In the Metropolis-Hastings algorithm, the Markov Chain starts at an initial position X and a candidate for the next step X' is suggested by a proposal distribution $Q(X' \mid X)$. The probability that X' is accepted as the next step depends on the ratio of the final and initial target densities:

$$a = \frac{\mathcal{P}(X')}{\mathcal{P}(X)} \frac{Q(X \mid X')}{Q(X' \mid X)}, \tag{2.8}$$

where a is the *acceptance ratio*. When $a > 1$, the suggested position X' is in a higher-density region of the target distribution than the initial position X and will be accepted. If $a < 1$, the suggested step is accepted with probability a, even though it leads in a direction away from the highest density. This stochasticity of occasionally accepting a 'wrong' move helps the MCMC algorithm to escape regions of local extrema. If the proposed position X' is accepted, this will become the new state of the chain. Alternatively, the chain will remain at X and the above routine is repeated.

Figure 2.2 – Markov Chain Monte Carlo methods use a chain of steps to efficiently navigate the probability landscape and find their way to the density peak of the target distribution. Image credit: Murray Foubister.

In Chapters 3 and 4, an affine-invariant MCMC method is used to approximate the posteriors of cosmological parameters, such as the sound horizon and Hubble constant. This class of MCMC algorithms is unaffected by affine transformations of space, which makes it well-suited to explore skewed density distributions (Goodman & Weare, 2010b).

Drawbacks of MCMC algorithms are that the chains can be inefficient in exploring high-dimensional parameter spaces, such that the burn-in phase can take a long time. Additionally, for multi-modal posteriors it can be difficult for the chains to converge to the true maximum of the posterior. Therefore, it is important to ensure that the Markov Chains have sufficiently explored the relevant parameter space. Several methods to assess whether the MCMC chains have converged include:

- A visual representation of multiple chains starting at widely varied positions can be inspected to see if they converge to the same region of the parameter space.
- The proposed steps should have a sufficiently high acceptance rate.
- The chain should have a short mixing time, i.e. they should progress rapidly to the region of interest. Since the initial samples corresponding to the burn-in phase of the chain are correlated, this duration should be small in comparison to the total sampling time.
- The Gelman-Rubin diagnostic (Brooks & Gelman, 1998; Gelman & Rubin, 1992) provides a measure of convergence by comparing the between-chains and within-chains variances, which should be comparable in size for a chain that has reached convergence.

2.2 NEURAL NETWORKS

This section draws material from Arendse, Kodi Ramanah, and Wojtak (2021), Kodi Ramanah, Arendse, and Wojtak (2021), and Kodi Ramanah, Wojtak, and Arendse (2021).

ASTRONOMICAL data sets have grown remarkably in size and complexity over the last decade, making it increasingly difficult to extract meaningful features from the data. For this reason, machine learning (ML) algorithms have emerged as promising tools to handle large data sets and perform a wide variety of tasks (Baron, 2019), such as classification, regression, clustering, outlier detection, and time series analysis. ML methods have been been employed for an impressive array of cosmological applications, ranging from the estimation of cosmological parameters directly from the cosmic large-scale matter distribution (Ravanbakhsh et al., 2017), generation of mock halo catalogues (Berger & Stein, 2019; Bernardini et al., 2020), classification of the structures of the cosmic web (Aragon-Calvo, 2019), photometric classification of supernovae (e.g. Charnock & Moss, 2017; Lochner et al., 2016), identification of strong lensing arc features from gravitationally lensed systems (e.g. Lanusse et al., 2018), estimation of lensing properties from photometric images (e.g. Hezaveh et al., 2017; Park et al., 2021), estimation of the combined mass of the Milky Way and Andromeda (Lemos et al., 2021), to the dynamical mass inference of galaxy clusters (e.g. Ho et al., 2019; Kodi Ramanah, Wojtak, Ansari, et al., 2020; Ntampaka et al., 2016). Emulators based on deep generative modelling techniques have been developed to predict the formation of cosmic structures in the Universe (He et al., 2019), to map 3D dark matter fields to their halo count distributions (Kodi Ramanah et al., 2019), and to obtain the high-resolution version of low-resolution dark matter simulations (Kodi Ramanah, Charnock, Villaescusa-Navarro, et al., 2020; Y. Li et al., 2021).

In this thesis, various types of *neural networks* are used to address selected problems in cosmology. An artificial neural network aims to uncover relationships and patterns in data via a process that is inspired by the inner-workings of a brain. It consists of several nodes, called *artificial neurons*, that build a weighted sum of their inputs, add a constant term (*bias*), apply an activation function, and pass the output along. Several layers of nodes can be stacked to form a *deep neural network*, in which each layer's output constitutes the subsequent layer's input. A *loss function* operating on the last layer quantifies how well the network's output matches the target output associated with the input data. During the training process, the loss function is minimised by adjusting the weights via stochastic gradient descent, a process known as *back-propagation*. Once the training is completed, unseen data can be passed into the network to obtain output predictions.

Neural networks were already theorised in 1943 (Mcculloch & Pitts, 1943), but their success only took flight around 50 years later, with the increase in computational power and parallel computing via graphics processing units (GPUs). Between 2009 and 2012, various types of neural networks won eight international competitions in ML and pattern recognition, after which neural networks became the standard method amongst all top contestants (Molnar, 2019; Russakovsky et al., 2014). The advantages of neural networks are that they are able learn non-linear and complex relationships, build high levels of abstraction, work well on large data sets, and are robust under incomplete information. However, caution should be exercised with the use of neural networks for several reasons. Firstly, they only return point estimates of the desired quantity, without meaningful uncertainties. Secondly, for supervised methods where a training set is used, it is essential to ensure similarity between the training set and actual observations, in order to avoid biases in the predictions for real input data. This problem can be addressed with *transfer learning*, in which knowledge from one data set can be generalised to a second data set with different properties. Another challenge relates to the interpretability of the results obtained with neural networks. It can often be difficult to understand exactly what the model has learned, and what information it used to reach those conclusions. Since most algorithms simply optimise an internal loss function, which can in principle be misaligned with the scientific goal, the results do not always correspond to the desired outcome. Several recent advances aim to make neural networks more interpretable, such as the visualisation of saliency maps and learned features, modification of the training set, and adversarial examples (Molnar, 2019).

Formally, a neural network is an arbitrarily complex and flexible model, $\mathbb{NN}(\omega, \eta) : \boldsymbol{D} \rightarrow \tilde{\boldsymbol{d}}$, which maps some input data $\boldsymbol{D}$ to a prediction of the desired target $\tilde{\boldsymbol{d}}$ associated with the data. ω and η correspond to the trainable model parameters, known as *weights*, and hyperparameters (e.g. network architecture, weights initialisation, type of activation and loss functions), respectively. The weights are optimised during training to minimise the loss function, which is a differentiable function of the data and the weight, often taken to be the negative log-likelihood:

$$-\ln \mathcal{L}(\tilde{\boldsymbol{d}}|\boldsymbol{D}, \omega, \eta). \tag{2.9}$$

As such, the training process is equivalent to determining the maximum likelihood estimate of the weights, resulting purely in a single point estimate by the neural network. The remainder of this section presents two types of neural networks that have been used for the work presented in this thesis; convolutional neural networks and recurrent neural networks. Section 2.3 discusses two methods, variational inference and simulation-based inference, which provide a reliable way to quantify the uncertainties associated with the network predictions.

2.2.1 Convolutional neural networks

Convolutional neural networks (hereafter CNNs) (LeCun, Bengio, et al., 1995; LeCun et al., 1998) are a particular type of artificial neural network, especially suited for problems where spatially correlated information is crucial. In essence, a CNN is designed as follows: A convolutional kernel of a given size, commonly referred to as a *filter*, is applied to each pixel (or voxel for 3D inputs) of the input image and its vicinity as it scans through the whole image. A given pixel in the output is only a function of the pixels in the input which are enclosed within the window defined by the kernel, known as the *receptive field* of the pixel. The output of a such a convolution on an image, called a *feature map*, contains high values in the pixels which match the pattern (or feature) encoded in the weights and biases of the corresponding kernel. These weights and biases are the trainable parameters that are optimised during training.

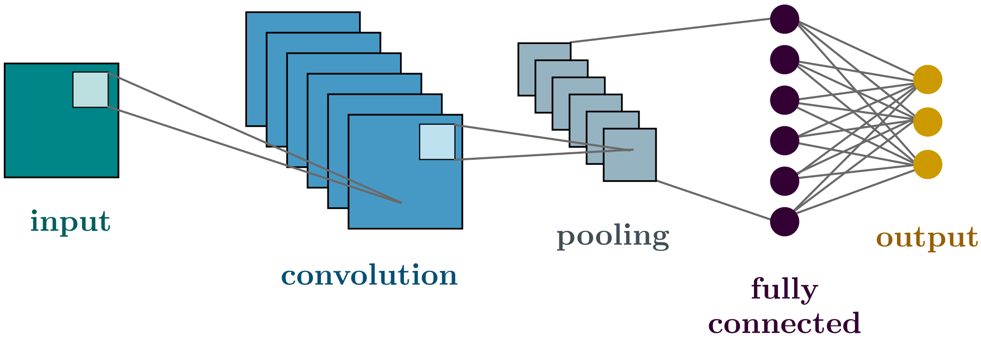


FIGURE 2.3 – A schematic representation of a simple convolutional neural network (CNN). The input image is fed to a convolutional layer, where a kernel scans over the whole image to produce a feature map. Subsequently, a pooling step is applied in order to downsample the data, after which it is flattened and passed along to a fully connected layer, eventually yielding an output.

To extract the series of distinct features of the input image, a convolutional layer generally employs several filters, resulting in a set of feature maps which are then fed as inputs to the subsequent layer. This convolutional operation is typically followed by a *pooling* layer as a subsampling or dimensionality reduction step (Goodfellow et al., 2016). The application of these two types of layers will reduce the initial input image to a compact representation of features, which can be reshaped as a vector. This feature vector is subsequently passed to the final layer, a *fully connected layer* in which every neuron receives input from every neuron of the preceding layer, to ultimately generate an output (LeCun et al., 2015). The architecture of a standard CNN is displayed in Figure 2.3.

In terms of the mathematical formalism, the convolutional operation may be described as a specialised linear operation, with the discrete convolution implemented via matrix multiplication. As such, a particular convolutional layer, denoted by ℓ, can be computed using

$$x_j^{\ell} = \mathcal{F}\left(\sum_{i\in\mathcal{M}_j} x_i^{\ell-1} \times k_{ij}^{\ell} + b_j^{\ell}\right), \tag{2.10}$$

where $\mathcal{F}$ denotes the activation function, k represents the convolutional kernel, $\mathcal{M}_j$ corresponds to the receptive field and b is the bias parameter (Goodfellow et al., 2016). The role of the activation function is to encode some non-linearity in the convolutional layers, so that a stack of such layers can be used as a generic function approximator. A commonly used activation function is the rectified linear unit (ReLU) activation function (Nair & Hinton, 2010), defined as follows:

$$\mathcal{F}(z_i) = \begin{cases} 0, & z_i < 0 \\ z_i, & z_i \geq 0. \end{cases} \tag{2.11}$$

The ReLU activation and its variants are less computationally expensive than other common activation functions, such as the sigmoid and hyperbolic tangent (tanh) functions, and mitigate the vanishing gradient issue in training deep neural networks (Glorot et al., 2011). The latter predicament arises when the gradient tends to zero due to a saturating activation function towards $f(z_i) \approx 0$ or $f(z_i) \approx 1$, as in the case of the sigmoid and tanh functions. This is inevitably detrimental to the effectiveness of gradient descent during training, resulting in poor performance.

The above described components allow a CNN to extract meaningful spatial features from the input image. By stacking several convolutional layers, the network is capable of building an internal hierarchical representation of features encoding the most relevant information from the input image, such that the network is able to identify increasingly complex patterns with the addition of more layers. Hence, convolutional layers provide a natural approach to take spatial context into consideration. A key aspect of such networks is that a stack of convolutional layers increases the sensitivity of subsequent layers to features on increasingly larger scales. In other words, the size of the receptive field becomes larger as we go deeper in the network. Moreover, convolutional layers retain the local information while performing the convolution on adjacent pixels, thereby allowing both local and global information to propagate through the network (LeCun et al., 2015).

CNNs are well-suited to tackle a multitude of problems in astronomy that involve visual data. In this thesis, they are employed to extract information from images of lensed supernovae (Chapters 5 and 6) and from the dynamical phase-space distribution of galaxy clusters (Chapter 7).

2.2.2 Recurrent neural networks

Recurrent neural networks (Lipton et al., 2015) (or RNNs) consist of collections of neurons, where the output of each neuron collection in a given layer is not only connected to the next layer but is also fed as input to itself. The temporal behaviour of such an architecture renders RNNs suitable for learning sequential models or for modelling time series. Long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) networks are a specific type of RNNs, which build upon these models by encoding a new element in the neurons, known as the *memory cell* (c_t). The role of the latter is to accumulate the information from the previous inputs x_{t-1}, thereby allowing the neuron to recall the information that it has already processed. By relying on a gating mechanism, the memory cell decides how much of the new input x_t to accumulate in the short-term memory c_t, and what it should forget from the prior state c_{t-1} and propagate c_t to the *hidden state* h_t, thereby resembling a long-term memory. LSTM is implemented via the following equations:

$$
\begin{aligned}
\text{Input gate} \quad & i_t = \sigma(W_i x_t + b_{ii} + U_i h_{t-1} + b_{hi}) && (2.12)\\
\text{Forget gate} \quad & f_t = \sigma(W_f x_t + b_{if} + U_f h_{t-1} + b_{hf}) && (2.13)\\
\text{Cell gate} \quad & g_t = \tanh(W_g x_t + b_{ig} + U_g h_{t-1} + b_{hg}) && (2.14)\\
\text{Output gate} \quad & o_t = \sigma(W_o x_t + b_{io} + U_o h_{t-1} + b_{ho}) && (2.15)\\
\text{Cell state} \quad & c_t = f_t \times c_{t-1} + i_t \times g_t && (2.16)\\
\text{Hidden state} \quad & h_t = o_t \times \tanh(c_t), && (2.17)
\end{aligned}
$$

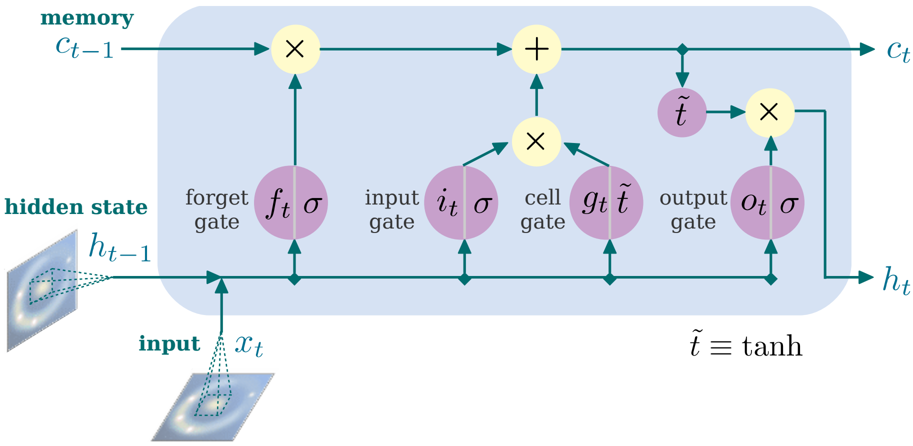


FIGURE 2.4 – Schematic representation of the ConvLSTM cell, illustrating the mechanism for updating the cell's memory (c_t) and hidden (h_t) states for a single time step. Figure from Kodi Ramanah, Arendse, et al. (2021).

where x_t is the input at time t, σ is the recurrent (sigmoid) activation function, W and U are the trainable weights, while b indicates the bias terms and $\times$ denotes element-wise product. h_{t-1} is the hidden state at the prior time step or its initialised state at $t = 0$, and similarly for c_{t-1}. In essence, h_t and c_t encapsulate the short (fast) and long (slow) correlations in the data, respectively. This dichotomy alleviates the issue of vanishing or exploding gradients that plague standard RNNs.

Convolutional LSTM (hereafter ConvLSTM) (Shi et al., 2015) is an extension of the LSTM network tailored to exploit spatially correlated features, in addition to temporal correlations, in data. The main change is that the 1D vectors representing the inputs, cell states and gates in eqs. (2.12)–(2.17) are now 2D tensors embedded in an LSTM network, and the products with the weight matrices W and U now represent convolutions over spatial dimensions, with the updates of the cell's states proceeding as before. A schematic representation of the ConvLSTM cell is displayed in Figure 2.4.

Some common applications of RNNs involve speech recognition, language translation and video tagging. In the field of astrophysics, the ConvLSTM model was recently used to infer galaxy properties from 21 cm lightcone images (Prelogović et al., 2021), thereby showing that the ConvLSTM network outperforms a standard 3D CNN, where the third dimension incorporates the temporal evolution, for this particular problem. In Chapter 5, we employ a ConvLSTM model for the identification of gravitationally lensed supernovae from time-series images obtained from multi-epoch observations.

2.3 UNCERTAINTY ESTIMATION FOR NEURAL NETWORKS

This section draws material from Arendse, Kodi Ramanah, and Wojtak (2021), Kodi Ramanah, Arendse, and Wojtak (2021), and Kodi Ramanah, Wojtak, and Arendse (2021).

Ideally, the neural network's point estimate of the weights should correspond to the global minimum on the likelihood surface. However, in practice this is generally not the case, as the likelihood surface is typically extremely complex, degenerate and non-convex. Therefore, it is much more likely that the weights will only converge to a local minimum on the likelihood surface, which is dictated to some extent by the initialisation of the weights. This is one of the major drawbacks of using neural networks, because it means that the network point estimates never truly correspond to the desired target $\tilde{\boldsymbol{d}}$. The techniques discussed in this section provide a way to mitigate this crucial limitation and to obtain scientifically accurate predictions, including reliable uncertainties, of the true targets $\tilde{\boldsymbol{d}}$ given the input data $\boldsymbol{D}$.

Uncertainties of neural networks can be characterised as either *aleatoric* or *epistemic*, encompassing the uncertainties in the data and model, respectively (Kendall & Gal, 2017). The former describes the inherent uncertainty by virtue of the random nature of the observations of interest, while the latter refers to the limitations of the model to accurately describe the set of observations. In the case that a neural network constitutes the model, epistemic uncertainties typically arise from insufficient network depth, flexibility, training epochs, data samples or suboptimal network architecture.

2.3.1 Variational inference

To overcome the limitation of point estimates for the weights and quantify the uncertainty associated with the model weights, neural networks combined with variational inference, commonly designated as *Bayesian neural networks* (Charnock et al., 2020; MacKay, 1992; Neal, 2012), cast the model parameters as probability distributions. Subsequently, they marginalise the network's output over these distributions within a Bayesian statistical framework to yield a posterior distribution. Hence, a trained Bayesian neural network represents an ensemble of networks, which allows the uncertainty on a specific prediction to be quantified.

The network output distribution captures, to some extent, a measure of the aleatoric uncertainties due to the noise intrinsic to the input data set. To more adequately account for this source of uncertainties, it must be ensured that the training data set is representative of the distribution of possible observations.

To account for the epistemic uncertainties associated with the choice of neural network's weights, the most common approach is to replace each network weight by a parametrised distribution to eventually infer the posterior distribution of the weights conditional on the input data during training via Bayes identity (eq. (2.2)). However, in practice, the posterior $\mathcal{P}(\omega|\mathcal{D})$ is intractable. Variational inference is an approach in which the true weight distribution is approximated by a variational distribution $q_\theta(\omega)$, where θ parameterises an ensemble of distributions. The training objective is then to ensure that the variational distribution $q_\theta(\omega)$ matches the posterior distribution $\mathcal{P}(\omega|\mathcal{D})$ as closely as possible, which can be achieved by minimising the metric distance (i.e. Kullback-Leibler (KL) divergence; Kullback & Leibler, 1951) as a quantitative measure of similarity:

$$\mathrm{KL}[q_\theta(\omega)\|\mathcal{P}(\omega|\mathcal{D})] \equiv \int q_\theta(\omega)\log\left[\frac{q_\theta(\omega)}{\mathcal{P}(\omega|\mathcal{D})}\right]\mathrm{d}\omega. \tag{2.18}$$

Using Bayes identity, the above loss function can be expressed in terms of the likelihood $\mathcal{L}(\mathcal{D}|\omega)$ and the prior distribution $\mathcal{P}(\omega)$ for the network weights (Blundell et al., 2015):

$$\mathrm{KL}[q_\theta(\omega)\|\mathcal{P}(\omega|\mathcal{D})] = \mathrm{KL}[q_\theta(\omega)\|\mathcal{P}(\omega)] - \mathbb{E}_{q_\theta(\omega)}[\log\mathcal{L}(\mathcal{D}|\omega)] + K, \tag{2.19}$$

where $\mathbb{E}_{q_\theta(\omega)}$ denotes the expectation under the variational distribution, with the constant K resulting from the Bayesian evidence and the second term being the standard negative log-likelihood. In simpler terms, variational inference implies the assumption of the form of the posterior distribution of the weights and the use of an optimisation routine to find the assumed probability distribution that is closest to the true posterior. This assumption simplifies the computation, resulting in some level of tractability.

Gal and Ghahramani (2015a) showed that when a Bernoulli distribution is assumed as the variational distribution, this can be approximated by performing dropout regularisation (Srivastava et al., 2014). Dropout is a technique whereby a fraction of model weights are randomly set to zero during each training step, effectively lowering the number of model parameters to ultimately mitigate risks of overfitting. Traditionally, dropout is deactivated at inference time, such that network predictions are deterministic. However, it was subsequently shown that when dropout is active during the evaluation phase, it is equivalent to implementing approximate Bayesian inference (Gal & Ghahramani, 2015b), whereby multiple model evaluations can be used to perform Monte Carlo integration of the posterior distribution. Under this formulation, it can be shown that including an ℓ_2 regularisation term, i.e. having Gaussian priors over the network weights, approximately entails computing the KL divergence with respect to an implicit prior (Gal & Ghahramani, 2015a). Consequently, training a neural network with dropout masks and ℓ_2 weight regularisation minimises the loss from eq. (2.19), such that it is feasible to perform proper variational inference without intensifying the computational workload (Charnock et al., 2020).

2.3.2 Simulation-based inference

Simulation-based inference (hereafter SBI), sometimes referred to as "likelihood-free inference", comprises a class of inference techniques relying on a simulator whose task is to generate high-fidelity simulations that can eventually be compared with observations (see, for e.g., Cranmer et al., 2019, for an in-depth review of recent developments). SBI techniques typically involve an estimation of the likelihood or posterior via informative summary statistics using classical density estimators (Diggle & Gratton, 1984) or neural density estimators (e.g. Germain et al., 2015; C.-W. Huang et al., 2018; Jimenez Rezende & Mohamed, 2015; Kingma et al., 2016;

Papamakarios & Murray, 2016; Papamakarios et al., 2017; Papamakarios et al., 2018; Uria et al., 2016) from recent advances in ML. These density estimators are used to approximate the distribution of summary statistics of the samples generated from the simulator. The obtained posterior encompasses the combined aleatoric and epistemic uncertainties to a reasonable extent.

The SBI framework adopted in this thesis is inspired by the approach presented in Charnock et al. (2018), where they demonstrate that parameter inference is feasible via SBI using summary statistics provided by a neural network. They developed an information maximising neural network to produce optimal summary statistics, but the approach presented therein may be employed with any neural network predicted summaries. This is because a neural network, by design, performs some form of data compression (or dimensionality reduction) to extract meaningful features from a given input data set $\boldsymbol{D}$ to yield informative summaries $\tilde{\boldsymbol{d}}$ of the data. Thereafter, a standard density estimator, such as a Gaussian kernel density estimator (KDE), or a neural density estimator (Germain et al., 2015; Papamakarios et al., 2017) can be used to approximate the desired likelihood or posterior from the distribution of network predicted summaries.

For a neural network that outputs a summary $\tilde{d}$, corresponding to a ground truth d, such an implementation of the SBI pipeline may be summarised via the following steps:

- A physical model, $\mathcal{F}(\boldsymbol{\phi}) : \boldsymbol{\Theta} \rightarrow \boldsymbol{D}$, is used to generate a training data set $\boldsymbol{D}$ from a set of model parameters $\boldsymbol{\Theta}$ and some initial conditions $\boldsymbol{\phi}$.
- A neural network is trained on the training set to obtain the desired summary $\tilde{d}$ from the input data.
- A separate test set, also generated with $\mathcal{F}(\boldsymbol{\phi})$, is fed to the trained neural network to obtain the corresponding summary estimates $\tilde{d}$.
- A standard or neural density estimator is used to compute the joint probability distribution of the ground truth d and summary prediction $\tilde{d}$ pairs.
- A slice through the above distribution at the network summary prediction $\tilde{d}_{\text{obs}}$, for a given input observation $\boldsymbol{D}_{\text{obs}}$, yields the approximate posterior predictive distribution:

$$\mathcal{P}(d|\tilde{d}_{\text{obs}}) \approx \mathcal{P}(d|\boldsymbol{D}_{\text{obs}}, \boldsymbol{\omega}, \boldsymbol{\eta}). \tag{2.20}$$

The above approach, which is visualised in Figure 2.5, has several key advantages. Although the posterior is conditional on the (trained) network weights and choice of hyperparameters, the SBI framework provides us with a posterior of the parameter of interest with statistically consistent uncertainties (Fluri et al., 2021; Makinen et al., 2021). If the performance and efficacy of the neural network are sub-optimal, as a result of not converging to the global optimal solution, the uncertainties on the network predictions will be inflated but not underestimated. Moreover, the density estimator can be precomputed, such that any slice through the likelihood can be computed nearly instantaneously to yield the desired approximate posterior for any given observation.

The main caveats of this framework are that there is a choice of density estimator with some hyperparameters, such as the bandwidth for the Gaussian KDE, and as for all SBI approaches,

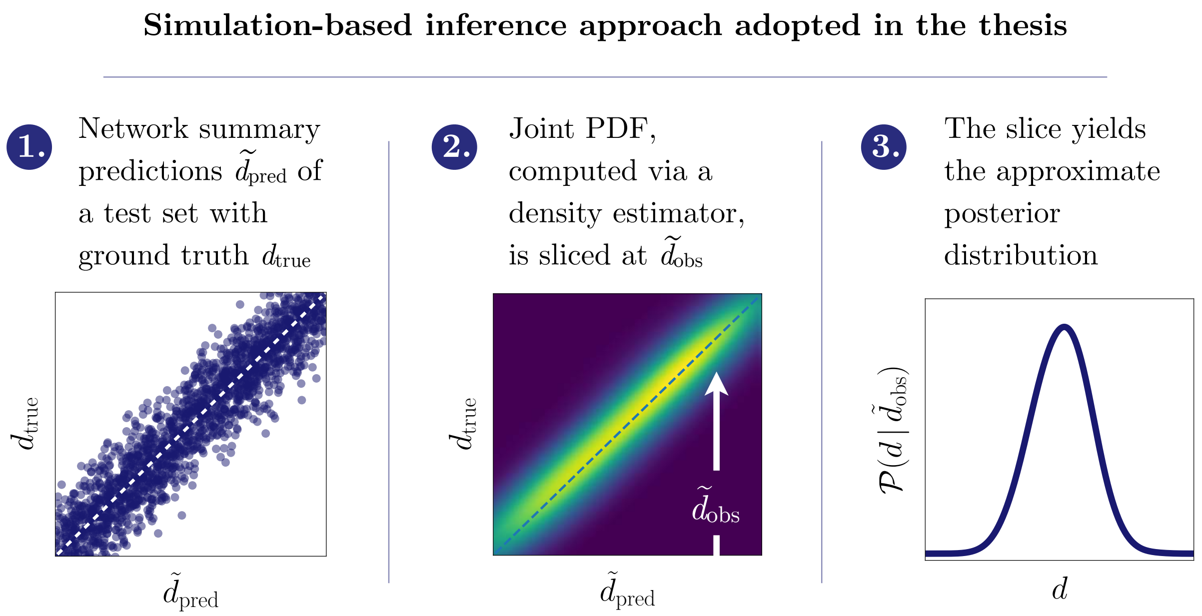


FIGURE 2.5 – Illustration of the simulation-based inference framework adopted throughout the thesis. Network predictions for a test set ($\tilde{d}_{\mathrm{pred}}$) are compared to the ground truth (d_{true}) and fed to a density estimator to obtain a joint 2D PDF. The posterior distribution of a new network prediction ($\tilde{d}_{\mathrm{obs}}$) can be obtained by slicing the joint PDF at $\tilde{d}_{\mathrm{obs}}$.

there is some dependence on the total number of simulations used to compute an approximation of the likelihood. Nevertheless, the Gaussian KDE is a fairly robust option as it is not very sensitive to the choice of hyperparameters. In contrast, sophisticated neural density estimators would require further (unsupervised) training and hyperparameter tuning, and would be prone to the shortcoming related to the training of conventional neural networks, i.e. convergence to local minima on the likelihood surface.

I

The Hubble constant & sound horizon tension

CHAPTER 3

Cosmology-independent measurements of the sound horizon

This chapter is based on the following article:
"Low-redshift measurement of the sound horizon through gravitational time-delays"

Published in Astronomy & Astrophysics (A&A), 632, A91 (2019).

Authors: Nikki Arendse, Adriano Agnello & Radosław J. Wojtak.

Abstract

The matter sound horizon can be inferred from the cosmic microwave background within the standard cosmological model. Independent direct measurements of the sound horizon are then a probe of possible deviations from the standard model. We aim at measuring the sound horizon r_d from low-redshift indicators, which are completely independent of CMB inference. We used the measured product $H(z)r_\mathrm{d}$ from baryon acoustic oscillations (BAO) together with supernovae Ia to constrain $H(z)/H_0$ and time-delay lenses analysed by the HoLiCOW collaboration to anchor cosmological distances ($\propto H_0^{-1}$). Additionally, we investigated the influence of adding a sample of quasars with higher redshift with standardisable UV-Xray luminosity distances. We adopted polynomial expansions in $H(z)$ or in comoving distances so that our inference was completely independent of any cosmological model on which the expansion history might be based. Our measurements are independent of Cepheids and systematics from peculiar motions

to within percent-level accuracy. The inferred sound horizon r_d varies between (133 ± 8) Mpc and (138 ± 5) Mpc across different models. The discrepancy with CMB measurements is robust against model choice. Statistical uncertainties are comparable to systematics. The combination of time-delay lenses, supernovae, and BAO yields a distance ladder that is independent of cosmology (and of Cepheid calibration) and a measurement of r_d that is independent of the CMB. These cosmographic measurements are then a competitive test of the standard model, regardless of the hypotheses on which the cosmology is based.

3.1 INTRODUCTION

THE SOUND HORIZON is a fundamental scale that is set by the physics of the early Universe and is imprinted on the clustering of dark and luminous matter of the Universe. The most precise measurements of the sound horizon are obtained from observations of the acoustic peaks in the power spectrum of the cosmic microwave background (CMB) radiation, although the inference partially depends on the underlying cosmological model. In particular, the recent *Planck* satellite mission yielded a sound horizon scale (at the end of the baryonic drag epoch) of $r_\mathrm{d} = 147.09 \pm 0.26$ Mpc. This was based on the spatially flat six-parameter ΛCDM model, which provides a satisfactory fit to all measured properties of the CMB (Planck Collaboration, Aghanim, et al., 2018), and on the standard model of particle physics.

The sound horizon remains fixed in the comoving coordinates since the last scattering epoch and its signature can be observed at low redshifts as an enhanced clustering of galaxies. This feature is referred to as baryon acoustic oscillations (BAO). When we assume that the sound horizon is calibrated by the CMB, BAO observations can be used to measure distances and the Hubble parameter at the corresponding redshifts. The resulting BAO constraints can then be extrapolated to $z = 0$, for instance, using type Ia supernovae (SNe), in order to determine the present-day expansion rate H_0. However, this inverse distance ladder procedure depends on the choice of cosmological model and on the strong assumption that the current standard cosmological model provides an accurate and sufficient description of the Universe at the lowest and highest redshifts. The robustness of the standard cosmological model has recently been questioned on the grounds of a strong and unexplained discrepancy between the local H_0 measured from SNe with distances calibrated by Cepheids and its CMB-based counterpart (currently a 4.4σ difference; Riess et al., 2019). The inverse distance ladder calibrated on the CMB should therefore be taken with caution. Recently, Macaulay et al. (2019) performed an inverse-distance-ladder measurement of H_0 adopting the baseline r_d from *Planck*, and therefore their inferred H_0 agrees with CMB predictions, as expected.

Observations of BAO alone only constrain a combination of the sound horizon and a distance or the expansion rate at the corresponding redshift, i.e.[1] $r_\mathrm{d}/D(z)$ and $r_\mathrm{d}H(z)$. Using SNe, we can propagate BAO observables to redshift $z = 0$ and obtain constraints on $r_\mathrm{d}H_0$ that are fully independent of the CMB (L'Huillier & Shafieloo, 2017; Shafieloo et al., 2018). The extrapolation to low redshifts can be performed using various cosmographic techniques, so that the final

[1] Distances are defined more precisely below.

measurement is essentially independent of cosmological model. Furthermore, combining BAO constraints with a low-redshift absolute calibration of distances or the expansion history, we can break the intrinsic degeneracy of the BAO between r_d and $H(z)$ and thus can determine the sound horizon scale. The resulting measurement is based solely on low-redshift observations, and it is therefore an alternative based on the local Universe to the sound horizon inferred from the CMB.

Several different calibrations of distances or the expansion history have been used to obtain independent low-redshift measurements of the sound horizon. The main results include the calibration of $H(z)$ estimated from cosmic chronometers (Heavens et al., 2014; Verde et al., 2017), the local measurement of H_0 from SNe with distances calibrated with Cepheids (Bernal et al., 2016), angular diameter distances to lens galaxies (Jee et al., 2016; Wojtak & Agnello, 2019), and adopting the Hubble constant from time-delay measurements (Aylor et al., 2019), although the last measurement is based on cosmology-dependent modelling (Birrer et al., 2019). Currently, the sound horizon is most precisely constrained by a combination of BAO measurements from the *Baryon Oscillations Spectroscopic Survey* (BOSS; Alam et al., 2017), with a calibration from the *Supernovae and H_0 for the Equation of State of dark energy* project SHoES; Riess et al., 2019. A significantly higher local value of the Hubble constant than its CMB-inferred counterpart implies a substantially smaller sound horizon scale than its analogue inferred from the CMB under the assumption of the standard ΛCDM model (Aylor et al., 2019). The discrepancy in H_0 and r_d may indicate a generic problem of distance scale at lowest and highest redshifts within the flat ΛCDM cosmological model (Bernal et al., 2016).

Here, we present a self-consistent inference of H_0 and r_d from BAO, SNe Ia, and time-delay likelihoods released by the HoLiCOW collaboration (Birrer et al., 2019; Suyu et al., 2017; Suyu et al., 2010; Suyu et al., 2014; Wong et al., 2017). We examine flat-ΛCDM models as a benchmark and different classes of cosmology-free models. Our approach allows us to determine the local sound horizon scale in a model-independent manner. A similar method was employed by Taubenberger et al. (2019), who used SNe to extrapolate constraints from time-delays to redshift $z = 0$, and thus to obtain a direct measurement of the Hubble constant that depends rather weakly on the adopted cosmology. Throughout this work, comoving distances, luminosity distances, and angular diameter distances are denoted by D_M, D_L, and D_A, respectively. We also adopt the distance duality relations $D_M(z_1 < z_2) = D_L(z_1 < z_2)/(1 + z_2)$, $D_A(z_1 < z_2) = D_M(z_1 < z_2)/(1 + z_2)$, which should hold in all generality and whose validity with current data sets has been tested (Wojtak & Agnello, 2019).

This chapter is organised as follows. The data sets, models, and inference are outlined in Section 3.2. Results are given in Section 3.3, and their implications are discussed in Section 3.4.

3.2 DATA SETS, MODELS, AND INFERENCE

We used a combination of different low-redshift probes to set different distance measurements and different models for the expansion history. All models inferred the following set of parameters: H_0, r_d, M_1 (normalisation of the SN distance moduli), and coefficients parametrising the expansion history or distance as a function of redshift. Curvature Ω_k is left as a free parameter in some models. The sample of high-redshift quasars introduces two additional free parameters: the normalisation M_2 and the intrinsic scatter $\sigma_{\rm int}$ of the quasar distance moduli.

3.2.1 Models

The first model, for homogeneity with previous literature, adopted a polynomial expansion of $H(z)$ in z:

$$H(z) = H_0\,(1 + \mathcal{B}_1 z + \mathcal{B}_2 z^2 + \mathcal{B}_3 z^3) + \mathcal{O}(z^4), \tag{3.1}$$

where the coefficients are related to the standard kinematical parameters, i.e. the deceleration q_0, jerk j_0 , and snap s_0, in the following way (Visser, 2004; Weinberg, 2008; Xu & Wang, 2011):

$$\begin{aligned}
\mathcal{B}_1 &= 1 + q_0 \\
\mathcal{B}_2 &= \frac{1}{2}(j_0 - q_0^2) \\
\mathcal{B}_3 &= \frac{1}{6}\left[3q_0^3 + 3q_0^2 - j_0(3 + 4q_0) - s_0\right] .
\end{aligned}$$

Model distances were computed through direct integration of $1/H(z)$. In our second chosen model family, $H(z)$ was expanded as a polynomial in $x = \log(1+z)$:

$$H(x) = H_0\,(1 + \mathcal{C}_1 x + \mathcal{C}_2 x^2 + \mathcal{C}_3 x^3) + \mathcal{O}(x^4). \tag{3.2}$$

Plugging the Taylor expansion of $z = 10^x - 1$ into eq. (3.1) and grouping the new terms by order, we find the following mapping between coefficients C_i and the kinematical parameters:

$$\begin{aligned}
\mathcal{C}_1 &= \ln(10)\,(1 + q_0) \\
\mathcal{C}_2 &= \frac{\ln^2(10)}{2}\,(-q_0^2 + q_0 + j_0 + 1) \\
\mathcal{C}_3 &= \frac{\ln^3(10)}{6}\left[3q_0^3 + q_0(1 - 4j_0) - s_0 + 1\right] .
\end{aligned}$$

Here, distances were also computed through direct numerical integration of $1/H(z)$. In our third model choice, comoving distances were computed through expansion in $y = z/(1+z)$, and $H(z)$ was obtained through a general relation (E.-K. Li et al., 2020),

$$H(z, \Omega_{\rm k}) = \frac{c}{\partial D_{\rm M}(z)/\partial z}\sqrt{1 + \frac{H_0^2\Omega_{\rm k}}{c^2} D_{\rm M}(z)^2}\,. \tag{3.3}$$

When a polynomial expansion

$$D_{\rm M}(y) = \frac{c}{H_0}\left[\, y + \mathcal{D}_2 y^2/2 + \mathcal{O}(y^3)\,\right] \tag{3.4}$$

is adopted, then the second-order coefficient $\mathcal{D}_2$ is related to the deceleration parameter q_0 through

$$q_0 = 1 - \mathcal{D}_2\,. \tag{3.5}$$

Adopting multiple families of parametrisations, for $H(z)$ and/or for model distances allowed us to quantify the systematics due to different ways of extrapolating the given distance measurements down to $z = 0$. This is equivalent to another common choice of adopting different cosmologies to extend the CDM model, but with the important difference that our chosen parametrisations are completely agnostic about what the underlying cosmological model should be.

Lastly, for the sake of comparison with widely adopted models, we also adopted a ΛCDM model class, with a uniform prior $\Omega_k = [-1.0, 1.0]$ on curvature, and with the constraint that $\Omega_\Lambda + \Omega_m + \Omega_k = 1$. A discrepancy in flat-ΛCDM ($\Omega_k = 0$) between CMB measurements and our low-redshift measurements would then indicate that more general model families are required, i.e. possible departures from concordance cosmology, or that the standard model needs to be extended.

3.2.2 Data sets

Our measurement relies on the complementarity of different cosmological probes. BAO observations constrain $r_d H(z)$ at several different redshifts and independently of the CMB. Standard candles play the role of the inverse distance ladder, by means of which the BAO constraints can be extrapolated to redshift $z = 0$. Finally, gravitational lensing time-delays place constraints on H_0, thus breaking the degeneracy between H_0 and r_d in the inverse distance ladder of BAO and standardisable candles.

In our study, we used pre-reconstruction (independent of cosmological model) consensus measurements of the BAO from the Baryon Oscillations Spectroscopic Survey (Alam et al., 2017). For the relative luminosity distances, we employed binned distance moduli of SNe Ia from the Pantheon sample (D. M. Scolnic et al., 2018). We excluded possible changes due to the choice of SN sample by re-running our inference on distance moduli from JLA (Betoule et al., 2014), and with the current quality of data, there is no appreciable change in the results. Finally, we used constraints on time-delays of four strongly lensed quasars observed by the HoLiCOW collaboration (see Birrer et al., 2019; Suyu et al., 2017, and references therein). Results from a fifth lens have recently been communicated by HoLiCOW (Rusu et al., 2019). We currently use only results that have been reviewed, validated, and released.

As an option that provides more precise distance indicators at high redshifts, we used distance moduli estimated from a relation between UV and X-ray luminosity quasars, which was proven to be an alternative standard candle at high redshift (Risaliti & Lusso, 2018). Risaliti and Lusso (2018) reported that quasar distances at high redshift show a deviation from ΛCDM; however, the lack of any corroborative pieces of evidence does not allow us to conclude if this deviation is a genuine cosmological anomaly or an unaccounted-for systematic effect. For this reason, we dismissed the quasar data at redshifts $z > 1.8$, which is the highest redshift of lensed quasars in our sample.

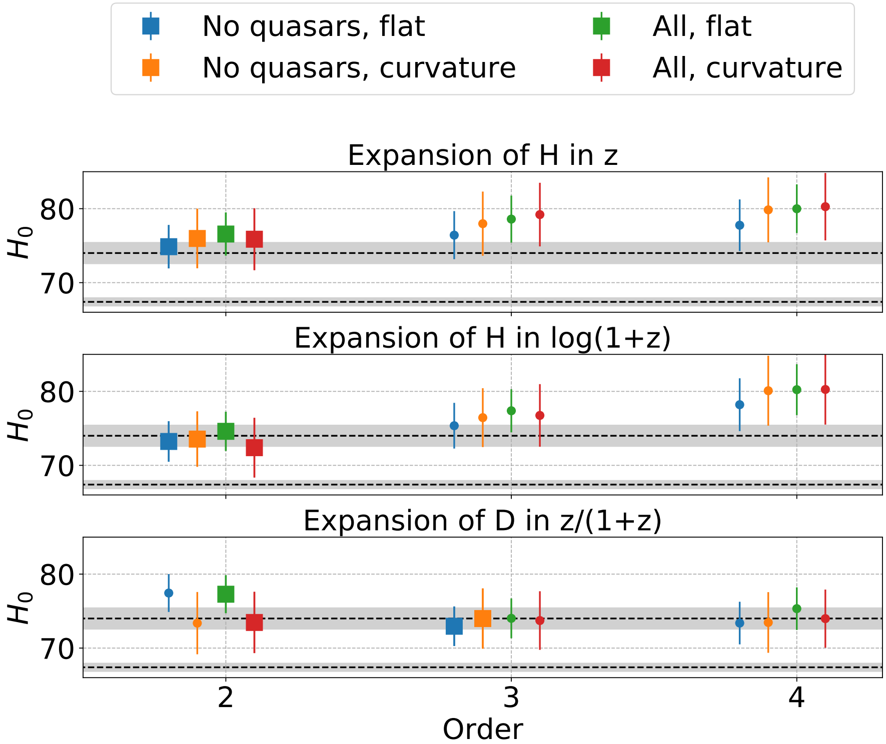


FIGURE 3.1 – Inferred Hubble constant H_0 (in km/s/Mpc) vs. the chosen model family and expansion truncation. The fiducial values from each expansion model (displayed as squares) are chosen by considering the change in BIC score and in $\ln \mathcal{L}_{m.a.p.}$ vs. the change in degrees of freedom. The upper dashed line corresponds to the local measurement value of $H_0 = 74.0$ km s^{-1} Mpc^{-1} with a Cepheid calibration, and the lower dashed line corresponds to the Planck value of $H_0 = 67.4$ km s^{-1} Mpc^{-1}. The shaded grey regions show the error bars.

3.2.3 Inference

The best-fit parameters and credibility ranges of the different expansion models were obtained by sampling the posterior using affine-invariant Monte Carlo Markov chains (Goodman & Weare, 2010a), and in particular with the python module emcee (Foreman-Mackey et al., 2013). For the BAO and SN data set, the uncertainties are given by a covariance matrix $\mathbf{C}$. The likelihood is obtained by

$$\begin{aligned} \mathcal{L} &= p(\text{data}|\text{model}) \propto e^{-\chi^2/2} \\ \chi^2 &= \mathbf{r}^\dagger \mathbf{C}^{-1} \mathbf{r}, \end{aligned} \tag{3.6}$$

where $\mathbf{r}$ corresponds to the difference between the value predicted by the expansion and the observed data.

The high-redshift quasar sample contains significant intrinsic scatter, σ_{int}, which has to be modelled as an additional free parameter. The total uncertainty on each quasar data point is the

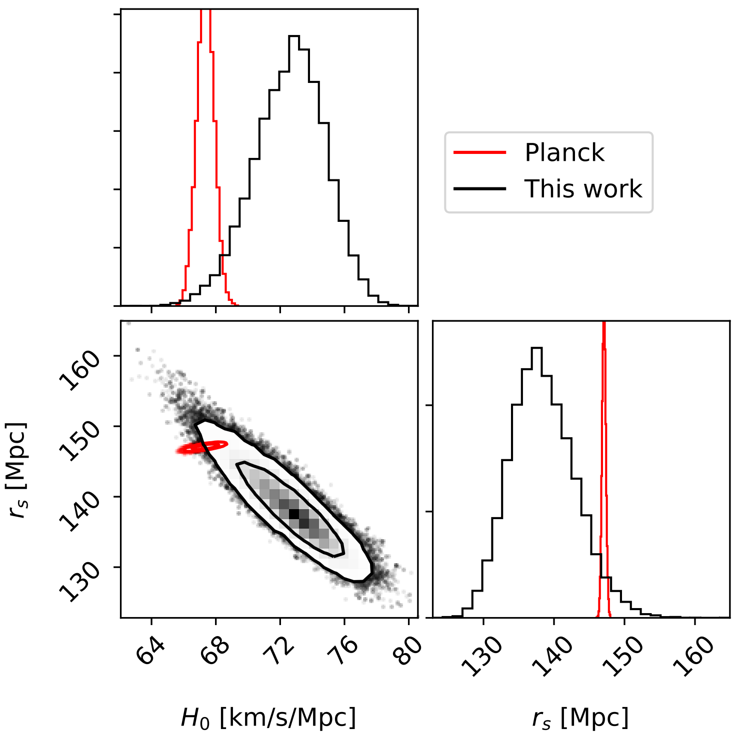


FIGURE 3.2 – Inference on cosmological parameters, including the Hubble constant H_0 and sound horizon r_d, for the baseline case of flat-ΛCDM models using time-delay lenses, SN Ia, and BAO as late-time indicators. The outermost credibility contour contains 95% of the marginalised posterior probability, and the innermost contour contains 68%.

sum of σ_i, the uncertainty of that data point, and σ_{int}. This leads to the following formula of the likelihood:

$$\mathcal{L}_{\text{quasars}} = \sum_{i=1}^{N} \frac{e^{-r_i^2/2(\sigma_i^2+\sigma_{\text{int}}^2)}}{\sqrt{(\sigma_i^2+\sigma_{\text{int}}^2)2\pi}}. \tag{3.7}$$

The likelihoods of the lensed quasars HE0435, RXJ1131, and B1608 of the HoLiCOW collaboration were given as skewed log-normal distributions of their time-delay distances $D_{\Delta t} = (1+z_l)D_l D_s/D_{ls}$ (see Section 1.4.6.2). For the lensed quasar J1206, both the angular diameter distance and the time delay distance were available in the form of a sample drawn from the model posterior distribution. A Gaussian kernel density estimator (KDE) was used to interpolate a smooth distribution between the posterior points.

The final log-likelihood that was sampled by emcee is a sum of the separate likelihoods of the SN, BAO, lensed quasars, and high-redshift quasars,

$$\ln(\mathcal{L}_{\text{total}}) = \ln(\mathcal{L}_{\text{SN}}) + \ln(\mathcal{L}_{\text{BAO}}) + \ln(\mathcal{L}_{\text{lenses}}) + \ln(\mathcal{L}_{\text{quasars}}). \tag{3.8}$$

We note that the high-redshift quasar likelihood is optional in our study. For all cosmographic models used in our work, parameter inference is carried out with or without quasar data, and both results are consistently reported.

A uniform prior was used for all the free parameters, except when the high-redshift quasar sample was used. In that case, the intrinsic scatter σ_{int} was also constrained to be larger than zero. This choice of priors does not seem to bias the inference according to current data and tests on flat-ΛCDM mocks.

To choose the right order of expansion for each model, the Bayesian information criterion (BIC) indicator (eq. (2.5)) was used to assign a score to each order. Table 3.1 displays the number of free parameters, maximum a posteriori likelihood, and the BIC score for four increasing orders of expansion. The expansion order with five free parameters provides the lowest BIC score. When more complexity is added to the model, the BIC value continuously increases, which supports the conclusion that higher expansion orders will be ruled out as well. When the high-redshift quasar sample was added to the data collection, it changed the preferred order of expansion of model 3 from third to second order.

Model 1

parameter	first order	second order	third order	fourth order
Free parameters	4	5	6	7
$\ln \mathcal{L}_{m.a.p.}$	-60.8	-55.8	-55.3	-55.2
BIC score	137.2	131.1	134.1	137.8

Model 2

parameter	first order	second order	third order	fourth order
Free parameters	4	5	6	7
$\ln \mathcal{L}_{m.a.p.}$	-67.1	-56.8	-55.7	-55.0
BIC score	149.8	133.2	134.9	137.4

Model 3

parameter	second order	third order	fourth order	fifth order
Free parameters	4	5	6	7
$\ln \mathcal{L}_{m.a.p.}$	-61.0	-56.1	-56.0	-54.5
BIC score	137.6	131.7	135.5	136.3

TABLE 3.1 – Overview of the number of free parameters, maximum a posteriori likelihood, and BIC score for different expansion orders for cosmographic models 1, 2, and 3. These numbers were calculated using the four lenses, SN, and BAO points and assuming a flat Universe. For expansion in H (models 1 and 2) the second order is preferred, and for expansion in distance (model 3) the third order is preferred. This corresponds to five free parameters in each of the models

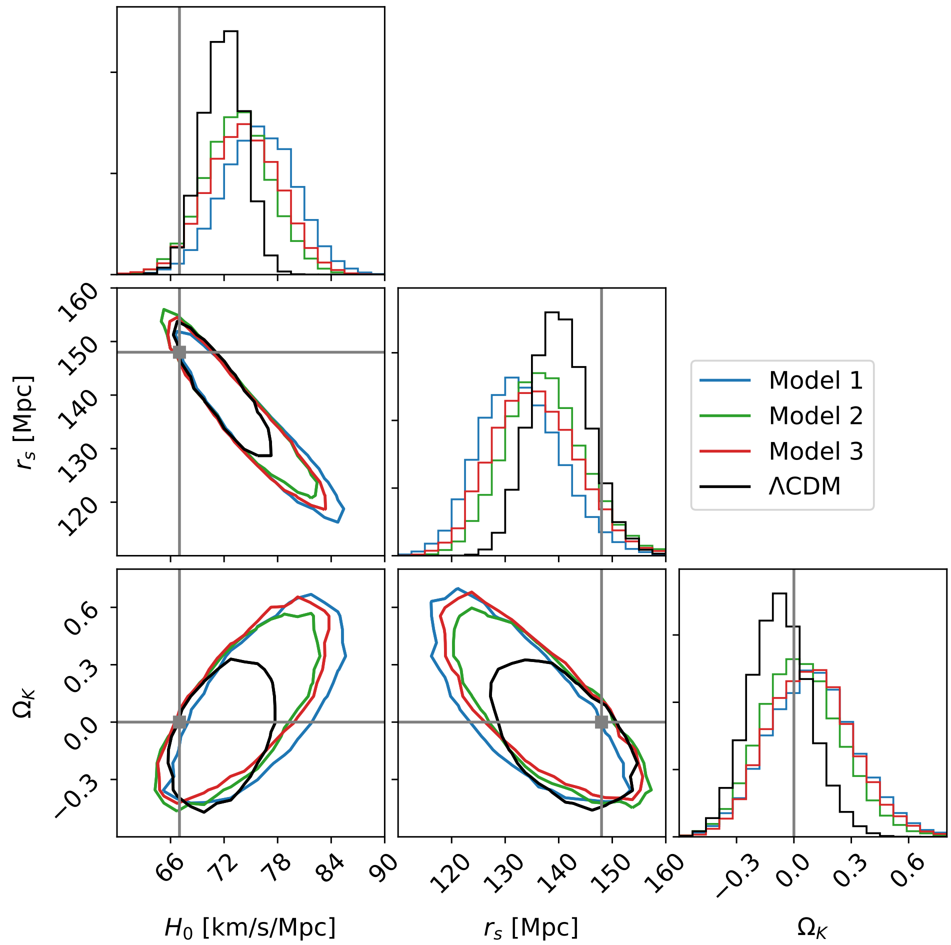


FIGURE 3.3 – Inference on the Hubble constant H_0 and sound horizon r_d for different models (at fiducial truncation order for models 1-3), with free Ω_k, using time-delay lenses, SN Ia, and BAO. While the inferred parameters can change among models and among truncation choices, the relative discrepancy with CMB measurements remains the same. The credibility contours contain 95% of the marginalised posterior probability. The grey point corresponds to the *Planck* value of H_0 and r_d and to a flat Universe.

3.3 RESULTS AND DISCUSSION

THE INFERRED values from our inference are given in Tables 3.2 and 3.3. For the sake of compactness, we report only the inferred values for each model that correspond to the lowest BIC scores (and to a ΔBIC > 2). Figure 3.1 shows the change in H_0 as inferred by different expansion orders. Plots of marginalised posteriors on selected cosmological parameters are given in Figures 3.2 and 3.3.

The inferred values of the Hubble constant from Table 3.2, both its maximum a posteriori and uncertainty, vary between (73.0 ± 2.7) km s^{-1} Mpc^{-1} and (76.0 ± 4.0) km s^{-1} Mpc^{-1}. They

	flat ($\Omega_k = 0$)			
parameter	model 1 (2[nd] order)	model 2 (2[nd] order)	model 3 (3[rd] order)	model 4 (fΛCDM)
r_d (Mpc)	135.26 ± 5.22	138.38 ± 4.97	137.76 ± 4.970	138.74 ± 4.67
$H_0 r_d$ (km s^{-1})	10091.06 ± 147.54	10095.11 ± 146.23	10069.64 ± 149.82	10046.10 ± 137.33
H_0 (km s^{-1} Mpc^{-1})	74.71 ± 2.92	73.06 ± 2.65	73.09 ± 2.67	72.48 ± 2.24
q_0	-0.62 ± 0.078	-0.72 ± 0.11	-0.57 ± 0.18	—
$\ln \mathcal{L}_{m.a.p.}$	-55.76	-56.80	-56.06	-56.31
BIC score	131.07	133.15	131.68	128.30
$\ln \tau$ (Planck ΛCDM)	3.1 (2.0σ)	2.3 (1.6σ)	2.3 (1.7σ)	2.5 (1.8σ)
	free Ω_k			
parameter	model 1 (2[nd] order)	model 2 (2[nd] order)	model 3 (3[rd] order)	model 4 (ΛCDM)
r_d (Mpc)	133.04 ± 7.57	137.57 ± 7.80	136.19 ± 8.05	139.91 ± 5.54
$H_0 r_d$ (km s^{-1})	10069.16 ± 156.97	10079.25 ± 158.20	10052.22 ± 162.32	10073.39 ± 155.18
H_0 (km s^{-1} Mpc^{-1})	75.91 ± 4.07	73.48 ± 3.86	73.82 ± 4.06	72.09 ± 2.41
Ω_k	0.099 ± 0.23	0.038 ± 0.21	0.079 ± 0.22	-0.066 ± 0.16
q_0	-0.62 ± 0.087	-0.71 ± 0.11	-0.55 ± 0.23	—
$\ln \mathcal{L}_{m.a.p.}$	-56.08	-57.10	-56.32	-56.19
BIC score	135.63	137.67	136.11	131.95
$\ln \tau$ (Planck ΛCDM)	2.3 (1.6σ)	1.6 (1.3σ)	1.5 (1.2σ)	2.3 (1.6 σ)

TABLE 3.2 – Inference on the cosmological parameters from BAO+SNe+lenses in our four model classes, with or without imposed flatness. We list the posterior mean and 68% uncertainties of the main parameters, the maximum a posteriori likelihood, the BIC score, and the odds τ that our measurements of H_0 and r_d are consistent with those from the *Planck* observations, as derived for the standard flat-ΛCDM cosmological model

are in full agreement with current results form the HoLiCOW and SHoES collaborations, even despite the choice of general and agnostic models in our method. This indicates that the discrepancy between Cepheid-calibrated H_0 and that inferred from CMB measurements is not due to (known and unknown) systematics in the very low redshift range. The inferred sound horizon r_d varies between (133 ± 8) Mpc and (138 ± 5) Mpc. The largest discrepancy with the value from CMB and standard model predictions (147.09 ± 0.26 Mpc) is more significant for models that are agnostic to the underlying cosmology.

The systematic uncertainties, due to different model choices, are still within the range allowed by statistical uncertainties. However, they may become dominant in future measurements aiming at percent-level precision. Adding UV-Xray standardisable quasars generally raises the inferred value of H_0 (and correspondingly lowers the inferred r_d), even though the normalisation of their Hubble diagram is treated as a nuisance parameter. The addition of the quasar sample also results in lower values of Ω_k. This suggests that the behaviour of distance modulus with redshift has sufficient constraining power on auxiliary cosmological parameters that in turn are degenerate with H_0 in the time-delay lensing standardisation. For all cosmographic models, the intrinsic scatter in quasar distance moduli found in our analysis is 1.45 mag, which is fully consistent with the estimate reported in Risaliti and Lusso (2018).

We quantified the tension with CMB measurements through the two-dimensional inference

parameter	model 1 (2^{nd} order)	model 2 (2^{nd} order)	model 3 (2^{nd} order)	model 4 (fΛCDM)
		flat ($\Omega_k = 0$)		
r_d (Mpc)	132.36 ± 5.05	135.67 ± 4.84	131.63 ± 4.45	138.24 ± 4.62
$H_0 r_d$ (km s^{-1})	10124.73 ± 143.40	10111.40 ± 147.68	10186.40 ± 145.68	9999.72 ± 134.38
H_0 (km s^{-1} Mpc^{-1})	76.59 ± 2.90	74.62 ± 2.67	77.38 ± 2.52	72.40 ± 2.21
q_0	-0.70 ± 0.074	-0.82 ± 0.105	-1.13 ± 0.11	—
$\ln \mathcal{L}_{m.a.p.}$	-2335.33	-2338.02	-2339.59	-2338.14
BIC score	4720.84	4726.22	4722.19	4719.30
$\ln \tau$ (Planck ΛCDM)	4.9 (2.7σ)	3.5 (2.2σ)	7.8 (3.5σ)	2.5 (1.7σ)
		free Ω_k		
parameter	model 1 (2^{nd} order)	model 2 (2^{nd} order)	model 3 (2^{nd} order)	model 4 (ΛCDM)
r_d (Mpc)	134.20 ± 8.00	140.74 ± 8.15	139.36 ± 8.40	143.70 ± 5.58
$H_0 r_d$ (km s^{-1})	10132.11 ± 160.61	10150.20 ± 155.94	10223.94 ± 152.08	10140.36 ± 157.6
H_0 (km s^{-1} Mpc^{-1})	75.74 ± 4.16	72.34 ± 3.89	73.37 ± 4.18	70.65 ± 2.29
Ω_k	-0.056 ± 0.22	-0.16 ± 0.20	-0.19 ± 0.17	-0.27 ± 0.14
q_0	-0.70 ± 0.082	-0.82 ± 0.11	-1.11 ± 0.17	—
$\ln \mathcal{L}_{m.a.p.}$	-2335.58	-2337.84	-2339.30	-2336.20
BIC score	4728.53	4733.03	4728.76	4722.56
$\ln \tau$ (Planck ΛCDM)	2.4 (1.7σ)	1.7 (1.3σ)	2.6 (1.8σ)	1.9 (1.4σ)

TABLE 3.3 – Same as for Table 3.2, but including UV-Xray quasars as standardisable distance indicators

on H_0 and r_d. Following Verde et al. (2013), we estimated the odds that both measurements are consistent by computing the following ratio:

$$\tau = \frac{\int\int \hat{p}_{\mathrm{CMB}}\, \hat{p}_{\mathrm{local}}\, \mathrm{d}H_0\, \mathrm{d}r_d}{\int\int p_{\mathrm{CMB}}\, p_{\mathrm{local}}\, \mathrm{d}H_0\, \mathrm{d}r_d}, \tag{3.9}$$

where p is the marginalised probability distribution for r_d and H_0 from the CMB (Planck Collaboration, Aghanim, et al., 2018) or our study (in both cases approximated by Gaussians), while $\hat{p}$ is a distribution shifted to a fixed arbitrary point so that both measurements have the same posterior probability means. A more intuitive scale representing the discrepancy between two measurements is a number-of-sigma tension, which can be derived from the odds ratio. This is done by calculating the probability enclosed by a contour such that $1/\tau = e^{-\frac{1}{2}r^2}$. The number of sigma tension can then be calculated from the probability by means of the error function. We list the logarithm of the odds and the number of sigma tension in Tables 3.2 and 3.3.

The tension with Planck measurements from CMB is approximately at a 2σ level. While the uncertainties from some model families are larger, the corresponding H_0 (r_d) optimal values are also higher (lower), and the tension remains the same. The curvature Ω_k slightly alleviates the tension through larger H_0 uncertainties, but the current data do not yield any evidence of a departure from flatness.

3.4 CONCLUSIONS AND OUTLOOK

CURRENT DATA enable a $\approx 3\%$ determination of key cosmological parameters, in particular, the Hubble constant H_0 and the sound horizon $r_{\rm d}$, resulting in a $\approx 2\sigma$ Gaussian tension with predictions from CMB measurements and the standard model. While this tension is robust against the choice of model family and is therefore independent of the underlying cosmology, the systematics due to different model choices are currently comparable to the statistical uncertainties and may dominate percent-level measurements of H_0. A simple estimate based on recent SHoES measurements (Riess et al., 2019) and very recent five-lens measurements by HoLiCOW (Rusu et al., 2019) indicates a $\approx 5\sigma$ tension with CMB measurements within a flat-ΛCDM model.

Our study also demonstrated the potential of constraining the curvature of the Universe solely based on low-redshift observations and in a cosmology-independent manner. The current precision of 0.20 is insufficient to test possible minimum departures from flatness, mainly due to the accuracy in H_0 from a small sample of well-studied lenses. Samples of lenses with suitable ancillary data are already being assembled (see e.g. Shajib et al., 2019). Future measurements of gravitational time-delays from the Large Synoptic Survey Telescope can reach percent-level precision (Liao et al., 2015), making this method a highly competitive probe (Denissenya et al., 2018).

AUTHOR CONTRIBUTIONS

NA led the project, performed the analysis, produced the figures, contributed to writing up the manuscript. AA proposed the main idea, provided the HoLiCOW data sets, validated the results, contributed to writing up the manuscript. RW proposed the main idea, validated the results, contributed to writing up the manuscript.

ACKNOWLEDGEMENTS

The authors were supported by a grant from VILLUM FONDEN (project number 16599). This project is partially funded by the Danish Council for independent research under the project "Fundamentals of Dark Matter Structures", DFF–6108-00470.
We are grateful for the public release of the time-delay distance likelihoods by the HoLiCOW collaboration, and interesting conversations with S. H. Suyu, F. Courbin and T. Treu.
We thank Guido Risaliti for sharing distance moduli measured from high-redshift quasars.
We also thank the anonymous referee for constructive comments which helped improve our work.

CHAPTER 4

Can modifications of ΛCDM solve the tension?

This chapter is based on the following article:
"Cosmic dissonance: are new physics or systematics behind a short sound horizon?"

Published in Astronomy & Astrophysics (A&A), 639, A57 (2020).

Authors: Nikki Arendse, Radosław J. Wojtak, Adriano Agnello, Geoff C.-F. Chen, Christopher D. Fassnacht, Dominique Sluse, Stefan Hilbert, Martin Millon, Vivien Bonvin, Kenneth C. Wong, Frédéric Courbin, Sherry H. Suyu, Simon Birrer, Tommaso Treu & Leon V.E. Koopman.

Abstract

Persistent tension between low-redshift observations and the Cosmic Microwave Background (CMB) radiation, in terms of two fundamental distance scales set by the sound horizon $r_{\rm d}$ and the Hubble constant H_0, suggests new physics beyond the standard model, departures from concordance cosmology, or residual systematics. The role of different probe combinations must be assessed, as well as of different physical models that can alter the expansion history of the Universe and the inferred cosmological parameters. We examine recently updated distance calibrations from Cepheids, gravitational lensing time-delay observations, and the Tip of the Red Giant Branch. Calibrating the Baryon Acoustic Oscillations (BAO) and type Ia supernovae with combinations of the distance indicators, we obtain a joint and self-consistent measurement of H_0 and $r_{\rm d}$ at low redshift, independent of cosmological models and CMB inference. In an attempt to alleviate the tension between late-time and CMB-based measurements, we consider

four extensions of the standard ΛCDM model. The sound horizon from our different measurements is $r_\mathrm{d} = (137 \pm 3^{stat.} \pm 2^{syst.})$ Mpc based on absolute distance calibration from gravitational lensing and the cosmic distance ladder. Depending on the adopted distance indicators, the *combined* tension in H_0 and r_d ranges between 2.3 and 5.1 σ, and is independent of changes to the low-redshift expansion history. We find that modifications of ΛCDM that change the physics after recombination fail to provide a solution to the problem, for the reason that they only resolve the tension in H_0, while the tension in r_d remains unchanged. Pre-recombination extensions (with early dark energy or the effective number of neutrinos $\mathrm{N_{eff}} = 3.24 \pm 0.16$) are allowed by the data, unless the calibration from Cepheids is included. Results from time-delay lenses are consistent with those from distance-ladder calibrations and point to a discrepancy between absolute distance scales measured from the CMB (assuming the standard cosmological model) and late-time observations. New proposals to resolve this tension should be examined with respect to reconciling not only the Hubble constant but also the sound horizon derived from the CMB and other cosmological probes.

4.1 INTRODUCTION

At the onset of matter-radiation decoupling after the Big Bang, photon-baryon fluid underwent oscillations whose characteristic physical scale is described by the so-called *sound horizon* r_d. This leaves a characteristic imprint on large scale distribution of baryons, with its characteristic size fixed in the comoving coordinates and equal to the sound horizon at the drag epoch z_d given by

$$r_\mathrm{d} \equiv r_\mathrm{s}(z_\mathrm{d}) = \int_{z_\mathrm{d}}^{\infty} \frac{c_\mathrm{s} \mathrm{d}z}{H(z)} \ , \tag{4.1}$$

where c_s is the sound speed in the primordial plasma and $H(z)$ is the Hubble parameter.

The sound horizon r_d is robustly determined from the Cosmic Microwave Background measurements (CMB), if the standard model of particle physics as well as the standard cosmological model in the pre-recombination Universe are adopted (Planck Collaboration, Aghanim, et al., 2018). Alternatively, it can be measured at later times, from the Baryon Acoustic Oscillation (BAO) peak in the two-point spatial correlation function of galaxies and quasars. The latter is an angular measurement, which can be converted into a physical r_d measurement through independent distance calibrations (see e.g. Arendse et al., 2019; Aylor et al., 2019; Bernal et al., 2016; Heavens et al., 2014; Verde et al., 2017). The parameter r_d is intimately linked to the current expansion rate of the Universe, the Hubble constant H_0, since BAO measurements constrain the product of H_0 and r_d.

Accurate distance measurements from CMB-independent observations can be used to determine r_d and H_0 in a way that is truly independent of early-Universe physics. Therefore, these measurements can test our understanding of the concordance cosmology and the standard model of particle physics, through low-redshift measurements only. Type Ia supernovae, calibrated by Cepheids with three independent distance anchors (parallaxes in the Milky Way, detached eclipsing binaries in the LMC and maser galaxy NGC 4258), provide the most precise distance calibration to date, as performed by the *Supernovae and H_0 for the Equation of State of dark energy* project (SHoES; Riess et al., 2019). Another powerful way of obtaining absolute distances is by using strongly lensed quasar systems, which extend to higher redshifts than the Cepheids. The *H_0 Lenses in COSMOGRAIL's Wellspring* collaboration HoLiCOW, Suyu et al., 2017 has provided few-percent-level precision constraints on H_0 from time-delay cosmology. Over the whole sample, the effect of known systematics is at $\lesssim 1\%$ level, currently negligible with respect to the statistical uncertainties (Millon et al., 2019). The latest results from SHoES and HoLiCOW indicate a strong tension in the Hubble constant H_0 between late-time observations (CMB-independent probes including primarily type Ia SNe, lensing and BAO) and CMB-based measurements, within a flat-ΛCDM model. Previous results based on four lenses alone (Arendse et al., 2019; Taubenberger et al., 2019) resulted in a 2σ discrepancy, while a six-lens analysis (Wong et al., 2020) gave a 3σ tension. When combined with the distance-ladder results by SHoES, the tension increases to a 5σ level, still adopting a flat ΛCDM cosmological model. It is worth noting that the tension between the late-time and CMB-based measurements of H_0 is mildly lowered by the recent measurement making use of precise distance calibration from the Tip of the Red Giant Branch (TRGB), as measured by the Carnegie-Chicago Hubble Project (herafter CCHP, Freedman et al., 2019). These measurements fall between those from SHoES and the CMB, at 1.7σ and 1.2σ differences respectively. For the sake of completeness, it is also worth mentioning that the Planck value of the Hubble constant is recovered in a CMB-independent but model-dependent analysis of BAO observations with the prior on the baryon density from the standard Big Bang Nucleosynthesis (Addison et al., 2018; Cuceu et al., 2019).

In this work, we revisit the claimed tension between late-time observations and the CMB in terms of the sound horizon and the Hubble constant, by making use of recent updated distance calibrations from gravitational time-delay lenses (HoLiCOW), Cepheids (SHoES), and TRGB (CCHP). Through our methods (summarised in Section 4.2.2), we obtain measurements of r_d for different combinations of late-time distance calibrations in a manner that is almost completely independent of any cosmological model. Moreover, we investigate selected extensions to the standard ΛCDM model that have recently been proposed as possible solutions to the Hubble tension. Such new models attempt to reconcile the tension by modifying the expansion history of the standard model either before or after recombination, hereafter *early-time* and *late-time* modifications, and thus increasing the Hubble constant derived from the CMB. We demonstrate that the late-time extensions fail to provide a solution to the problem, for the reason that they only succeed in alleviating the tension in H_0, while the tension in r_d remains unchanged. Our analysis emphasises the importance of comparing at least H_0 and r_d derived from late-time observations and the CMB when testing new models devised to mitigate the Hubble constant tension.

This chapter is structured as follows. Section 4.2 describes the late-time measurements of r_d and H_0, including the different data sets, models and inference methods that are used. In Section 4.3 we outline how the late-time measurements are compared with CMB inference and extensions of the concordance scenario. Our results are described in Section 4.4 and our conclusions in Section 4.5.

4.2 LATE-TIME MEASUREMENTS: DATA AND METHODS

The values of r_{d} and H_0 can be constrained by employing several CMB-independent probes at $0 < z < 2$, referred to as late-time measurements in this chapter. In Section 4.2.1, we provide an overview of the data sets that we use in our analysis. Section 4.2.2 introduces the models we choose to fit the Hubble diagram and interpolate up to redshift zero. By choosing models that are independent of cosmology we minimise the systematic uncertainty associated with cosmological model choices. Details about the inference are discussed in Section 4.2.3, and functional tests are shown in Appendix A.2.

4.2.1 Data sets

The shape of the late-time expansion of the Universe has been mapped precisely with type Ia supernovae (SNe). In this work, we use relative distance moduli from the Pantheon sample (D. M. Scolnic et al., 2018)

Information about r_{d} is introduced by adding BAO measurements, which constrain the product of H_0 and r_{d}. Our main results are obtained for the Hubble parameters $H(z)$ and the transverse comoving distances $D_{\mathrm{M}}(z)$ determined from the Baryon Oscillations Spectroscopic Survey (BOSS; Alam et al., 2017). Additionally, we look into the effect of adding BAO constraints from the correlation of Lyα forest absorption and quasars in the extended Baryon Oscillation Spectroscopic Survey (eBOSS; Blomqvist et al., 2019; de Sainte Agathe et al., 2019) and several isotropic BAO measurements. The isotropic measurements do not contain sufficient statistics to measure $H(z)$ and $D_{\mathrm{M}}(z)$ separately, but combine them in the volume-averaged distance $D_{\mathrm{V}} = \left[czD_{\mathrm{M}}^2(z)H^{-1}(z)\right]^{1/3}$. We include two measurements from the reconstructed 6-degree Field Galaxy Survey (Carter et al., 2018), two from eBOSS by (Ata et al., 2018; Bautista et al., 2018), and three from the WiggleZ Dark Energy Survey (Kazin et al., 2014).

Both SNe and BAO measurements provide only relative distances, thus their distance scale needs to be calibrated with absolute distance measurements. Time-delay and angular-diameter distances to strongly lensed quasars, obtained by the HoLiCOW collaboration, provide such an absolute calibration of cosmological distances (see e.g. Suyu et al., 2017, and references therein). Results from a fifth and a sixth lensed quasar system have been recently obtained (Bonvin et al., 2019; Chen et al., 2019; Rusu et al., 2019; Sluse et al., 2019), including new distance measurements on previous lensed quasar systems using new data and analysis (**Jee2019lenses**; Chen et al., 2019). In this work, we use complete constraints on distances from observations of the 6 lensed quasars systems, as summarised in Wong et al. (2020). The information from the lensed quasars is modelled self-consistently, together with the relative distance indicators (SNe, BAO).

Keeping the lensing data as our primary calibration of the absolute distance scale in all fits, we also include two optional priors given by recent local determinations of the Hubble constant. The first is the latest SHoES measurement yielding $H_0 = 74.03 \pm 1.42$ km s^{-1} Mpc^{-1} (Riess et al., 2019). The second is based on calibrating distances with the Tip of the Red Giant Branch (TRGB), a standard candle alternative to Cepheids. Here, analyses carried out by two separate groups have resulted in different values for H_0: Yuan et al. (2019) found 72.4 ± 2.0 km s^{-1} Mpc^{-1}, while CCHP obtained 69.6 ± 2.0 km s^{-1} Mpc^{-1} (Freedman et al., 2019; Freedman et al., 2020). In order to include both the highest and lowest late-time measurements of H_0, we have chosen to use the CCHP results for the TRGB and SHoES results for Cepheids in our analysis. Since there is a partial overlap in the galaxy samples considered for the TRGB and Cepheid measurements, the two calibrations will only be applied separately.

Finally, quasars are optionally used as secondary standard candles at high redshifts, by means of a relation between their UV and X-ray luminosities (Risaliti & Lusso, 2018). We do so in one of our inference runs in Table 4.4, as an independent check.

Our constraints on the late-time expansion is largely based on data sets and models that we explored in previous work (Arendse et al., 2019). The difference with previous data sets is the inclusion of two additional quasar-lens measurements (Chen et al., 2019; Rusu et al., 2019), Lyα BAO measurements at $z = 2.34$ and 2.35, several volume-averaged BAO measurements (D_V BAO), and the combination with the Cepheid distance-ladder or the TRGB calibration.

4.2.2 Models

Measuring r_d and H_0 from the observations described above requires adopting a model of the expansion history. This is usually done by means of employing the standard ΛCDM model, but any tension among different r_d and H_0 measurements in the ΛCDM framework may mean that the ΛCDM expansion history is not necessarily an adequate model choice. Instead of employing different extensions to ΛCDM to overcome this issue, we use three different models of polynomial parametrisations, that are completely agnostic about the underlying expansion history. This allows us to make an inference of r_d and H_0 that is based solely on observational data, and does not rely on cosmology.

The specifications of the three polynomial parametrisations (hereafter referred to as Models 1, 2 and 3) are listed in Table 4.1. Model 1 adopts a polynomial expansion of $H(z)$ (Visser, 2004; Weinberg, 1972), Model 2 expands the luminosity distance D_L[1] as a polynomial in $\log(1+z)$ (Risaliti & Lusso, 2018), and Model 3 describes transverse comoving distances D_M by polynomials in $z/(1+z)$ (Cattoën & Visser, 2007; E.-K. Li et al., 2020). For Model 1, comoving distances are obtained from $H(z)$ through direct numerical integration of

$$d_c(z) = \int_0^z \frac{c}{H(z)}\, dz, \tag{4.2}$$

and for Models 2 and 3, $H(z)$ is obtained through

$$H(z, \Omega_k) = \frac{c}{\partial D_M(z)/\partial z} \sqrt{1 + \frac{H_0^2 \Omega_k}{c^2} D_M(z)^2} \tag{4.3}$$

(Weinberg, 1972).

[1] Where the distance measures are related to each other according to $D_L = (1+z)D_M = (1+z)^2 D_A$.

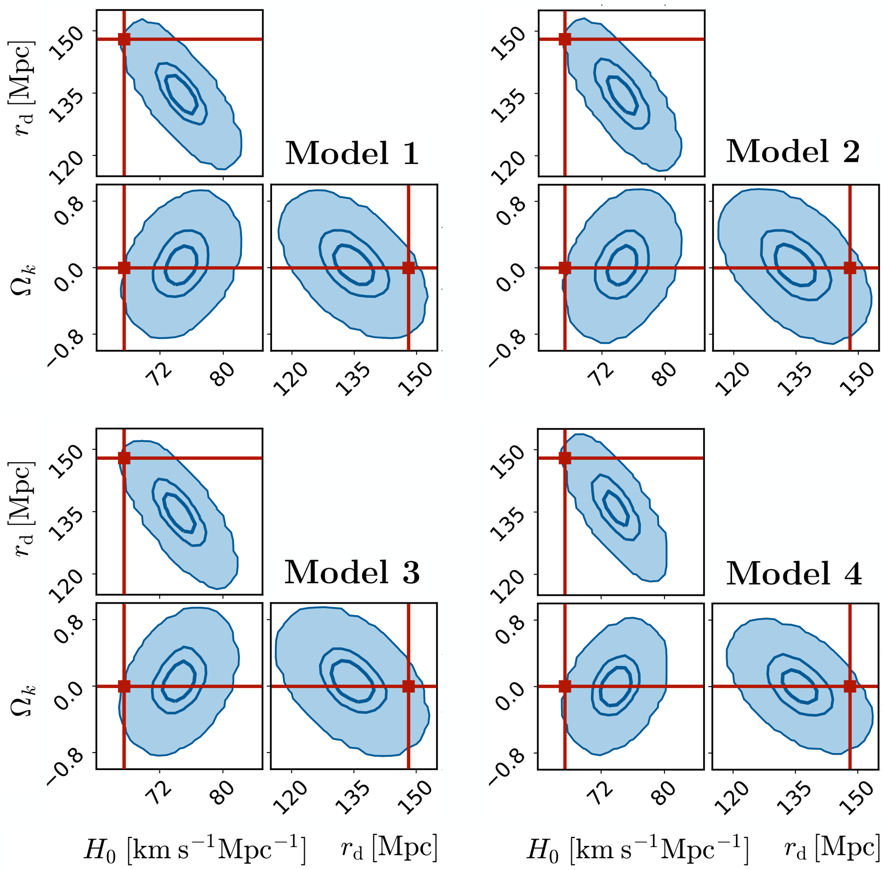


FIGURE 4.1 – Constraints on the sound horizon r_d, the Hubble constant H_0 and Ω_k from late-time observations including BAO (BOSS), type Ia supernovae (Pantheon), gravitational lensing (HoLiCOW) and the cosmic distance ladder calibrated with Cepheids (SH0ES). The panels show results for three cosmology-independent models listed in Table 4.1 and a ΛCDM cosmological model. The red lines indicate the best fit values obtained from Planck for a flat ΛCDM cosmological model. The contours indicate 1-, 2- and 5σ confidence regions of the posterior probability (the latter obtained by Gaussian extrapolation). All panels demonstrate a 5σ tension between r_d and H_0 measured from the CMB and the late-time observations.

We truncate all polynomials at the lowest expansion order required by the condition that Models 1, 2 and 3 recover distances in a ΛCDM model, if their free coefficients are fixed at values found by Taylor expanding the corresponding functions in the fiducial ΛCDM model (see more in Appendix A.2). This guarantees that expansion histories derived from the employed models converge to ΛCDM once observations become consistent exclusively with the standard model. Distances in ΛCDM are recovered with a minimum accuracy of 2 percent at $z < 1.8$, where the accuracy limit is set by the current precision of the Hubble constant measurements

Model	Formula
1	$H(z) = H_0\left(1 + b_1 z + b_2 z^2\right)$
2	$D_L(z) = \frac{c\ln(10)}{H_0}\left\{\log(1+z) + c_2[\log(1+z)]^2 + c_3[\log(1+z)]^3 + c_4[\log(1+z)]^4\right\}$
3	$D_M(z) = \frac{c}{H_0}\left(\frac{z}{1+z} + d_2\left[\frac{z}{1+z}\right]^2 + d_3\left[\frac{z}{1+z}\right]^3 + d_4\left[\frac{z}{1+z}\right]^4\right)$
4	$H(z) = H_0\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2}$

TABLE 4.1 – The three polynomial parametrisations (Models 1, 2 and 3) adopted in this study to place cosmology-independent constraints on r_d and H_0. The fourth case is a ΛCDM cosmological model.

and the upper limit of redshift is given by the most distant lensed quasar. Including higher order terms is disfavoured by Bayesian Information Criterion (BIC). In Appendix A.2, we also show that this convergence criterion ensures that biases in H_0 are at sub-percent level, and biases in q_0 at a few-percent level.

Finally, in order to compare Models 1-3 with the most commonly adopted cosmological model, the fourth family (Model 4) adopts a ΛCDM parametrisation. In all cases, both flatness and departures from it are considered.

4.2.3 Inference

We fit four models listed in Table 4.1 to observational data of type Ia supernovae, BAO and lensed quasars. Constrained model parameters include r_d, H_0 and all remaining free polynomial coefficients (or density parameters in the case of ΛCDM model). The posterior distributions of the parameters are obtained using Affine-Invariant Monte Carlo Markov Chains (MCMC) (Goodman & Weare, 2010a), and in particular the python module emcee (Foreman-Mackey et al., 2013). For the sake of completeness, we also derive constraints on the deceleration parameter q_0 using the MCMC samples. Appendix A.2 outlines the relations between polynomial coefficients, which are primary parameters in our fits, and q_0.

The likelihoods of the distances measured from lensed quasars are either given as a skewed log-normal distribution[2] (for B1608) or as samples of points from the HoLiCOW model posteriors (for RXJ1131, HE0435, PG1115, J1206 and WFI2033). The probability density is obtained by constructing a Gaussian kernel density estimator (KDE). For the lens systems HE0435 and WFI2033 only a robust measurement of their time-delay distance (eq. (1.76)) is provided, which is the only robust distance currently derived from time-delay lensing in the presence of significant perturbers at lower redshift. For the remaining four lenses (B1608, RXJ1131, PG1115,

[2] Full names and coordinates of each lens are given in the HoLiCOW XIII paper (Wong et al., 2020).

J1206), information on both their time-delay distances and their angular diameter distances is available. For the remaining observables (BAO, SNe, quasars, and SHoES or CCHP), the general form of the likelihood for each data set is given by

$$\begin{aligned} \mathcal{L} &= p(\text{data}|\text{model}) \propto e^{-\chi^2/2} \\ \chi^2 &= \mathbf{r}^\dagger \mathbf{C}^{-1} \mathbf{r}, \end{aligned} \tag{4.4}$$

where **C** is the covariance matrix of the data and **r** corresponds to the difference between the predicted and the observed values. The final likelihood is a product of the separate likelihoods corresponding to each data set.

A uniform prior is used for the parameters, for ease of comparison with previous work. In particular, the value of r_d is kept between 0 and 200 Mpc and, if applicable, Ω_k between -1 and 1 and Ω_m between 0.05 and 0.5, to ensure consistency with the priors on Ω_m by HoLiCOW. These priors do not skew the inference, at least with the current uncertainties. The upper and lower boundaries of r_d do not influence any of the results. For the coefficients of the expansion (b_i, c_i and d_i in Table 4.1), we used a uniform prior without limits. In all cases, best fit values are given by the posterior mean and errors provide 68.3 percent confidence intervals. The code to generate the results in this paper is publicly available on Github[3].

4.3 COMPARISON WITH THE CMB: DATA AND MODELS

THE SOUND HORIZON and the Hubble constant are independently measured from the CMB. For the standard flat ΛCDM cosmological model, the Planck observations yield $r_\mathrm{d} = 147.2 \pm 0.3$ Mpc and $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} (Planck Collaboration, Aghanim, et al., 2018). As we shall demonstrate, both parameters are strongly discrepant with their counterparts determined from late-time observations. In the following subsections, we describe how we quantify this tension and outline a few popular extensions of the standard cosmological model devised to reduce the discrepancy.

4.3.1 Quantifying the tension

In order to check whether or not our results for r_d and H_0 are in agreement with those obtained by *Planck*, the Gaussian odds indicator τ is used (Bernal et al., 2016; Verde et al., 2013):

$$\tau = \frac{\int \bar{P}_A \bar{P}_B \, \mathrm{d}x}{\int P_A P_B \, \mathrm{d}x}. \tag{4.5}$$

Here, P_A and P_B denote the posterior distributions of experiments A and B, while $\bar{P}_A$ and $\bar{P}_B$ correspond to the same distributions after a shift has been performed, such that the maxima of P_A and P_B coincide. A high value for τ means that it is unlikely that both experiments measure the same quantity. In an idealised situation, when experiment A yields a measurement with infinite precision (P_A is given a δ function), the odds indicator equals the ratio of probability P_B

[3] https://github.com/Nikki1510/cosmic_dissonance

evaluated at best fit values returned by both experiments. Eq. (4.5) generalises this interpretation to cases where both measurements have non-zero uncertainties.

A more intuitive scale representing the discrepancy between two measurements is a number-of-sigma tension, and it can be directly derived from the odds ratios (see e.g. Bernal et al., 2016). First, the odds indicator is used to calculate the probability enclosed by a contour r such that $1/\tau = e^{-\frac{1}{2}r^2}$. The probability is then converted to a number of sigma tension, using a one-dimensional cumulant (the error function).

4.3.2 Extensions of the ΛCDM model

Any tension between late-time measurements and CMB-based model-dependent inference may be caused by unknown systematics, or it can mean that our knowledge of the physics underlying the expansion history is incomplete. The standard flat ΛCDM model can be extended by either changing physics in the early Universe (pre-recombination; this will be referred to as *early-time* modifications) or at later epochs (post-recombination; this will be referred to as *late-time* modifications). In the first case, one can decrease the sound horizon inferred from the CMB observations by adding an energy-momentum tensor beyond the standard model, which effectively increases $H(z)$ in the early Universe. In order to keep the observed angular scales imprinted in the CMB unchanged, this alteration automatically implies an increase in the value of H_0. Therefore, the overall effect of early-time modifications is a shift of both r_d and H_0 towards the measurements from late-time observations. In the second approach, one may obtain higher values of H_0 by decreasing the expansion rate at intermediate redshifts. This can be done by modifying the dark energy density such that it increases over time. Although many late-time extensions of the standard model can quite easily increase H_0 inferred from the CMB, r_d cannot be modified as appreciably as H_0 – as it is primarily driven by physics in the early Universe.

In order to explore different resolutions of the tension in H_0 and r_d on the grounds of new physics, we consider several extensions of the standard ΛCDM model. Although the selected models do not exhaust all possible proposals from the literature, they are sufficiently representative in terms of covering most possible model-dependent alterations of H_0 and r_d inferred from the CMB. In what follows, the inference for early dark energy and PEDE (described below) have been obtained using a *Planck* compressed likelihood, as detailed in Appendix A. For the remaining models, we use publicly available MCMC chains (based on Planck's temperature and polarisation data) from the Planck Legacy Archive[4] (Planck Collaboration, Aghanim, et al., 2018).

4.3.2.1 *Early-time (pre-recombination) extensions*

• *Effective Number of Relativistic Species (N_{eff}).*
In this extension of ΛCDM, there are additional relativistic particles that contribute to the radiation density of the early Universe, resulting in $\mathrm{N}_\mathrm{eff} > 3$. An increased radiation density leads to a

[4] https://pla.esac.esa.int

later matter-radiation equality and to an increased expansion rate in the early Universe, leaving an observational imprint on the CMB (D. Eisenstein & White, 2004; Hannestad, 2003; Mörtsell & Dhawan, 2018). This in turn reduces the value of the sound horizon r_d at recombination and increases H_0 derived from the CMB, thereby relieving some of the tension between late-time and CMB measurements (Carneiro et al., 2019; Gelmini et al., 2019).

- *Early Dark Energy.*

The expansion rate in the early Universe could also be increased by the presence of a more general form of dark energy. This additional dark energy should have a noticeable contribution to the energy budget at high redshifts, but should dilute away faster than radiation to leave the evolution of the Universe after recombination unchanged (Doran et al., 2007; Linder & Robbers, 2008). As a promising example of this class of models, we consider early dark energy which behaves nominally as a scalar field ϕ with a potential $V(\phi) \propto [1 - \cos(\phi/f)]^3$ (Poulin et al., 2019). In the effective fluid description, the energy density ρ_{EDE} evolves as

$$\rho_{\rm EDE}(a) = \frac{2\rho_{\rm EDE}(a_c)}{1 + (a/a_c)^{9/2}}, \tag{4.6}$$

with the scale factor a (Poulin et al., 2018). The early dark energy equation of state approaches asymptotically -1 for $a \ll a_c$ and $1/2$ for $a \gg a_c$. When fitting the model to the CMB data, we adopt the following flat priors in $\log_{10}(a_c)$ and $f_{\rm EDE} = \Omega_\phi(a_c)/\Omega_{\rm tot}(a_c)$: $-4.0 < \log_{10}(a_c) < -3.2$ and $0.1 > f_{\rm EDE} > 0$.

4.3.2.2 *Late-time (post-recombination) extensions*

- *Time-dependent dark energy (wCDM).*

The wCDM cosmology introduces the equation of state parameter w as a free parameter (as opposed to the fixed ΛCDM value of $w = -1$), so that the dark energy density $\rho_{\rm DE}$ can change as a function of redshift as

$$\rho_{\rm DE}(z) = \rho_{\rm DE,0}(1+z)^{3(1+w)}. \tag{4.7}$$

- *Phenomenologically Emergent Dark Energy (PEDE).*

In the PEDE model, dark energy has no effective role in the early Universe but emerges at later times (X. Li & Shafieloo, 2019). The redshift evolution of the dark energy density is described by

$$\rho_{\rm DE}(z) = \rho_{\rm DE,0} \times [1 - \tanh(\log_{10}(1+z))], \tag{4.8}$$

giving it the same number of degrees of freedom as ΛCDM. We emphasise that this parametrisation is mostly *ad hoc*.

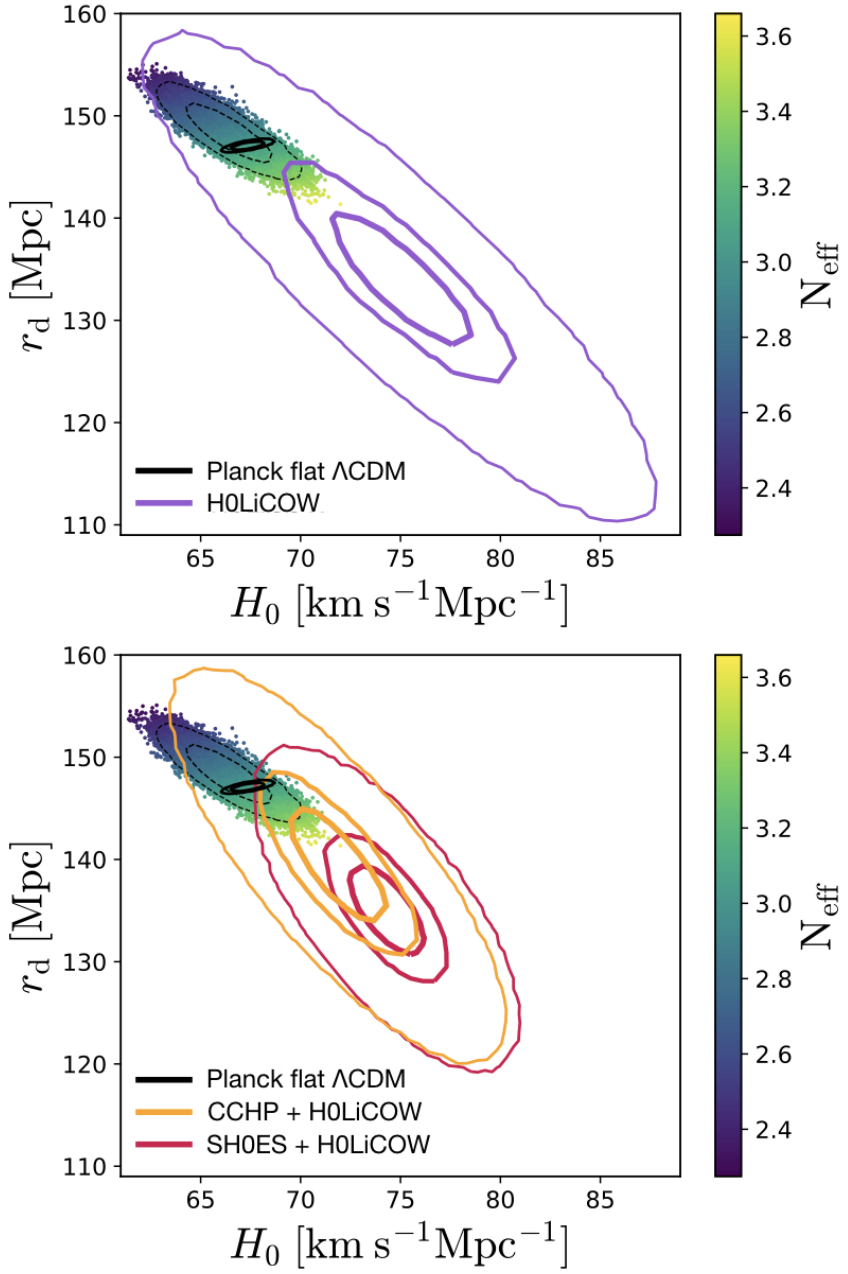


FIGURE 4.2 – Comparison between the sound horizon r_d and the Hubble constant H_0 measured from *Planck* observations of the CMB (assuming a flat ΛCDM) and late-time observations (using flat Model 3) obtained by calibrating SN and BAO measurements with three different absolute distance calibrations from: gravitational lensing (HoLiCOW), the cosmic distance ladder with Cepheids (SHoES) or the TRGB (CCHP). For the late-time data, the contours show 1-, 2- and 5σ confidence regions of the posterior probability (the latter obtained by Gaussian extrapolation). The Planck constraints (1- and 2σ confidence regions) are obtained for the standard effective number of neutrinos (black solid line) and a model with a free effective number of neutrinos (black dashed lines, colour points).

4.4 RESULTS AND DISCUSSION

THE VALUES of the sound horizon and other parameters inferred from the six lenses, Pantheon SN sample and BAO measurements (BOSS) using three models that employ polynomial parametrisation or a ΛCDM model are listed in Table 4.2. The tension with *Planck* flat ΛCDM and late-time extension models is displayed in the last rows and ranges from 2σ to 3σ. When combining the distance calibration from the lensed quasars with that from SHoES (the distance ladder with Cepheids), the constraints on r_d are tighter and the tension with *Planck* increases to 5σ, as can be seen in Table 4.3. The corresponding Bayesian Information Criterion (BIC) values are the lowest for Model 4 (ΛCDM). However, the differences in BIC do not exceed 6 (substantial level on the Jeffreys scale), with a minimum of 1 for Model 1 (barely worth mentioning level on the Jeffreys scale). Figure 4.1 compares constraints on H_0, r_d and Ω_k from late-time observations including the prior from SHoES to the best fit parameters derived from Planck assuming a flat ΛCDM model. For all models, the *Planck* parameters lie on the 5σ contour in the $H_0 - r_\mathrm{d}$ plane, demonstrating that the tension is independent of the chosen expansion family.

In Table 4.4, some other combinations of data sets have been explored. This includes a calibration of lenses + CCHP instead of SHoES, inclusion of several volume-averaged and Ly-α BAO and the addition of high redshift quasars as secondary standard candles. Considering all results based on the main data sets (HoLiCOW, SN, BAO/BOSS) with the cosmic distance ladder (SHoES or CCHP), we find $r_\mathrm{d} = (137 \pm 3^{stat.} \pm 2^{syst.})$ Mpc, where the systematic error accounts for differences between SHoES and CCHP distance calibration. In addition, we run an inference free of any SN data, thus only using lensed quasars and BAO measurements from BOSS, D_V and Ly-α with a flat ΛCDM model.[5] This results in the following values for the cosmological parameters: $r_\mathrm{d} = 138.6 \pm 3.8$ Mpc, $H_0 r_\mathrm{d} = 10166 \pm 142\ \mathrm{km\,s^{-1}}$, $\Omega_\mathrm{m} = 0.29 \pm 0.02$.

4.4.1 Early-time extensions

A possible solution for the tension is an extension to the early Universe physics, such as an additional component of relativistic species. *Planck* 2018 chains with free $\mathrm{N_{eff}}$ (based on full temperature and polarisation data) have been used to investigate this scenario. In Figure 4.2, *Planck* + free $\mathrm{N_{eff}}$ is compared to results from Model 3 using SN + BAO with only the HoLiCOW lenses as calibrator (upper panel) and using a combination of HoLiCOW lenses and either SHoES or CCHP as calibrators (lower panel). A higher value of $\mathrm{N_{eff}}$ is shown to move the *Planck* value to a lower r_d and a higher H_0, therefore alleviating the tension to some extent. In this case, the combined analysis of *Planck* and low-redshift data yields $\mathrm{N_{eff}} = 3.24 \pm 0.16$. This effect is only convincing when the late-time measurements are calibrated with HoLiCOW and CCHP, since the alternative Cepheid calibration is still in tension with the *Planck*+$\mathrm{N_{eff}}$ extension (see Table 4.3).

[5] For the flat ΛCDM model, we adopted a prior of $\Omega_\mathrm{m} = \mathcal{U}[0.05, 0.5]$.

TABLE 4.2 – Posterior mean and standard deviation for the sound horizon r_d, $H_0 r_\mathrm{d}$ and q_0 inferred from late-time observations including HoLiCOW lensing observations, Pantheon SN sample and BAO measurements (BOSS). The fit quality is summarised in terms of log-likelihood at the maximum posterior probability, $\ln \mathcal{L}_{m.a.p.}$, and the Bayesian Information Criterion BIC $= \ln(N)k - 2\ln(\mathcal{L}_{m.a.p.})$, where N is the number of data points and k is the number of free parameters. The odds indicator τ quantifies the tension between r_d and H_0 measured from late-time observations and the *Planck* data (for the standard flat ΛCDM model and its two extensions with a free effective number of neutrinos or early dark energy).

	flat ($\Omega_\mathrm{k} = 0$)			
parameter	Model 1	Model 2	Model 3	Model 4 (fΛCDM)
r_d (Mpc)	132.7 ± 4.2	132.9 ± 4.4	134.2 ± 4.4	136.9 ± 3.7
$H_0 r_\mathrm{d}$ (km s^{-1})	10107 ± 147	10065 ± 150	10052 ± 152	10038 ± 136
q_0	-0.7 ± 0.07	-0.5 ± 0.2	-0.4 ± 0.3	-0.55 ± 0.03
$\ln \mathcal{L}_{m.a.p.}$	-86.3	-86.1	-86.7	-87.7
BIC score	193	196	198	192
$\ln \tau$ (Planck ΛCDM)	6.6 (3.2σ)	5.7 (2.9σ)	5.0 (2.7σ)	5.7 (2.9σ)
$\ln \tau$ (Planck ΛCDM+N$_\mathrm{eff}$)	6.3 (3.1σ)	5.6 (2.9σ)	4.9 (2.7σ)	5.0 (2.7σ)
$\ln \tau$ (Planck early DE)	5.1 (2.8σ)	4.4 (2.5σ)	3.7 (2.3σ)	3.7 (2.2σ)

	free Ω_k			
parameter	Model 1	Model 2	Model 3	Model 4 (ΛCDM)
r_d (Mpc)	129.2 ± 5.7	130.6 ± 5.9	131.2 ± 6.1	137.2 ± 4.8
$H_0 r_\mathrm{d}$ (km s^{-1})	10045 ± 155	10033 ± 157	10017 ± 160	10041 ± 156
Ω_k	0.18 ± 0.2	0.13 ± 0.2	0.15 ± 0.2	-0.01 ± 0.2
q_0	-0.6 ± 0.1	-0.4 ± 0.2	-0.4 ± 0.3	-0.56 ± 0.07
$\ln \mathcal{L}_{m.a.p.}$	-86.1	-85.9	-86.4	-87.7
BIC score	196	200	201	196
$\ln \tau$ (Planck ΛCDM)	5.6 (2.9σ)	4.6 (2.6σ)	4.0 (2.4σ)	4.2 (2.4σ)
$\ln \tau$ (Planck ΛCDM+N$_\mathrm{eff}$)	5.7 (2.9σ)	4.7 (2.6σ)	4.2 (2.4σ)	3.9 (2.3σ)
$\ln \tau$ (Planck early DE)	4.5 (2.6σ)	3.6 (2.2σ)	3.1 (2.0σ)	2.6 (1.8σ)

TABLE 4.3 – The same as Table 4.2, but for fits based on the HoLiCOW lensing, Pantheon SN sample, BAO measurements (BOSS) and H_0 from SHoES.

	flat ($\Omega_k = 0$)			
parameter	Model 1	Model 2	Model 3	Model 4 (fΛCDM)
r_d (Mpc)	135.1 ± 2.8	135.0 ± 2.9	135.1 ± 2.9	136.1 ± 2.7
$H_0 r_d$ (km s^{-1})	10079 ± 143	10055 ± 148	10038 ± 153	10037 ± 136
q_0	-0.6 ± 0.07	-0.4 ± 0.2	-0.4 ± 0.3	-0.55 ± 0.03
$\ln \mathcal{L}_{m.a.p.}$	-86.6	-86.4	-86.8	-87.7
BIC score	193	197	198	192
$\ln \tau$ (Planck ΛCDM)	15.1 (5.1σ)	15.0 (5.1σ)	13.9 (4.9σ)	15.1 (5.1σ)
$\ln \tau$ (Planck ΛCDM+N_{eff})	9.9 (4.1σ)	9.7 (4.0σ)	9.2 (3.9σ)	9.1 (3.9σ)
$\ln \tau$ (Planck early DE)	9.4 (3.9σ)	9.2 (3.9σ)	8.6 (3.7σ)	8.7 (3.8σ)
	free Ω_k			
parameter	Model 1	Model 2	Model 3	Model 4 (ΛCDM)
r_d (Mpc)	134.8 ± 3.2	134.7 ± 3.3	134.6 ± 3.3	136.1 ± 3.2
$H_0 r_d$ (km s^{-1})	10067 ± 156	10042 ± 161	10021 ± 161	10035 ± 152
Ω_k	0.04 ± 0.2	0.03 ± 0.2	0.06 ± 0.2	0.003 ± 0.2
q_0	-0.6 ± 0.09	-0.4 ± 0.2	-0.4 ± 0.3	-0.55 ± 0.07
$\ln \mathcal{L}_{m.a.p.}$	-86.7	-86.5	-86.8	-87.7
BIC score	198	201	202	196
$\ln \tau$ (Planck ΛCDM)	13.3 (4.8σ)	13.2 (4.8σ)	12.7 (4.7σ)	12.8 (4.7 σ)
$\ln \tau$ (Planck ΛCDM+N_{eff})	9.2 (3.9σ)	9.0 (3.9σ)	8.9 (3.8σ)	8.2 (3.6σ)
$\ln \tau$ (Planck early DE)	8.3 (3.7σ)	8.2 (3.7σ)	8.0 (3.6σ)	7.3 (3.4σ)

4.4.2 Tension between the CMB and late-time observations

Figure 4.3 demonstrates the potential of the selected extensions of the standard ΛCDM model outlined in Section 4.3.2 to resolve the tension between r_d and H_0 measured from the CMB and late-time observations. The shaded grey contours show constraints from late-time observations using Model 3 with $\Omega_k = 0$. Thanks to using a polynomial parametrisation, these measurements are marginalised over a wide class of the expansion history and in this sense they are independent of cosmological model. We show results for distance calibrations based on the HoLiCOW lenses combined with SHoES or CCHP. The contours in colour show constraints from Planck for the flat ΛCDM model (black contours) and its four extension.

As clearly seen from Figure 4.3, none of the ΛCDM extensions manage to convincingly unify the *Planck* measurements with the late-time ones if the SHoES calibration is used to anchor the distance ladder. In particular, late-time extensions involving different generalisations of the cosmological constant can increase the H_0 value inferred from the CMB, but they leave r_d unchanged. Although early-time extensions can potentially match both H_0 and r_d from low-redshift probes and the CMB, that this may happen by expanding the posterior proba-

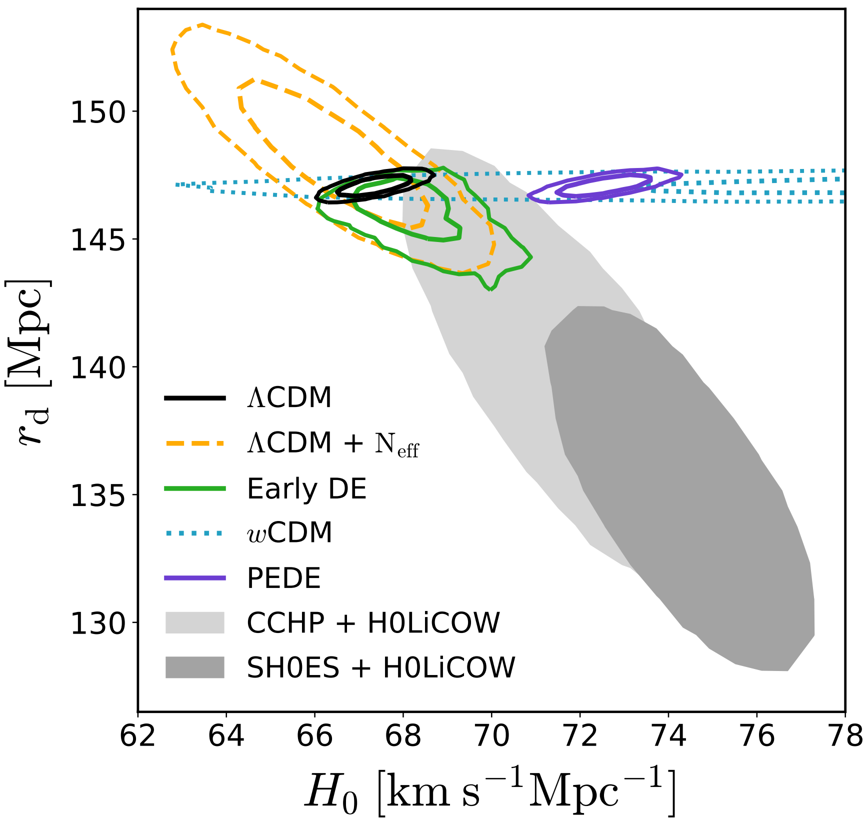


FIGURE 4.3 – The effect of four different extensions of the flat ΛCDM model on the sound horizon and the Hubble constant measured from the *Planck* data. The models considered here are ΛCDM + free N_{eff}, early dark energy, wCDM and PEDE. The CMB-based constraints are compared to the measurements from late-time observations (SN + BAO + HoLiCOW + SHoES/CCHP) shown with the grey shaded contours. The late-time measurements are obtained with Model 3 (see Table 4.1) and show the 2σ credibility regions.

bility contours rather than shifting the best fit values (see also Bernal et al., 2016; Karwal & Kamionkowski, 2016), as demonstrated in Figure 4.3. In this respect, both early dark energy models and extensions with extra relativistic species are quite similar. The apparent difference between their probability contours reflect differences in the priors. While a free effective number of relativistic species can either decrease or increase the sound horizon, early dark energy (with positive energy density) can only increase the energy budget, and thus decrease the sound horizon.

Figure 4.4 summarises the tension in the $H_0 - r_d$ plane between late-time measurements and *Planck* with different extensions of ΛCDM. To ensure a fair comparison, the same ΛCDM extensions are used in the late-time and CMB-based inference. Therefore, the *Planck* PEDE-CDM results have been compared to late-time results obtained with PEDE-CDM, and the *Planck* wCDM results to late-time results using wCDM. For the early-time extensions this is not of great

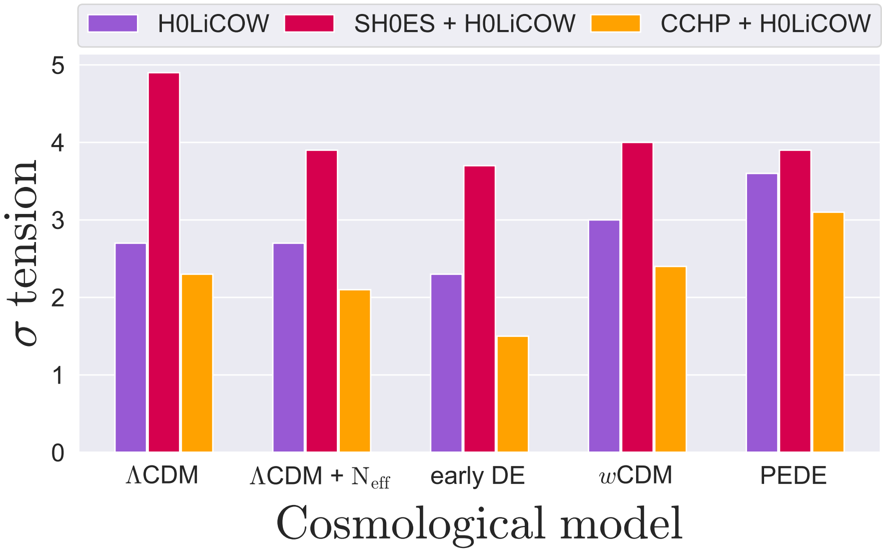


FIGURE 4.4 – Tension between the sound horizon and the Hubble constant measured from late-time observations and the CMB for the following cosmological models: ΛCDM, ΛCDM + N_{eff}, early DE, wCDM, PEDE-CDM (flatness assumed in all cases). Late-time observations include BAO, type Ia supernovae and three different absolute distance calibrations from gravitational lensing (HoLiCOW), the cosmic distance ladder with Cepheids (SHoES) or the TRGB (CCHP).

importance, since their effects do not influence the low-redshift measurements.

By adopting different models of polynomial parametrisations (Models 1, 2 and 3), we minimise the dependence on a cosmological model. Although our inference with these models does not depend on ΛCDM, it does have a weak dependency on General Relativity (GR). The lensed quasars that are used to calibrate the distance ladder need GR in order to calculate the angular diameter distance, through the *Ansatz* that the lensing potential (used in the time-delay inference) is exactly twice the gravitational potential (used to obtain $D_A \propto c^3 \Delta t/\sigma^2$ from stellar kinematics). However, the role of this GR dependence is subdominant with current D_A uncertainties (10% – 20%). On the other hand, GR also enters the early-Universe expansion through the 'abundances' of different components (Ω_m, Ω_{de}, N_{eff}).

4.4.3 One lens at a time

Since H_0 and r_d are constants, they must be independent of the chosen indicators. If they are inferred from each indicator separately, any trend will signal residual systematics, either in the indicators themselves or in the parametrisation that is chosen to extrapolate $H(z)$ down to H_0.

The HoLiCOW collaboration have shown that, if H_0 is obtained from lenses in a flat-ΛCDM model, there is a weak trend in its inferred value versus lens redshift, with lower-redshift (resp. higher-redshift) lenses being more (resp. less) discrepant with the *Planck* measurements (Wong et al., 2020). Even though this trend is currently not significant (given current uncertainties), it may be indicative of intrinsic systematics in the lensing inference, or in the way that time-delay distances are converted into H_0 values through a flat-ΛCDM parametrisation.

Here we repeat this test using more general models of the expansion history, specifically flat Model 3 and flat PEDE-CDM model. Figure 4.5 shows the sound horizon r_d measured from combining BAO and SNe data with lensing constraints from each lens separately. The results demonstrate that the distance calibration from HoLiCOW lenses shows a similar trend with lens redshift as the one shown by Wong et al. (2020) for a flat ΛCDM cosmology. Based on the sample-wide analysis by Millon et al. (2019), this weak trend cannot be explained simply on the basis of known systematics in the lens models or kinematics of each lens. We should emphasise, however, that this trend is *not* statistically significant (1.6σ) yet.

Although the current weak trend of r_d with redshift of gravitatonal lens is consistent with being a statistical fluke, it is instructive to investigate if there any expansion models that can re-absorb this (weak) trend. For example, a recent ($z \approx 0.4$) change in dark energy may produce this behaviour, if the data are interpreted with expansion histories that are 'too' smooth. For this reason, we examine the same lens-by-lens determination within the PEDE model family. The results are shown as dotted error-bars in Figure 4.5. Even the PEDE model with accelerated late-time expansion cannot eliminate the (weak) trend in r_d. The constraints set by the relative distance moduli of SN enforce PEDE to closely resemble the ΛCDM case, but with a higher matter content ($\Omega_\mathrm{m} \approx 0.345$) and smaller sound horizon ($r_\mathrm{d} \approx 138$ Mpc). Therefore, PEDE does not resolve the current tension.

4.5 CONCLUSIONS AND OUTLOOK

We have combined the newest available low-redshift probes to obtain an estimate of the sound horizon at the drag epoch, r_d. In order to minimise the dependence on a cosmological model, we have used a set of polynomial parametrisations that are almost entirely independent of the underlying cosmology, as well as the standard ΛCDM model. In the $H_0 - r_\mathrm{d}$ plane, we have found a tension of 5σ between *Planck* results using flat ΛCDM and late-time observations calibrated with HoLiCOW lenses and SHoES. This tension reduces to 2.4σ if CCHP results are used as a distance-ladder anchor instead of SHoES. We have investigated whether early- or late-time extensions to the standard ΛCDM model can resolve the tension and examined models with free $\mathrm{N_{eff}}$, early dark energy, wCDM and PEDE-CDM. None of these model extensions provide a satisfying solution to the Hubble tension problem (see also Aylor et al., 2019; Knox & Millea, 2020), except for free $\mathrm{N_{eff}}$ or early dark energy in combination with low redshift data calibrated by CCHP + HoLiCOW.

These findings can indicate that: (1) extensions of early-time physics are necessary; and/or (2) that systematics from different late-time probes are becoming comparable to the statistical uncertainties. Arguments based on local under-densities or peculiar velocities cannot resolve the tension: the $\approx 3\sigma$ tension persists if the inverse-distance ladder is restricted to $z \geq 0.03$, where the role of peculiar velocities is $\lesssim 0.1\%$ (see also Wojtak & Agnello, 2019). Multiple secondary sources of errors in redshift measurements were studied by Davis et al. (2019), but none of them seem to have any noticeable effect. Another explanation may be that the standardisation

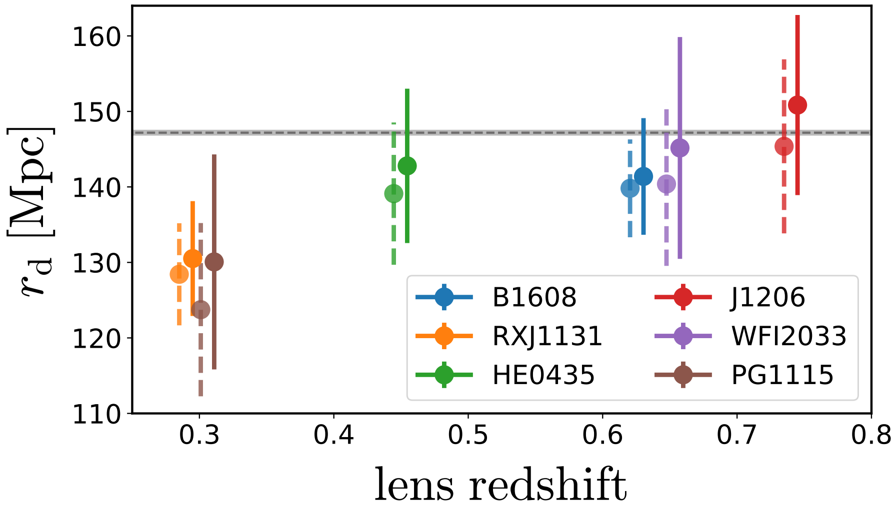


FIGURE 4.5 – The sound horizon r_d measured from combining BAO and SNe data with HoLiCOW lensing observations of each lens separately. Here the distance calibration is set solely by the lensing observations of each individual lens. The measured sound horizon is shown as a function of lens redshift for fits with a flat Model 3 (solid error bars) and a flat PEDE-CDM model (dashed error bars). For both models, the measurements show a slight trend of r_d increasing with lens redshift. The inference from Models 1 and 2 is fully consistent with the Model 3 results. The grey dashed line with shaded region shows *Planck*'s value of r_d and its (sub-percent) uncertainty obtained for the standard flat ΛCDM model.

of SNe Ia is not properly understood yet (as a caveat see e.g. Rigault et al., 2015, also Khetan et al. 2020), or that there is some (hitherto undiscovered) source of systematics in one of the other used data sets. If all astrophysical systematics are exhausted, one can also consider proposals involving non-standard physics in the local Universe such as screened fifth forces, which may bias H_0 measurements high via modulation of gravity-dependent pulsation periods of Cepheids (for more details see Desmond et al., 2019). For these reasons, we also provide a measurement that relies only on lenses and BAO, without any additional constraint from SNe, in Section 4.4.

The weak trend in Figure 4.5 may indicate residual systematics in the lens models, or the need for different low-z expansion models, or it may vanish entirely with larger lens samples. In order to check the robustness of the trend, cosmography-grade models of more lenses are needed, over the whole $0.3 \lesssim z \lesssim 0.7$ current redshift interval and beyond. Finally, the role of systematics in the lens mass models can be assessed once high-S/N spatially-resolved kinematics are available (Shajib et al., 2018; Yıldırım et al., 2020), which would enable more flexible dynamical models than the ones used so far on aperture-averaged velocity dispersions.

As a final remark, we emphasise that resolving the H_0 tension alone is not sufficient, since different models that can shift this value are still at tension with the inferred r_d from BAO and low-redshift indicators. Also, a direct combination of the inference from late-time and CMB-based measurements that may be at $> 3\sigma$ tension, hence hardly compatible with one another, should be justified. Therefore, any new proposal to resolve the discrepancy between

TABLE 4.4 – The same as Table 4.2, but for various combinations of late-time observations including two local determinations of H_0 (SHoES or CCHP), measurements of isotropic BAO (D_V) and anisotropic BAO from the Lyman-α forest of quasars (Ly-α), and estimates of distance moduli from high-redshift quasars. The parameters are determined using Model 3 with $\Omega_k = 0$.

flat ($\Omega_k = 0$)		
parameter	**CCHP + HoLiCOW + SN + BAO (BOSS)**	**SHoES + HoLiCOW + SN + BAO (BOSS + D_V + Ly-α)**
r_d (Mpc)	139.5 ± 3.6	138.1 ± 2.7
$H_0 r_d$ (km s^{-1})	10019 ± 152	10197.1 ± 135
q_0	-0.4 ± 0.4	-0.9 ± 0.3
$\ln\tau$ (Planck ΛCDM)	3.8 (2.3σ)	12.8 (4.7σ)
$\ln\tau$ (Planck ΛCDM+N$_{\rm eff}$)	3.4 (2.1σ)	8.0 (3.6σ)
$\ln\tau$ (Planck early DE)	2.1 (1.5σ)	7.2 (3.4σ)
flat ($\Omega_k = 0$)		
parameter	**HoLiCOW + SN + BAO (BOSS + D_V + Ly-α)**	**SHoES + HoLiCOW + SN + BAO (BOSS) + high-z quasars**
r_d (Mpc)	138.6 ± 3.8	134.0 ± 2.8
$H_0 r_d$ (km s^{-1})	10191 ± 138	10011 ± 149
q_0	-0.8 ± 0.3	-0.2 ± 0.3
$\ln\tau$ (Planck ΛCDM)	4.4 (2.5σ)	17.1 (5.5σ)
$\ln\tau$ (Planck ΛCDM+N$_{\rm eff}$)	4.2 (2.4σ)	11.0 (4.3σ)
$\ln\tau$ (Planck early DE)	2.8 (1.9σ)	10.8 (4.3σ)

CMB-based and late-time measurements should consider both H_0 and r_d, and examine the *separate* inference upon late-time and CMB-based data.

SUPPORTING MATERIAL

Appendix A.1: Planck compressed likelihood
Appendix A.2: Convergence tests of polynomial parametrisations

DATA AVAILABILITY

The source code repository, containing python scripts to generate the results and figures of this paper, is available at https://github.com/Nikki1510/cosmic_dissonance.

AUTHOR CONTRIBUTIONS

NA led the project, performed the analysis, produced the figures, revised the paper after comments from the referee, wrote the manuscript. RW proposed the main idea, implemented the extensions of ΛCDM in the *Planck* compressed likelihood, validated the results, revised the paper after comments from the referee, contributed to writing up the manuscript. AA proposed the main idea, provided the HoLiCOW data sets, validated the results, contributed to writing up the manuscript. GC, CF, DS, SH, MM, VB, KW, FC, SS, SB, TT & LK provided the HoLiCOW data sets and delivered valuable feedback on the manuscript.

ACKNOWLEDGEMENTS

We thank Chiara Spiniello for useful comments on an earlier version of this manuscript, and Inh Jee and Eiichiro Komatsu for discussions on the lensing distance likelihoods. These time delay cosmography observations are associated with programs HST-GO-9375, HST-GO-9744, HST-GO-10158, HST-GO-12889, and HST-14254. Support for programs HST-GO-10158 HST-GO-12889 HST-14254 was provided to members of our team by NASA through grants from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. AA and RJW were supported by a grant from VILLUM FONDEN (project number 16599). This project is partially funded by the Danish council for independent research under the project "Fundamentals of Dark Matter Structures", DFF–6108-00470. This project has received funding from the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (COSMICLENS; grant agreement No. 787886) and from the Swiss National Science Foundation (SNSF). This work was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. SH acknowledges support by the DFG cluster of excellence 'Origin and Structure of the Universe' (www.universe-cluster.de). CDF acknowledges support for this work from the National Science Foundation under Grant No. AST-1715611. SHS thanks the Max Planck Society for support through the Max Planck Research Group. TT acknowledges support by the Packard Foundation through a Packard Research fellowship and by the National Science Foundation through NSF grants AST-1714953 and AST-1906976. LVEK is partly supported through an NWO-VICI grant (project number 639.043.308).

II

Gravitationally lensed supernovae

CHAPTER 5

Finding gravitationally lensed supernovae

This chapter is based on the following article:
"AI-driven spatio-temporal engine for finding gravitationally lensed supernovae"

Under review at Nature Astronomy.
ArXiv e-prints: 2107.12399.

Authors: Doogesh Kodi Ramanah, Nikki Arendse & Radosław J. Wojtak.

Abstract

We present a spatio-temporal AI framework that concurrently exploits both the spatial and time-variable features of gravitationally lensed supernovae in optical images to ultimately aid in the discovery of such exotic transients in wide-field surveys. Our spatio-temporal engine is designed using recurrent convolutional layers, while drawing from recent advances in variational inference to quantify approximate Bayesian uncertainties via a confidence score. Using simulated Young Supernova Experiment (YSE) images as a showcase, we find that the use of time-series images yields a substantial gain of nearly 20 per cent in classification accuracy over single-epoch observations, with a preliminary application to mock observations from the Legacy Survey of Space and Time (LSST) yielding around 99 per cent accuracy. Our innovative deep learning machinery adds an extra dimension in the search for gravitationally lensed supernovae from current and future astrophysical transient surveys.

5.1 INTRODUCTION

STRONG gravitational lensing of astrophysical transient sources, such as supernovae and active galactic nuclei (AGN), involves bending of light from the background source as the light travels towards the observer through a lens galaxy, producing multiple images of the strongly lensed object. Time delays between these images are particularly sensitive to the cosmic expansion rate and observations of several gravitationally lensed quasars were successfully used to place competitive constraints on the Hubble constant (Birrer et al., 2020; Wong et al., 2020). Time delay distances combined with observations of type Ia supernovae were also used to measure curvature of the Universe (Arendse et al., 2019; T. Collett et al., 2019).

Type Ia supernovae, as considered in this work, possess several advantages over AGN, such as quasars, as time delay indicators. Since they are standardisable candles, they can be used to directly compute the lensing magnification, thereby mitigating issues posed by the degeneracy between the lens potential and H_0 (Oguri & Kawano, 2003). With exceptionally well-characterised light curve morphology, extracting the time delays is less complicated relative to quasars which display significant variation in their spectral sequences (Nugent et al., 2002; Pereira et al., 2013; S. A. Rodney et al., 2016). Finally, after the supernova has faded away, follow-up observations may be conducted to better constrain the lens galaxy properties, such as the mass profile, without contamination from the supernova (Ding et al., 2021). On the other hand, the expected time delays from type Ia supernovae are typically much shorter and image separations are smaller than in known lensed quasars, making the precise measurement of time delays more challenging.

Lensed supernovae remain, nevertheless, exotic astrophysical objects, with the challenges inherent to their discovery rendering them extremely difficult to observe. The two main indicators are the multiply-imaged lensing signature (*image multiplicity*) (Oguri & Marshall, 2010) and exceptionally amplified (*magnification*) flux (Goldstein & Nugent, 2017). As of date, only three multiply-imaged resolved systems have been discovered (Goobar et al., 2017; Kelly et al., 2015; S. Rodney et al., 2021), with two of them lensed by galaxy clusters and one by a galaxy lens, along with a handful of highly magnified unresolved detections (Amanullah et al., 2011; Patel et al., 2014; Quimby et al., 2014; S. A. Rodney et al., 2015), all discovered serendipitously rather than by means of dedicated searches. A major impediment is that these supernovae are visible for at most hundred days after explosion, in contrast to the much longer timescales of quasars. Consequently, it is of paramount importance to flag lensed candidates while the supernova is still active so that high-resolution imaging or spectroscopy may be used to confirm the lensed nature and to measure a time delay, thereby further limiting the window of opportunity. The situation is yet exacerbated due to most strong gravitational lenses yielding image separations that are below the resolution of ground-based optical surveys (Oguri, 2006).

Given the scarcity of gravitationally lensed astrophysical transients and the typical volume of imaging data sets, machine learning (ML) techniques, adept at detecting subtle patterns and correlations in data, provide a natural solution to this problem. As such, various ML algorithms (Avestruz et al., 2019; Cañameras et al., 2020; Cheng et al., 2020; Davies et al., 2019; Gentile et al., 2021; X. Huang et al., 2021; X. Huang et al., 2020; Lanusse et al., 2018;

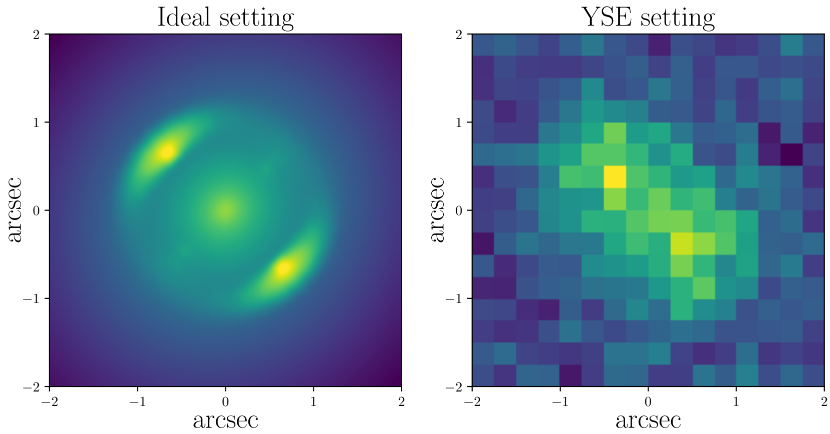


FIGURE 5.1 – Contrast between a simulated gravitationally lensed supernova as imaged by a perfect instrument (*left panel*) and the Young Supernova Experiment (YSE, *right panel*), highlighting the challenging nature of discovering lensed supernovae from YSE observations. The differences between the idealised and YSE settings lie mainly in the resolution, background noise level and quality of seeing, as characterised by the instrument's PSF.

Schaefer et al., 2018) for finding strongly lensed systems, particularly well-suited for hosting lensed supernovae or quasars, in photometric surveys have been recently developed. These ML studies typically employ deep convolutional neural networks that are tuned to identify the characteristic signature of the arc features constituting the Einstein rings.

In this work, we aim to harness the power of artificial intelligence (AI), in particular, ML algorithms, to design a new tool specifically tailored for finding gravitationally lensed supernovae from the Young Supernova Experiment (D. O. Jones et al., 2021). Our approach introduces some innovative features with respect to the existing ML-based lens finders to cope with the distinct challenges posed by the Young Supernova Experiment, as outlined below. Our proposed methodology is, nevertheless, also of practical relevance for other current or next-generation surveys, as long as the images constituting the training set accurately emulates the characteristics of the actual observations.

The Young Supernova Experiment (D. O. Jones et al., 2021) (hereafter YSE) constitutes a recent endeavour of a novel optical time-domain survey on the Pan-STARRS telescopes. YSE is geared towards the discovery of fast-rising supernovae within a few hours to days of explosion, thereby complementing other currently ongoing surveys. Moreover, YSE is presently the only time-domain survey with observations in four bands, with the capacity to discover faint transients (~ 21.5 mag in *gri* and ~ 20.5 mag in *z*) that allow the study of the earliest phases of stellar explosions. According to survey forecasts from simulations, the full-capacity operation of YSE will result in the discovery of around 5000 new supernovae per year, with at least a couple of them within three days of explosion per month. The latter YSE attribute is especially pertinent as it is essential to identify new lensed supernova candidates sufficiently early so as to initiate the follow-up sequence in a timely fashion. The substantial volume of new observations from YSE is, therefore, an exciting hunting ground for lensed supernovae.

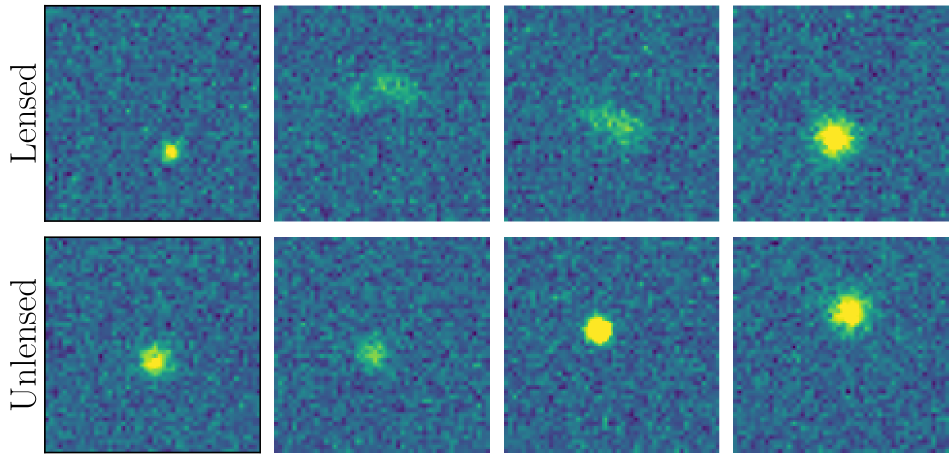


Figure 5.2 – A random set of lensed (*top row*) and unlensed (*bottom row*) supernovae simulated according to YSE i-band specifications. The colour scale is anchored for all images, with each image stamp extending over 12 arcsecs. Our image simulation pipeline is designed such that the properties of the mock unlensed images closely match those of the actual YSE observations. These two rows illustrate that most lensed sources in YSE will be unresolved, thereby appearing similar to unlensed ones. Since the YSE data pipeline provides difference images of transients directly, i.e. the subtraction of a historic reference image from a newly observed image, we do not incorporate the lensing galaxy in our image simulation pipeline.

Such an undertaking is, however, riddled with challenges in the YSE context. It is extremely unlikely that YSE will fully resolve the particularly conspicuous signature of the Einstein rings in lensed systems. This is mainly due to the instrumental capabilities of the Pan-STARRS telescopes, such as the point spread function (PSF) and the typical background noise level. As an illustration of the innate difficulty of this task, Figure 5.1 depicts the stark contrast between a gravitationally lensed supernova in an ideal setting with almost perfect observing conditions, i.e. remarkably high resolution, very sharp PSF and non-existent background noise, and the same source as observed in typical YSE settings. Furthermore, unresolved or partially resolved lensed images are a common feature for YSE-like observations. An additional limitation emanates from the fact that for typical lensing configurations predicted for the survey, the second brightest image has flux near or below the detection limit (Wojtak et al., 2019). As a consequence, the aforementioned ML-based lens finders are not immediately adequate for our particularly strenuous problem due to the absence of the distinctive features of such lensed systems. Figure 5.2, displaying a random sample of simulated lensed and unlensed (i.e. normal) supernovae further corroborates the challenging nature of finding lensed supernovae from the YSE survey.

To make tangible progress in the search for lensed supernovae in the YSE context, there is a pressing need for highly effective detection algorithms. To this end, we design a neural network that is sensitive to both spatial and temporal correlations and takes into account the temporal evolution of an astrophysical transient over multiple epochs, in addition to its spatial information from a single-epoch image. An example of an extra distinctive trait, even for unresolved images, is a spatial shift in the light centroid of the lensed point sources over sequential epochs, contrary to unlensed sources. Searches for sources with spatially extended and time-variable sources were also proposed as a method to select lensed quasar candidates (Chao et al., 2020; Kochanek et al., 2006; Lacki et al., 2009). This constitutes the rationale behind our novel approach where we employ time-series images in training a spatio-temporal engine, based on recurrent convolutional layers, to perform a binary classification of lensed and unlensed supernovae. We build on recent advances pertaining to variational inference to yield approximate Bayesian uncertainties on the neural network predictions. A measure of the network confidence associated with a given prediction yields an additional metric to aid in swiftly pinpointing promising lensed candidates that can be prioritised during the subsequent human vetting step. The above innovative aspects distinguish our approach from previous ML-based lens finders proposed in the literature.

The remainder of this chapter is organised as follows. Section 5.2 provides an overview of the methods employed in this work, including all relevant details pertaining to the background, numerical implementation and optimisation of our Bayesian neural classifiers, the generation of simulated lensed and unlensed YSE-like time-series images constituting our training, validation and test sets, and the time-series image compression algorithm. In Section 5.3, we validate and demonstrate the performance of the classifier on YSE and LSST data. Finally, we provide an overview of the main findings of our work in Section 5.4.

5.2 METHODS

To illustrate the improvement in classification accuracy, by virtue of the informative features encoded in the respective neural network inputs, we design and implement three separate neural classifiers:

(A) Single-epoch model trained using single-epoch images randomly selected from the time-series images;
(B) Compressed temporal model trained using compressed 2D temporal representations of the time-series images;
(C) Spatio-temporal model trained using all time-series images.

All above ML models output a network probability score, but are fed distinct types of input images. The time-series images may be unevenly distributed in time. To obtain the compressed temporal representation, we implement a variant of the smooth manifold extraction algorithm (Shihavuddin et al., 2017), recently proposed in the field of medical image processing, to combine the arbitrary number of time-series images per source into a single informative image. This section begins by outlining the image simulation procedure, followed by a description of the smooth manifold extraction algorithm, and finally, provides details about the neural classifiers presented above.

Image property	YSE	LSST
PSF (FWHM)	$\mathcal{N}^{S}(5, 0.95, 0.375)$	$\mathcal{N}^{S}(4, 0.55, 0.299)$
Pixel size ($''$)	0.25	0.2
Pixel background noise ($\sigma_{\rm bkg}$)	6.5	6.1
Exposure time (s)	27	30
Image stamp size ($''$)	12.0	9.6
Image stamp area (arcsec^2)	144	92
Number of pixels	48×48	48×48
Cadence (days; i-band)	6	9
Zero-point magnitude (i-band)	24.74	27.79
Limiting magnitude (i-band)	21.40	23.90

TABLE 5.1 – **Simulated image characteristics.** The images constituting the training, validation and test sets are generated in accordance with the YSE survey specifications, thereby closely emulating real YSE observations. The image properties for our preliminary LSST set-up are also provided. $\mathcal{N}^{S}(a, \mu, \sigma)$ denotes a skew-normal distribution with skewness parameter a.

5.2.1 Image simulation procedure

We employ the multi-purpose lens modelling package LENSTRONOMY [1] (Birrer & Amara, 2018) to generate images that accurately reflect the actual YSE observations. The YSE survey is designed to acquire well-sampled light curves in four bands (*griz*) for several thousands of transient astrophysical sources up to a redshift of around 0.2. The current field of view of YSE is approximately 750 square degrees of the sky with a cadence of three days, with the area covered by the survey to be doubled in the near future, with a median seeing (FWHM of PSF) of 1.28 arcsec. The image properties and YSE-like survey characteristics, as adopted in this work, are detailed in Table 5.1. YSE utilises the *difference imaging* technique for the identification of new sources in optical images, whereby a historic reference image is subtracted from a given input image to excise the static and non-varying sources, such that transient sources manifest as residual flux that can be detected and measured photometrically with conventional methods.

We consider a supernova as a point source in its host galaxy. Since YSE provides difference images that contain only the astrophysical transients, we simulate images that consist of solely the lensed or unlensed supernova, i.e. the host and lens galaxies are not represented. As such, their respective light profiles are not of relevance to our image simulation pipeline. We assume a standard ΛCDM cosmological model, as characterised by the latest best-fit values from the Planck Collaboration (Planck Collaboration, Aghanim, et al., 2018). While the actual YSE image stamp extends over 75 arcsecs in length, we consider only the central portion covering an area of 144 arcsec^2, thereby obviating image artefacts, such as strips of dead pixels, in the real observations. This cutout size is sufficiently large to encode all relevant lensing features.

[1] https://lenstronomy.readthedocs.io/en/latest/

Parameter	Distribution
Lens redshift	$z_{\text{lens}} \sim \mathcal{N}(0.4, 0.1)$
Lensed source redshift	$z_{\text{src}} \sim \mathcal{N}^{\mathcal{S}}(2.5, 0.67, 0.1)$
Unlensed source redshift	$z_{\text{src}} \sim \mathcal{N}^{\mathcal{S}}(5, 0.08, 0.1)$
Lensed source position ($''$)	$\alpha, \delta \sim \mathcal{U}(-\theta_{\text{E}}, \theta_{\text{E}})$
Unlensed source position ($''$)	$\alpha, \delta \sim \mathcal{N}(0, 0.65)$
Lens galaxy	
Elliptical power-law mass	
Lens centre ($''$)	$x_{\text{lens}}, y_{\text{lens}} \equiv (0, 0)$
Einstein radius ($''$)	$\theta_{\text{E}} \sim \mathcal{N}^{\mathcal{S}}(8, 0.35, 0.75)$
Power-law slope	$\gamma_{\text{lens}} \sim \mathcal{N}(2.0, 0.1)$
Axis ratio	$q_{\text{lens}} \sim \mathcal{N}(0.7, 0.15)$
Orientation angle (rad)	$\phi_{\text{lens}} \sim \mathcal{U}(-\pi/2, \pi/2)$
Environment	
External shear modulus	$\gamma_{\text{ext}} \sim \mathcal{U}(0, 0.05)$
Orientation angle (rad)	$\phi_{\text{ext}} \sim \mathcal{U}(-\pi/2, \pi/2)$
Light curve	
Stretch	$x_1 \sim \mathcal{N}^{\mathcal{S}}(-3.6, 0.96, 1.2)$
Colour	$c \sim \mathcal{N}^{\mathcal{S}}(5.5, -0.11, 0.13)$
Absolute magnitude	$M_{\text{abs}} \sim \mathcal{N}(-19.43, 0.12)$
Milky Way extinction	$E(B-V) \sim \mathcal{U}(0, 0.1)$

TABLE 5.2 – **Parameter distributions for lensed and unlensed systems.** The distribution of input parameters employed in the image simulation pipeline to generate the training, validation, and test data sets. $\mathcal{N}(\mu, \sigma)$ corresponds a normal distribution with mean μ and standard deviation σ, $\mathcal{N}^{\mathcal{S}}(a, \mu, \sigma)$ implies a skew-normal distribution with skewness parameter a, while $\mathcal{U}(x, y)$ denotes a uniform distribution with bounds x and y.

A key component in our image generation pipeline is the choice of a model profile with the prerequisite flexibility to sufficiently characterise lensing systems in practice. Previous ML-based studies involving lens modelling made use of the singular isothermal ellipsoid (SIE) lens mass profile (Hezaveh et al., 2017; Madireddy et al., 2019; Maresca et al., 2021; Pearson et al., 2019; Perreault Levasseur et al., 2017; Schuldt et al., 2021). An extension to this model is the power-law elliptical lens mass distribution (PEMD) (Barkana, 1998; Kormann et al., 1994), where the 3D power-law mass slope γ_{lens} is allowed to vary. The PEMD model has been adopted in recent studies (Park et al., 2021; Pearson et al., 2021; Wagner-Carena et al., 2021), and we employ the same model in our work. The PEMD profile can be expressed in terms of six parameters, as given by eq. (1.81). We also account for the external shear component that is characterised by the shear modulus γ_{ext} and the shear angle ϕ_{ext}. To circumvent issues arising from cyclic boundary conditions due to the 2π-periodic property of the angles, the target lens mass ellipticity and external shear are expressed as follows:

$$e_1 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \cos(2\phi_{\text{lens}}) \tag{5.1}$$

$$e_2 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \sin(2\phi_{\text{lens}}) \tag{5.2}$$

$$\gamma_1 = \gamma_{\text{ext}} \cos(2\phi_{\text{ext}}) \tag{5.3}$$

$$\gamma_2 = \gamma_{\text{ext}} \sin(2\phi_{\text{ext}}). \tag{5.4}$$

To generate an ensemble of simulated YSE-like images, we sample the relevant parameters from their respective distributions, as detailed in Table 5.2, motivated by observational and theoretical considerations via Monte Carlo simulations (Oguri & Marshall, 2010; Wojtak et al., 2019). Specifically, we employ a joint probability distribution of Einstein radii, lens galaxy redshifts and supernova redshifts for strongly lensed type Ia supernovae detectable in YSE. Using this probability distribution as a prior, we generate lensing configurations by drawing random positions of the source within the area of strong lensing, while keeping the position of the lens fixed to the centre of the image. For every lensing configuration, we then verify whether the total magnification enables detection of the lensed supernova in i-band. The peak magnitude (without magnification) is calculated assuming type Ia supernovae with light curves simulated as outlined in the next section. Given a particular lensing configuration, the in-built lens equation solver in LENSTRONOMY can be used to compute the positions of the multiple lensed images of the supernova and their associated magnifications. We also account for a microlensing (Dobler & Keeton, 2006) effect, whereby the strongly lensed supernova is microlensed by the stars in the lens galaxy, by incorporating a stochastic perturbation in the computed magnification. The perturbation does not vary in time and thus, it neglects the evolution of supernova photosphere, which may additionally modify the shape of light curves and change the actual time delays measured from the arrival time of the light curve peaks (Pierel & Rodney, 2019). However, keeping in mind that the secondary images of lensed supernovae expected in YSE are just around or below the detection limit, the adopted microlensing model is sufficient for the purpose of generating reliable images in the conditions limited by the survey's design. The stochastic perturbation in the macrolens magnification field is computed assuming a Gaussian distribution with a standard deviation of 0.05, similar to what was assumed in Park et al. (2021) for generating simulated images of gravitationally lensed AGN.

In this work, we consider only i-band images, and we therefore adopt the zero-point magnitude for i-band (Tonry et al., 2012) in line with the Pan-STARRS1 specifications. Unlensed supernovae have a magnification of unity and are rendered at the position of the host galaxy. Figure 5.2 depicts a random set of simulated realisations of lensed and unlensed supernovae, generated in accordance with YSE i-band specifications so as to closely resemble the actual YSE observations. An analogous procedure is used to simulate LSST-like observations with the image properties specified in Table 5.1.

5.2.2 Simulating time-series images

In order to closely emulate the YSE data output, which consists of multiple images taken at different observational epochs, we generate time series of images. Compared to single-epoch observations, this provides the advantage of fully capturing the multiplicity of the lensed sources, even when the multiple lensed images do not all appear simultaneously.

We model the supernova variability using synthetic light curves. In this work, we only consider type Ia supernovae, since their characteristic light curves are easy to model and they constitute a large fraction ($\sim 40\%$) of the predicted Pan-STARRS lensed supernova population (Wojtak et al., 2019). The light curves are simulated using SNCosmo[2](Barbary et al., 2016) and its built-in parametric light curve model SALT2(Guy et al., 2007), which takes as input an amplitude parameter x_0, stretch parameter x_1, and a colour parameter c. We sample the x_1 and c parameters from asymmetric Gaussian distributions (D. Scolnic & Kessler, 2016) that have been derived for the Pan-STARRS1 data release(Rest et al., 2014). Then, the Tripp formula(Tripp, 1998) provides a relation for the absolute B-band peak magnitude M_B that a type Ia supernova, based on its stretch and colour parameters, is expected to have:

$$M_B = -\alpha x_1 + \beta c + M_{abs}, \tag{5.5}$$

where $M_{abs} \sim \mathcal{N}(-19.43, 0.12)$ is the expected absolute magnitude of a supernova with $x_1 = c = 0$, following the standard Planck 2018 ΛCDM calibration(Planck Collaboration, Aghanim, et al., 2018). The coefficients $\alpha = 0.14$ and $\beta = 3.1$ (D. Scolnic & Kessler, 2016) specify the correlation of absolute magnitude with the stretch and colour parameters, respectively. We assume a Milky Way dust extinction model (Fitzpatrick, 1999), with optical total-to-selective extinction ratio $R_V = 3.1$ and low $E(B-V)$ values, since YSE chooses fields with high Galactic latitude and low Milky Way extinction. The adopted distribution for Milky Way dust extinction and other input parameters to the light curve generation routine are provided in Table 5.2.

After the light curves have been generated, they are used to model the apparent magnitude (including K-corrections) for both the lensed and unlensed sources. For the lensed systems, the brightness of each image follows the variability of the light curve, with a correction for the computed magnification, stochastic microlensing perturbations and time delays. Finally, we transform the apparent magnitude to data counts per second using the zero-point magnitude of the instrument, which, when multiplied with the exposure time, yields the amplitude in the desired units for LENSTRONOMY. The observational epochs for each configuration are sampled with a 6-day cadence in the time interval that the observed supernova is brighter than the limiting Pan-STARRS magnitude, resulting in a varying number of images per system. The temporal evolution of a given lensed source, as observed under YSE-like settings, is indicated in Figure 5.3. The Mock Lenses in Time software package (Vernardos, 2021), published towards the end of our study, presents another alternative for generating time-series images and incorporates a more rigorous treatment of microlensing.

To optimise the three neural networks, we generate a balanced training data set, i.e. with around 50 per cent lensed and unlensed sources, containing a total of 40000 realisations, where

[2] https://sncosmo.readthedocs.io/en/stable/

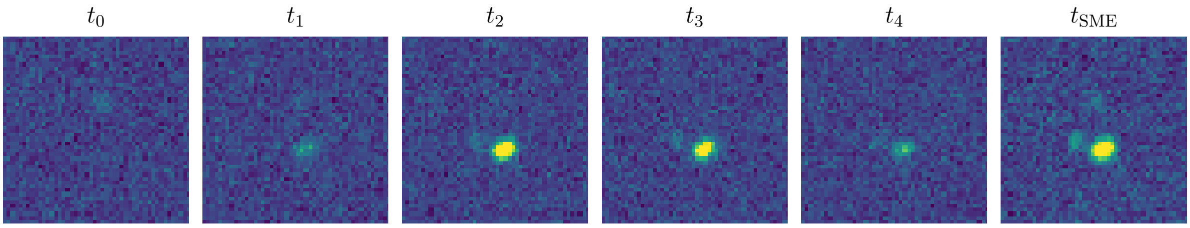


FIGURE 5.3 – Temporal evolution of the image stamp, as observed by YSE, of a simulated realisation of a lensed supernovae. The image on the far right corresponds to the compressed time-series representation via smooth manifold extraction (SME).

20 per cent is kept for validation. For model performance evaluation, we generate a separate test set consisting of two subsets containing purely lensed and unlensed sources, respectively, with each subset holding 10000 realisations. The images are simulated as detailed in the previous section, along with an extra layer of complexity to incorporate the temporal evolution aspect in the pipeline. The number of time-series observations for each source will depend on its light curve variation and YSE cadence. A new value of PSF is drawn for each multiple-epoch observation as we do not expect correlations in PSF due to the low YSE cadence. Our simulated data set consists solely of sources whose apparent magnitudes are within the YSE (i-band) limiting magnitude. To train the single-epoch model, single-epoch observations are randomly drawn from the time-series observations of each source. For the compressed temporal model, we employ the smooth manifold extraction technique, as described in the following section, to combine the series of an arbitrary number of images per source into a single informative image. The spatio-temporal model, in contrast, is trained using all the time-series images of a given source, with arbitrary time intervals between the single-epoch observations.

For real data applications, the preprocessing of the images as obtained from the YSE data pipeline is straightforward, as the YSE pipeline already provides difference images that are centred on the transient. The only requirement is to extract the central portion extending over 12 arcsecs, in accordance with the image stamp size of the images in our training set.

5.2.3 Smooth manifold extraction

In order to compress our time-series of images into a single 2D representation, we employ the technique of smooth manifold extraction (Shihavuddin et al., 2017) (SME). The SME algorithm is a technique that allows the extraction of the signal embedded in a stack of 2D images, while preserving the local spatial relationships in the original volume of data. It was originally proposed in the field of medical image processing to compress the series of images obtained via fluorescence microscopy into a single 2D representation. The underlying motivation was to improve upon the standard maximum intensity projection (MIP) method that extracts a discontinuous layer of pixels from a 3D stack of images, thereby resulting in artefacts in the final compressed or projected image that may lead to misleading interpretations.

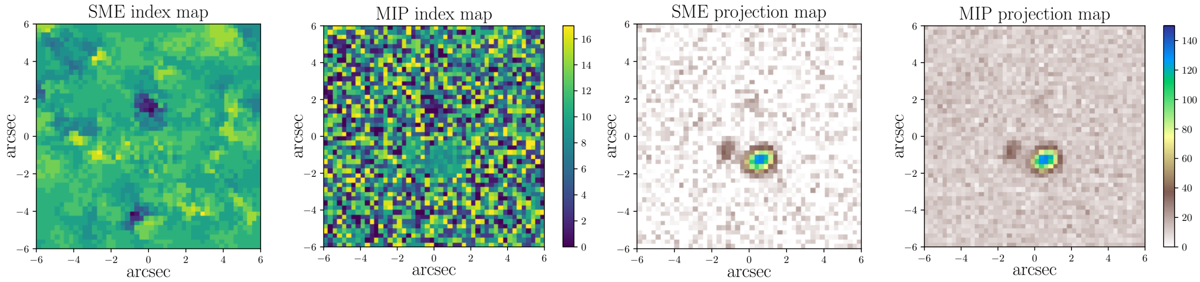


FIGURE 5.4 – Comparison between smooth manifold extraction (SME) and maximum intensity projection (MIP) in terms of their respective temporal index maps and corresponding projection maps. The SME index map, unlike the highly discontinuous MIP version, smoothly extracts the signal embedded in the stack of time-series images, thereby enforcing spatial consistency. Consequently, the compressed SME image representation shows a significantly improved contrast with respect to its MIP counterpart. Moreover, the background properties of the SME image are not influenced by the number of time-series images for a given transient, such that this does not artificially bias the neural classifier during training.

In this work, we implement a variant of the original SME algorithm, tailored for our specific problem, in order to scan the time-series images of a given source and extract the signal without modifying the image properties, such as the background noise, of single-epoch observations. In contrast, the implementation of MIP is straightforward as it consists in retrieving the level of maximum intensity along the temporal axis for each (x, y) spatial position, with the map of levels referred to as the index map $\mathbf{Z}$. MIP is not suited to our classification problem as it biases the network learning process. This is due to the fact that the simulated lensed sources have, on average, a higher number of time-series snapshots than the unlensed ones, such that MIP results in distinct background characteristics for the two cases.

The rationale of the SME optimisation routine is to fit a smooth 2D manifold onto the foreground signal, while disregarding the background, thereby propagating the index map from the foreground to the local background. To this end, we constrain each pixel in the index map by minimising the distance between the highest intensity region and the local variance of the index map. The former ensures foreground proximity of the index map, while the latter enforces its smoothness. The optimal index map is, therefore, obtained by minimising the cost function:

$$\mathbf{Z}_{\mathrm{SME}} = \underset{\mathbf{Z}}{\operatorname{argmin}} \sum_{(x,y)} \mathcal{W}(x,y) |\mathbf{Z}_{\max}(x,y) - \mathbf{Z}(x,y)| + \sigma_{\mathbf{Z}}(x,y), \tag{5.6}$$

where $\mathbf{Z}_{\max}$ corresponds to the MIP index map and $\sigma_{\mathbf{Z}}$ is the local spatial standard deviation computed for a 3×3 window centred on (x, y). The $\mathcal{W}$ operator assigns a weight to each pixel to quantify whether it encodes a signal. As weighting scheme, we opt for the following `softplus` function:

$$\mathcal{W}[f_{i\in(x,y)}] = \log\left\{1 + \sqrt{a \exp\left[k(f_i - b)\right]}\right\}/k, \tag{5.7}$$

where the constants are set to $a = 10^{-3}$, $b = 30$ and $k = 0.125$, on the basis of numerical experiments. Finally, we adopt a tolerance threshold ϵ that is sufficiently stringent to allow the

Algorithm 1 Smooth manifold extraction

procedure SME($\mathcal{I}$)
 ▷ Input is an image stack $\mathcal{I}$ with dimensions $W \times H \times D$
 $\mathbf{Z}_{\text{max}}(x, y) = \underset{z}{\text{argmax}}\, \mathcal{I}(x, y, z)$ ▷ MIP index map
 $\mathbf{I}_{\text{MIP}} = \mathcal{I}[x, y, \mathbf{Z}_{\text{max}}(x, y)]$ ▷ MIP 2D image
 $\mathbf{Z}_0 = \mathbf{Z}_{\text{max}}(x, y)$ ▷ Initialise $\mathbf{Z}$ with MIP index map
 ▷ Initial step size (T_0), step factor (ΔT), tolerance (ϵ)
 $i = 1$, $T_0 = D/100$, $\Delta T = 0.99$, $\epsilon = 5 \times 10^{-3}$
 while $\|\mathbf{Z}_i - \mathbf{Z}_{i-1}\| / \|\mathbf{Z}_i\| > \epsilon$ **do**
  for $\forall (x, y) \in W \times H$ **do**
   ▷ All quantities below are a function of (x, y)
   $\Delta \mathbf{Z}_i \in \{-T_i, 0, T_i\}$
   $\tilde{\mathbf{Z}} \leftarrow \mathbf{Z}_{i-1} + \Delta \mathbf{Z}_i$
   $\mathbf{Z}_i \leftarrow \underset{\tilde{\mathbf{Z}}}{\text{argmin}}\, \mathcal{W}(\mathbf{I}_{\text{MIP}}) |\mathbf{Z}_{\text{max}} - \tilde{\mathbf{Z}}| + \sigma_{\tilde{\mathbf{Z}}}$
  end for
  $T_i \leftarrow T_i \times \Delta T$
  $i \leftarrow i + 1$
 end while
 return $\mathcal{I}[x, y, \mathbf{Z}_i(x, y)] \equiv \mathbf{I}_{\text{SME}}$ ▷ SME 2D image
end procedure

cost function to converge. The numerical implementation of our variant of the SME algorithm, as outlined in Algorithm 1, employs a distinct weighting strategy and convergence scheme with respect to the original implementation (Shihavuddin et al., 2017), and results in smooth convergence of the cost function. A comparison of the temporal index maps for SME and MIP, and their respective compressed image representations, is illustrated in Figure 5.4.

5.2.4 Neural classifiers

Here, we describe our neural classifiers that are specialised in handling single-epoch, compressed temporal, or spatio-temporal input. The single-epoch and compressed temporal models are cast as a convolutional neural network (CNN) (LeCun et al., 2015; LeCun, Bengio, et al., 1995), as described in Section 2.2.1. The CNN is composed of convolutional and max-pooling layers, and eventually a fully-connected layer leading to the final output layer. Each convolutional layer has a set of kernels that learn to detect a certain type of feature present in the network input. We made use of 2D kernels whose weights are first randomly initialised and subsequently updated during the network training routine. The input data set for the single-epoch and compressed temporal models comprises a collection of lensed and unlensed supernova images (single-epoch or compressed time-series representation, respectively), while the ground truth labels are binary in nature. As depicted in Figure 5.5, the CNN architecture consists of three blocks of two convolutional layers employing 6, 12 and 24 kernels of sizes 5×5, 3×3 and 1×1, respectively, with unit stride and zero-padding, followed by a max-pooling operation. The first layer learns features on scales of 1.25 arcsecs. With the successive stacking of convolutional layers, low-level local features gleaned in the initial layers are merged into high-level global features by the following layers, rendering the network sensitive to features on larger scales. This allows both local and global information to propagate through the deep learning machinery. We applied the non-linear rectified linear unit (ReLU) (Nair & Hinton, 2010) activation function, $f(x) = \max(0, x)$, to every feature map, except for the output layer. The feature maps are then fed to a max-pooling layer that extracts the maximum value over 2×2 non-overlapping regions of the feature maps for dimensionality reduction. The output from the final convolutional layer and its corresponding pooling layer is flattened into a 1D vector and fed to two fully-connected layers with 12 and 4 neurons, respectively. The final output layer has a sigmoid activation to yield outputs in the range $[0, 1]$, as per the convention for networks designed for binary classification.

For the spatio-temporal network, we employ a Convolutional LSTM (hereafter ConvLSTM) (Shi et al., 2015) as described in Section 2.2.2, since these models are especially well-suited to exploit spatial and temporal correlations in data. The ConvLSTM architecture of the spatio-temporal network is illustrated in Figure 5.5. The input data set contains a collection of time-series images of lensed and unlensed supernovae, with the ground truth labels and final network output being identical to those of the CNN models. The input time-series images are first processed through three ConvLSTM layers, encoding 6, 12 and 24 cells, respectively, combining 2D convolutions in the spatial domain with recurrent LSTMs across the temporal dimension. For dimensionality reduction, max-pooling operations are applied after each ConvLSTM layer, followed by a stack of two LSTM layers after a flattening step. The final section of the ConvLSTM architecture, i.e. fully-connected layers and network output, follows that of its CNN counterpart.

For a classification problem, the network's prediction accuracy and overall efficacy is generally evaluated via a performance metric, with the conventional choice for a classification task being the *accuracy*, i.e. the fraction of correctly assigned labels. Nevertheless, this is not sufficient to assess the reliability of any individual network prediction and it is, hence, imperative to quantify the confidence associated with a given prediction. For a standard neural classifier, the final output layer is an N-dimensional vector, corresponding to a classification problem with N classes, whose components sum to unity, such that they can be interpreted as probabilities. The trained network, for a given test realisation, then assigns the label to the one with largest output probability score. A caveat is that the network probability score should not be interpreted as the confidence in the prediction as it is explicitly dependent on the maximum likelihood estimates of the weights, the training data and the set of hyperparameters adopted during the network training procedure.

In this work, we adopt the variational inference technique known as Monte Carlo (MC) Dropout (Gal & Ghahramani, 2015a), which is outlined in more detail in Section 2.3.1. We use the posterior predictions obtained via the MC Dropout method to derive a confidence score, as a single summary statistic, that encapsulates the epistemic uncertainty for each network classification. To this end, we make use of the information entropy $\mathcal{H}$ in the binary classification

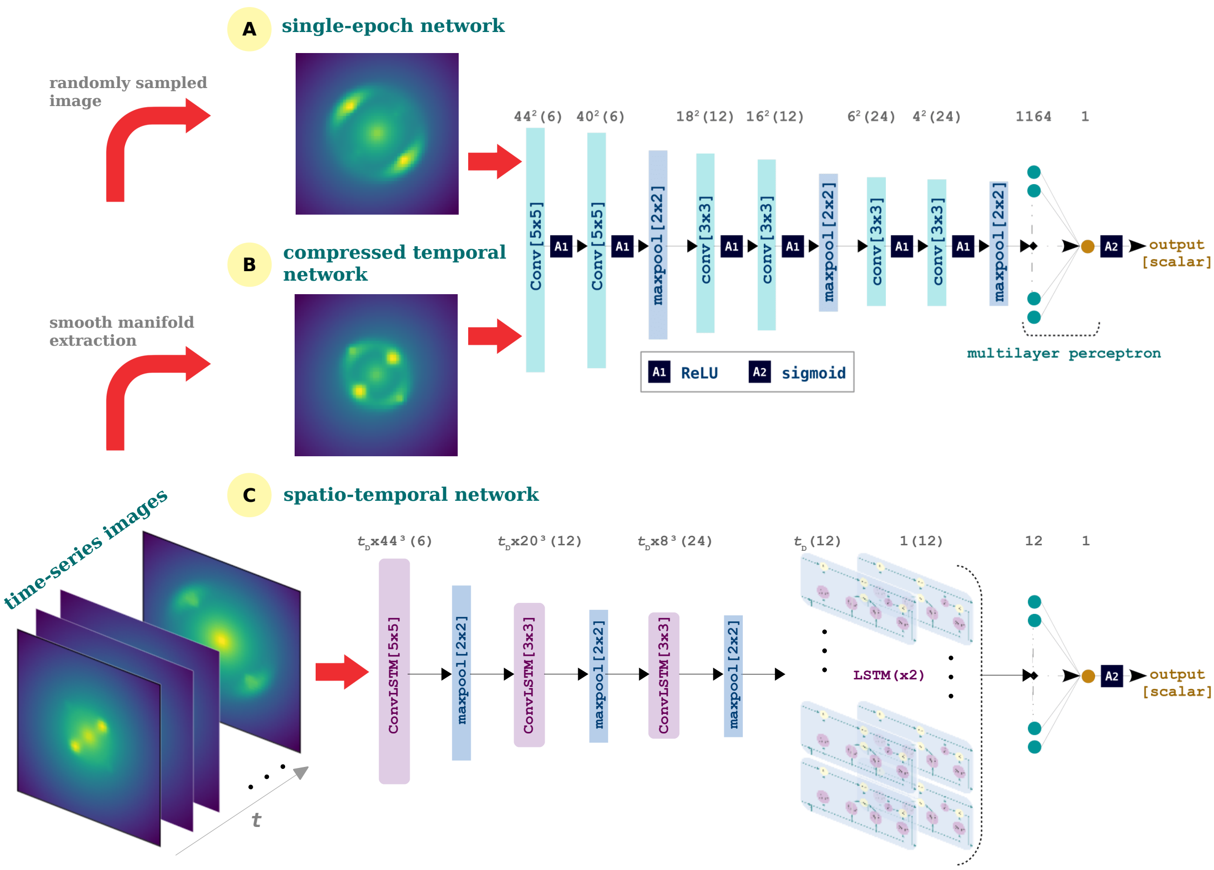


FIGURE 5.5 – Schematic representation of our Bayesian neural classifiers. The three models have distinct inputs, but with a similar scalar output, corresponding to a probability score, that is used for classification. The spatio-temporal model, unlike the single-epoch and compressed temporal models, optimally exploits the information characterised by the temporal evolution of the astrophysical transient. We combine our neural classifiers with variational inference to provide a confidence score, thereby quantifying approximate Bayesian uncertainties associated with each network prediction. The dimensions of the image slices, resulting from the various convolutional, maxpooling or LSTM operations are indicated above the architecture schematics, with the number of feature maps per layer displayed in parentheses and t_{D} denoting the temporal dimension (number of time-series images per source) for the spatio-temporal network. Dropout masks are applied after each convolutional or dense layer, except for the final output layer. The AI engines have a relatively low model complexity with $\mathcal{O}(10^4)$ trainable parameters.

context (Houlsby et al., 2011):

$$\mathcal{H}(p) = -p\log_2 p - (1-p)\log_2(1-p), \tag{5.8}$$

where p is the network output probability. The neural classifier confidence $\mathcal{C}$ can then be computed from the average entropy of N_{MC} posterior samples, as follows (Killestein et al., 2021):

$$\mathcal{C} = 1 - \frac{1}{N_{\rm MC}} \sum_{i=1}^{N_{\rm MC}} \mathcal{H}_i, \tag{5.9}$$

where $\mathcal{H}_i$ denotes the binary entropy of the $i^{\rm th}$ sample, yielding a confidence score in the range $[0, 1]$. While this approach provides uncertainties that are correlated with the network probability scores, the dispersion resulting from multiple samples is still informative, with ten MC samples found to improve classification performance relative to a deterministic network (cf. Figure 6 in Killestein et al. (2021)). In our work, we set $N_{\rm MC} = 50$.

Our choice of variational method is driven primarily by the intended practical utility of our AI tool for sifting through huge volumes of plausible lensed transients. Having a separate confidence score for a specific source, in addition to the network probability score, is informative for both human vetting and automated filtering pipelines. The numerical implementation is also straightforward, with the extra requirements being the inclusion of dropout masks after the convolutional and dense layers, except for the final output layer, with ℓ_2 weight regularisation. The network training routine is also unchanged with respect to a conventional neural network, with the additional computational workload being the multiple model evaluations at inference time for a given input image. The MC Dropout method is robust as long as we restrict our neural network to low model complexity and employ a large training data set, such that it is justified to assume minimal dropout rates and ℓ_2 regularisations. A more rigorous way of quantifying epistemic uncertainties is by training an ensemble of networks with randomly drawn architectures, variational distributions and hyperparameters, but this is beyond the scope of the practical tool developed in this work.

5.2.5 Network implementation and training

We implement our Bayesian neural classifiers using the Keras framework (Chollet et al., 2015) via a TensorFlow backend (Abadi et al., 2016). To train the neural networks, we use the adam (Kingma & Ba, 2014) optimiser for robust and reliable convergence, and the binary cross-entropy loss function designed for binary classification problems. We set the learning rate to $\eta = 10^{-4}$, along with the default values, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, for the first and second moment exponential decay rates, respectively. We use a fixed batch size of 100 samples and train our neural classifiers for around 50 epochs, with the training routine for the CNN and ConvLSTM running to completion in around ten and forty minutes, respectively, on an NVIDIA V100 Tensor Core GPU.

To limit network overfitting, we adopt the early stopping regularisation technique (Goodfellow et al., 2016), and implement an early stopping criterion of five epochs. To this end, 20% of the training data set is randomly selected to comprise a validation set. Training is terminated when the validation loss ceases to improve for five consecutive epochs, with the weights of the previously saved best fit model restored. We use a marginal dropout probability of 10^{-2} between the convolutional and fully-connected layers, along with ℓ_2 weight regularisation penalty of 10^{-4}, which provide further regularisation. The inclusion of the dropout masks and ℓ_2 regularisation, however, is primarily for the purpose of variational inference to quantify the confidence score associated with a given network classification. Our CNN architecture, consisting of 12075 trainable parameters, has a relatively low model complexity and requires negligible regularisation given the large volume of training data. Our ConvLSTM architecture has roughly five times more parameters with 65961 trainable weights, but is still generally of low complexity (in machine learning terms). We leave any (automated) hyperparameter tuning to a future work.

5.3 RESULTS

5.3.1 Evaluating neural classifier performance

In order to evaluate the performance of our trained Bayesian neural classifiers on the test set, we employ several standard classification metrics. To quantify the accuracy of our model, we first compute the true positive rate (TPR) and false positive rate (FPR). The former, also known as *completeness*, is defined as the fraction of detected lensed sources, whilst the latter, also referred to as the *contamination* rate, is defined as the fraction of normal supernovae incorrectly classified as lensed sources:

$$\mathrm{TPR} = \frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}}, \quad \mathrm{FPR} = \frac{N_{\mathrm{FP}}}{N_{\mathrm{FP}} + N_{\mathrm{TN}}}, \tag{5.10}$$

where N_{TP}, N_{FN}, N_{FP} and N_{TN} denote the number of true positives, false negatives, false positives and true negatives, respectively. These summary statistics are a function of the detection threshold applied to the neural classifier's probability score. The overall performance of the trained classifier is, hence, conventionally assessed by the receiver operating characteristic (ROC) curve. This diagnostic curve depicts the TPR as a function of the FPR and is generated by gradually increasing the detection threshold in the range $[0, 1]$, as illustrated in Figure 5.6 for our three models. Random predictions by the classifier will produce a diagonal line, with the area under the curve, denoted by the AUC score, equal to 0.5, whilst a perfect classifier will have an AUC score of unity. All three models achieve decent AUC scores, with that of the spatio-temporal one approaching unity, showcasing the gradual improvement of the use of the full time-series images over single-epoch images, as a result of the additional temporal information.

In this work, we adopt the default value of 0.5 for the detection threshold for both of our models. Given this threshold, we can visualise the overall classification accuracy using the confusion matrix, as shown in Figure 5.7. The confusion matrix describes the percentage of samples from each class that are accurately classified and simultaneously expresses that of erroneous classifications. As a robust measure of the quality of binary classifications, the Matthews correlation coefficient (MCC) is a commonly used metric to summarise the confusion matrix:

$$\mathrm{MCC} = \frac{N_{\mathrm{TP}}N_{\mathrm{TN}} - N_{\mathrm{FP}}N_{\mathrm{FN}}}{\sqrt{(N_{\mathrm{TP}} + N_{\mathrm{FP}})(N_{\mathrm{TP}} + N_{\mathrm{FN}})(N_{\mathrm{TN}} + N_{\mathrm{FP}})(N_{\mathrm{TN}} + N_{\mathrm{FN}})}}, \tag{5.11}$$

with the limiting values of $\mathrm{MCC} = \{-1, 1\}$ corresponding to predictions in total disagreement and perfect agreement, respectively, with observations, while $\mathrm{MCC} = 0$ implies random classifier

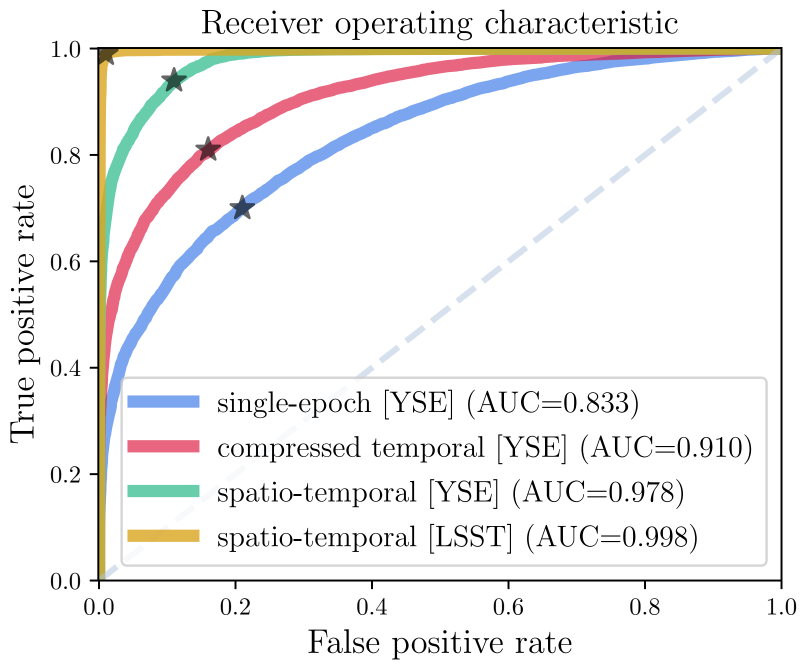


FIGURE 5.6 – Receiver operating characteristic curves for the three Bayesian neural classifiers, showing the true positive rate as a function of the false positive rate, both computed for varying detection thresholds in the range [0,1]. The chosen detection threshold of 0.5 used to compute the confusion matrix in Figure 5.7 is indicated using black stars. All classifiers achieve decent AUC scores, with the AUC score of the spatio-temporal model highlighting the progressive improvement in classification efficacy with the addition of temporal information. Its performance on the LSST mock observations is also extremely promising.

predictions. The MCC scores for our models are given in Figure 5.7, further demonstrating the improved classification efficacy with the inclusion of temporal information.

An essential aspect to consider when using a neural classifier is whether its output, i.e. the network probability score, accurately reflects the probability of a source being lensed or unlensed. In order to verify how closely the network's output correlates with probability, we perform a classifier calibration test in Figure 5.8. All three models display nearly perfect calibration, implying that the network probability scores may be interpreted as the probability that a certain source is lensed. Moreover, our Bayesian neural classifiers quantify the reliability of the network classifications via a confidence score, computed from the average information entropy of an ensemble of Monte Carlo posterior samples from the trained networks (cf. eq. (5.9)). The distribution of such scores associated with wrongly classified transients depicted in Figure 5.9. We find that there is a clear trend of erroneously classified sources having very low confidence scores, thereby justifying the utility of this additional metric when evaluating potential lensed supernovae.

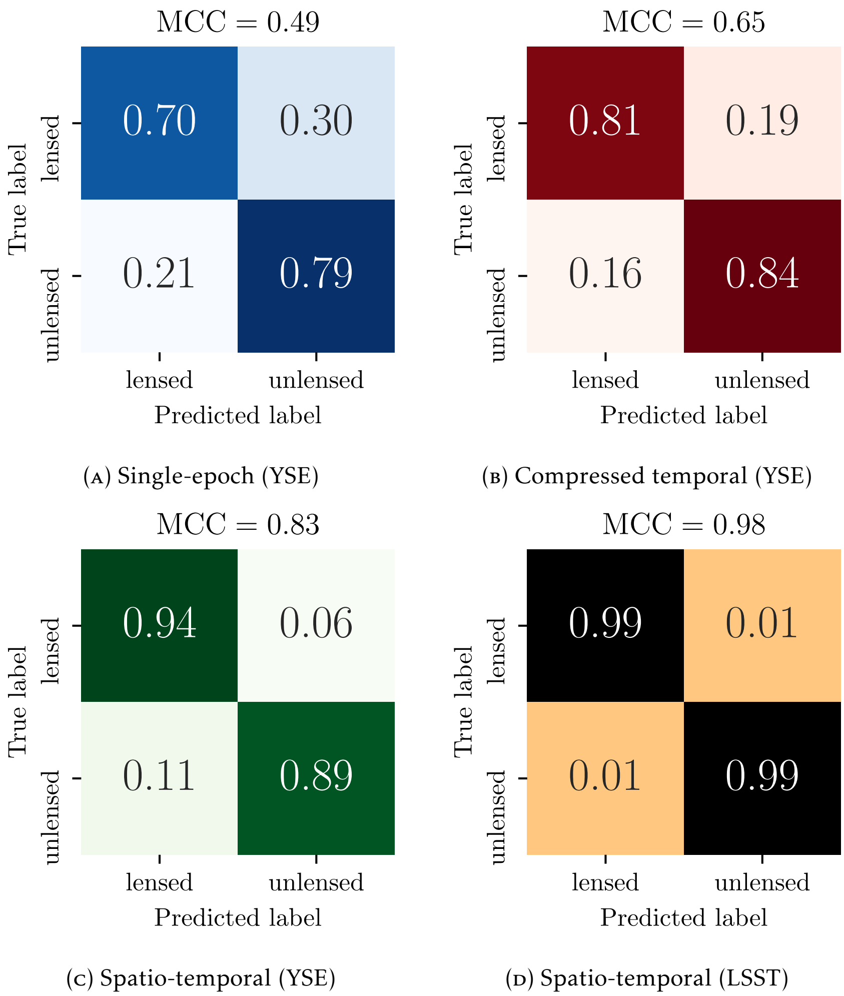


FIGURE 5.7 – Confusion matrix showing the classification efficacy of the three models employing single-epoch images, the compressed temporal representation and all the time-series images, respectively, for the YSE set-up. The spatio-temporal model yields an improvement of nearly 20% in overall classification accuracy relative to the single-epoch model, with 99% accuracy for LSST images.

5.3.2 Application to mock LSST observations

Next-generation transient surveys are poised to drastically increase the number of known lensed supernovae. For instance, the Legacy Survey of Space and Time (LSST) from the Vera C. Rubin Observatory (Ivezic et al., 2008) should lead to the imminent discovery of several hundreds of lensed supernovae (Goldstein & Nugent, 2017; Wojtak et al., 2019), while the Nancy Grace Roman Space Telescope (Spergel et al., 2015) is expected to find a few dozens (Oguri & Marshall, 2010). For such surveys with enhanced resolution and seeing, we expect an improved classification performance of the spatio-temporal model. To showcase the relevance

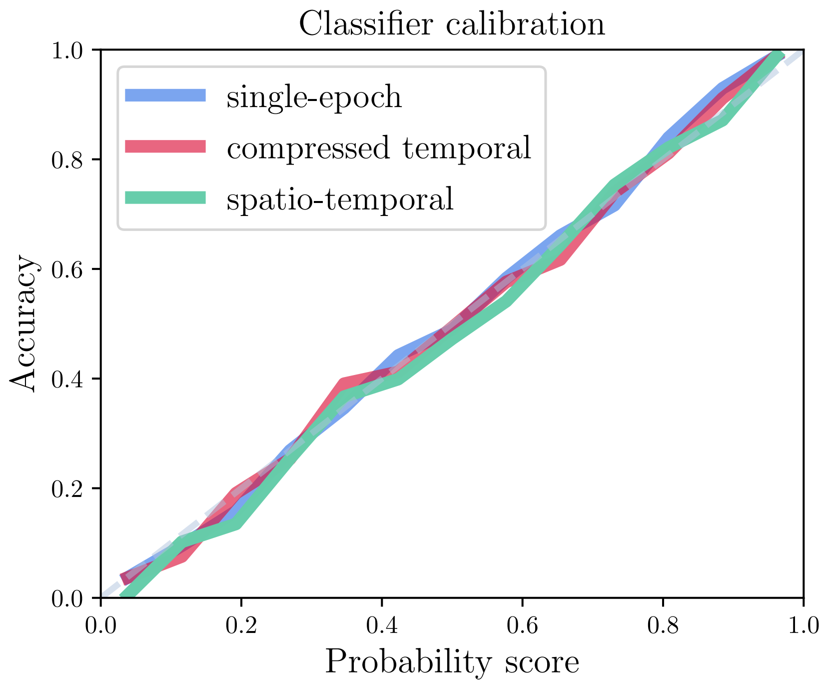


FIGURE 5.8 – Classifier calibration indicating how well the network probability score corresponds to probability. The diagonal dashed line implies perfect calibration, i.e. a perfect match between the probability score and accuracy. The calibration curves for the three Bayesian neural classifiers show a remarkable degree of calibration, such that it is justified to use the network probability score as a proxy for the probability that a source is lensed.

of our spatio-temporal engine for such upcoming surveys, we verify the classification efficacy for mock observations generated within a preliminary LSST-like set-up (cf. Table 5.1). The gold line in Figure 5.6 depicts the resulting ROC curve with an AUC score of almost unity (AUC = 0.998), with the corresponding confusion matrix provided in Figure 5.7 (MCC = 0.978) indicating an overall classification accuracy of around 99 per cent. Over the lifetime of LSST, $\mathcal{O}(10^2)$ lensed supernovae are expected out of the $\mathcal{O}(10^5)$ new supernova discoveries per year. With the spatio-temporal network, the odds of finding a lensed supernova rise from 1 in 1000 to 1 in 10, representing an improvement by two orders of magnitude.

5.4 DISCUSSION AND CONCLUSION

We presented a novel AI-assisted spatio-temporal engine, based on recurrent convolutional layers, to identify gravitationally lensed supernovae from the presently ongoing YSE survey. Our approach draws from recent advances in variational inference to quantify approximate Bayesian uncertainties, thereby assigning a confidence score to each model prediction that accurately reflects the uncertainty inherent to the network classification. Incorporating the temporal information encoded in the evolution of a given source led to a significant gain of almost 20

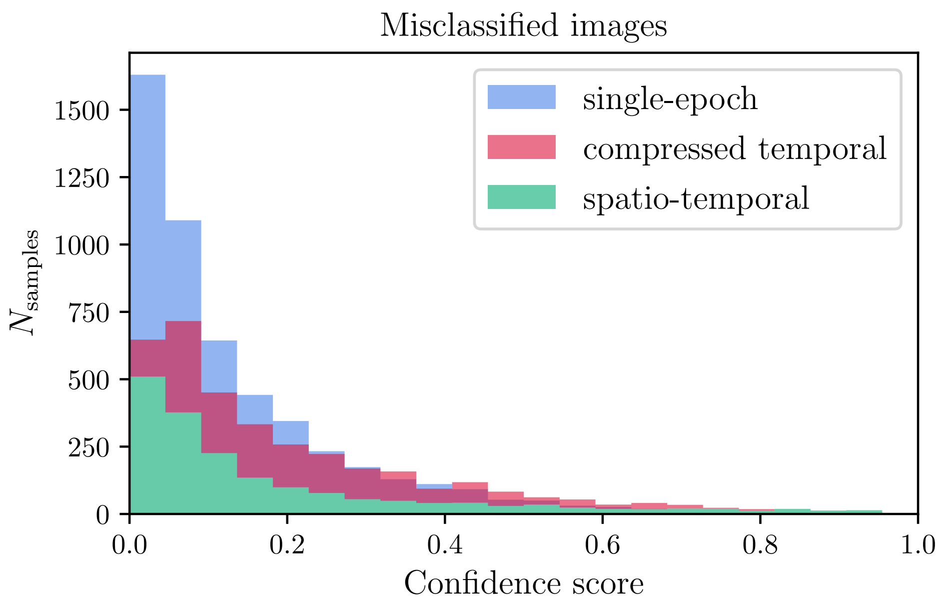


FIGURE 5.9 – Distribution of confidence scores for misclassified (false positives and false negatives) images. On average, for all three models, only around 5% of wrongly classified sources have a confidence score larger than 0.5, thereby demonstrating the reliability of this metric to reflect the neural network uncertainty associated with a given classification.

per cent in classification accuracy relative to single-epoch observations for the test data set generated within the YSE set-up. Such neural classifiers are complementary to the standard selection based on typical lens galaxy redshifts and lens-supernova angular separations, with the combination of these two distinct approaches crucial for rapidly identifying promising lensed supernova candidates, such that follow-up spectroscopic observations can be initiated in a timely manner.

Our spatio-temporal model is tailored to detect time-variable lensing features in the time series of difference images. To illustrate that the temporal correlations in the time-series images of a given source are conducive to the classification accuracy, we trained a CNN using the compressed time-series representation. To this end, we implemented a variant of the smooth manifold extraction technique, originally proposed for processing a stack of images produced via 3D fluorescence microscopy, to compress the observed time-series images of an astrophysical transient into a single informative image. We find that the spatio-temporal model results yet in an improvement of around 10 per cent in accuracy relative to its compressed temporal counterpart. This gain in accuracy matches our intuitive expectations since the variation in observed brightness of lensed supernovae has a particular trend.

Having an estimate of the confidence associated with a given neural classifier prediction undeniably brings some additional insights when sifting through an avalanche of plausible transients as recorded by an instrument. Human vetting, as the typical final step after the

ML-based pre-filtering, should benefit from the extra information about the neural network uncertainties when assessing and prioritising potential lensed supernovae. Moreover, the confidence metric can also be incorporated alongside the network probability score on the data acquisition platform for prompt detection alerts concerning extremely promising lensed candidates. This opens the possibility of automating decision-making pertaining to follow-up observations and reporting.

The AI machinery presented here may be naturally adapted in the context of upcoming transient surveys that would deliver unprecedented volumes of lensed candidates. Indeed, we performed a preliminary application to LSST-like difference images and showed that the classification accuracy of the spatio-temporal model rises to around 99 per cent, thereby demonstrating its efficacy for next-generation surveys. The classification accuracy can be even further enhanced when images in the remaining filters are included, thereby improving an effective temporal resolution of the image series. As such, a straightforward extension is to include additional channels in our network so as to work at the level of multi-band images.

An interesting application of our spatio-temporal engine is to determine gravitational time-delays from multiple-epoch images of lensed supernovae and subsequently infer the Hubble constant (Arendse et al., 2021), which constitutes our ongoing study. This yields a complementary framework to standard cosmographic analyses in the quest for independent measurements of the cosmic acceleration.

DATA AVAILABILITY

The source code repository and catalogue of simulated images will be made available at https://github.com/doogesh/spatio_temporal_glSNe_finder upon publication.

AUTHOR CONTRIBUTIONS

DKR led the project, built the image generation pipeline, implemented a variant of the SME algorithm for combining time-series snapshots, designed and optimised the neural network architecture (three ML models), performed the analysis, produced the figures, wrote the paper. NA implemented the code for generating time-series images, assisted the network optimisation and contributed to writing up the manuscript. RW proposed the main idea, validated the image simulation pipeline, and contributed to writing up the manuscript.

ACKNOWLEDGEMENTS

We thank Jens Hjorth, David Jones and Ken Smith for their assistance with our YSE-related queries. We thank Tom Charnock for his views on variational inference and David Prelogović for interesting exchanges about recurrent networks. DKR is a DARK fellow supported by a Semper Ardens grant from the Carlsberg Foundation (reference CF15-0384). This work was supported by a VILLUM FONDEN Investigator grant (project number 16599). This work has made use of the Infinity Cluster hosted by Institut d'Astrophysique de Paris and an HPC facility funded by a grant from VILLUM FONDEN (project number 16599).

CHAPTER 6

Inferring the Hubble constant from lensed supernovae

This chapter presents the following work in progress:
"Inferring the Hubble constant from lensed supernovae in LSST with spatio-temporal neural networks"

Authors: Nikki Arendse, Doogesh Kodi Ramanah & Radosław J. Wojtak.

ABSTRACT

Gravitationally lensed supernovae are a promising new probe to obtain independent measurements of the Hubble constant (H_0). Here, we present a machine learning pipeline that constrains the Hubble constant from simulated time-series images for the Legacy Survey of Space and Time (LSST), ultimately aimed at cosmological inference on real LSST data. We use a convolutional neural network that is sensitive to both the spatial and temporal features of the input data, along with a simulation-based inference approach to quantify the uncertainties on the network predictions. In this proof-of-concept demonstration, we forecast to find a sample of 400 lensed type Ia supernova systems during the 10 year survey duration of LSST. From this sample, we obtain an unbiased 1.2% measurement of H_0 using only LSST i-band data without any follow-up observations. We find that the majority of the constraining power emanates from doubly imaged supernovae, whereby the largest residual scatter is due to the source position and time delays between the lensed images. This work underlines the usefulness of pure LSST data for time-delay cosmography in a joint population analysis. A continuation of this study will take into account the effects of microlensing on the inferred precision of H_0.

6.1 INTRODUCTION

RECENT measurements of the Hubble constant (H_0) from the Cosmic Microwave Background (CMB; Planck Collaboration, Aghanim, et al., 2018) radiation are in tension with observations from low-redshift probes, such as type Ia supernovae calibrated by Cepheids (Riess et al., 2021) or gravitationally lensed quasars (Wong et al., 2020). In addition to characterising the cosmic expansion rate at any epoch, the Hubble constant is an essential quantity for calculating the age of the Universe and calibrating the cosmic distance scale. Furthermore, the discrepancy between low-redshift and CMB-based methods could point to new physics beyond the standard ΛCDM model.

Nevertheless, the tension is lowered considerably by several recent low-redshift measurements, such as distance calibrations from the Tip of the Red Giant Branch (TRGB), as measured by the Carnegie-Chicago Hubble Project (Freedman, 2021) and the new analysis of seven gravitationally lensed quasars by the TDCOSMO collaboration (Birrer et al., 2020) with less restrictive mass model priors than in Wong et al. (2020). To determine whether new physics or residual systematics are behind the Hubble tension, there is a pressing need for new, independent measurements of H_0. One promising avenue is the use of gravitationally lensed supernovae for cosmology, as described in more detail in Section 1.4.6.5. Although the current estimates of detection rates of similar lensed supernovae (Goldstein & Nugent, 2017; Oguri & Marshall, 2010; Wojtak et al., 2019) in ongoing transient surveys, such as the Zwicky Transient Factory (ZTF; Bellm et al., 2019) and the Young Supernova Experiment (YSE; D. O. Jones et al., 2021), are only a few objects per year, the predicted rates are more than two orders of magnitude higher for next-generation surveys, such as the the Legacy Survey of Space and Time (LSST) from the Vera C. Rubin Observatory (Ivezic et al., 2008).

As the aforementioned upcoming time-domain surveys will receive several terabytes worth of observations per night, processing the data becomes increasingly challenging. Machine learning (ML) methods have emerged as a way to make sense of the unprecedented volumes of data and to automate the detection and analysis of lensed sources. Specifically, convolutional neural networks (CNNs) have been successfully employed to find strongly lensed systems in images from transient surveys (Avestruz et al., 2019; Cañameras et al., 2020; Cheng et al., 2020; Davies et al., 2019; Gentile et al., 2021; X. Huang et al., 2021; X. Huang et al., 2020; Kodi Ramanah, Arendse, et al., 2021; Lanusse et al., 2018; Schaefer et al., 2018) and to infer lens galaxy properties without costly Markov Chain Monte Carlo (MCMC) analyses (Hezaveh et al., 2017; Park et al., 2021; Wagner-Carena et al., 2021).

Another application of lensed transient images is to infer time delays between different images, which are primarily sensitive to the Hubble constant. A higher value of H_0 corresponds to shorter time delays, as illustrated for two simulated lensed supernova systems in Figure 6.1. Recently, Huber, Suyu, Ghoshdastidar, et al. (2021) developed an ML approach, based on fully connected neural networks and random forests, to obtain point estimates of the time delays of simulated lensed supernovae from their observed light curves. However, so far there has not been a complete, automated pipeline to take time-series images of lensed supernovae as input and combine the inference of the time delays and lens model properties into an estimate

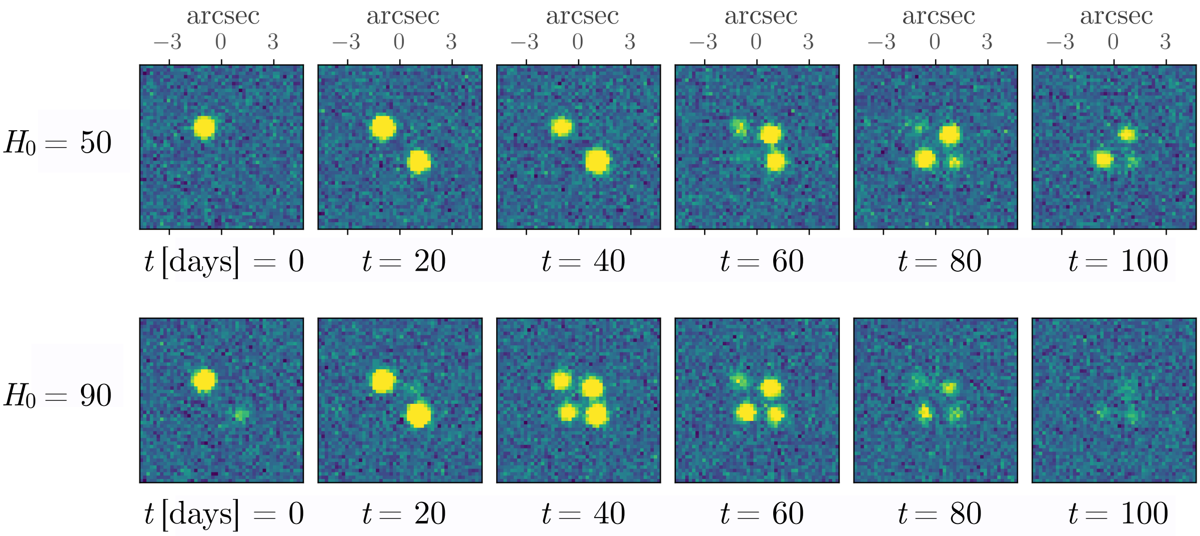


FIGURE 6.1 – Temporal evolution of difference images of two identical lensed supernova systems, as observed by LSST, simulated using a different value of the Hubble constant. Since a higher value of H_0 corresponds to shorter time delays, observations of time-series images can be employed to determine the cosmic expansion rate. The lens system depicted in the images corresponds to $z_{\text{lens}} = 0.3$, $z_{\text{src}} = 0.6$ and $\theta_{\text{E}} = 1.2$, which is at the higher end of the Einstein radius distribution expected for typical LSST systems.

of the Hubble constant. We present such a framework in this chapter, in the context of LSST observations. First, we train a spatio-temporal CNN on simulated LSST time-series images and employ a simulation-based inference approach to quantify the uncertainties on the neural network predictions. Then, by using realistic expectations for the lensed type Ia supernovae discovery rate, we obtain predictions for the accuracy and precision of future H_0 estimates, using *only* LSST i-band data and no follow-up observations. Although the precision from a single lensed supernova system in LSST data will not be competitive, we demonstrate that the combined posterior of 400 lensed type Ia supernovae yields a 1.2% unbiased measurement of the Hubble constant. This prediction of the lensed supernova rate corresponds to LSST observations over a time period of 10 years.

Throughout this work, we assume a standard flat ΛCDM model with $\Omega_{\text{m}} = 0.3$. The remainder of this chapter is organised as follows. Section 6.2 provides a description of our image simulation procedure, assumptions regarding the lens mass profile and supernova light curves, as well as details pertaining to the spatio-temporal CNN architecture and simulation-based inference approach. In Section 6.3, we present our results for a joint H_0 inference and verify the validity of the uncertainty estimates provided by the neural network. Finally, we provide a discussion of the main findings of our study and possible avenues for future work in Section 6.4.

LSST image properties	
PSF (FWHM)	$\mathcal{N}^{\mathcal{S}}(4.05, 0.55, 0.30)$
Pixel size ($''$)	0.2
Pixel background noise (σ_{bkg})	6.1
Exposure time (s)	30
Postage stamp size ($''$)	9.6
Postage stamp area (arcsec^2)	92
Number of pixels	48×48
Median cadence (days; i-band)	11.0
Zero-point magnitude (i-band)	27.79
Limiting magnitude (i-band)	23.9

TABLE 6.1 – **Simulated image characteristics.** The images constituting the training, validation, test and evaluation sets are generated in accordance with the LSST survey specifications, thereby closely emulating real LSST observations. $\mathcal{N}^{\mathcal{S}}(a, \mu, \sigma)$ denotes a skewed normal distribution with skewness parameter a.

6.2 METHODS

6.2.1 Lens galaxy mass profile assumptions

WE EMPLOY the multi-purpose lens modelling package LENSTRONOMY[1] (Birrer & Amara, 2018) to generate images of lensed supernovae, while assuming a flat ΛCDM model with $\Omega_{\text{m}} = 0.3$. The mass profile that we adopt for the lens galaxies is a power-law elliptical mass distribution (PEMD; Barkana, 1998; Kormann et al., 1994), as given in eq. (1.81), which is an extension of the singular isothermal ellipsoid mass profile whereby the 3D power law mass slope γ_{lens} is allowed to vary. To circumvent issues arising from cyclic boundary conditions due to the 2π-periodic property of the angles when we train our network, we express the target lens mass ellipticity as follows:

$$e_1 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \cos(2\phi_{\text{lens}}) \tag{6.1}$$

$$e_2 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \sin(2\phi_{\text{lens}}), \tag{6.2}$$

where q_{lens} is the projected axis ratio of the lens and ϕ_{lens} is the lens orientation angle. We model the external shear from the line-of-sight structures with a shear modulus γ_{ext} and a shear angle ϕ_{ext}. The adopted parameter distributions are given in Table 6.2. In our current implementation, we do not account for microlensing effects from substructures in the lens galaxy. At the moment, we are developing a framework to include an approximate, yet realistic,

[1] https://lenstronomy.readthedocs.io/en/latest/

Parameter	Distribution
Hubble constant	$H_0 \sim \mathcal{U}(20, 100)$
Lens redshift	$z_{\rm lens} \sim \mathcal{N}^{\mathcal{S}}(3.88, 0.13, 0.36)$ *
Lensed source redshift	$z_{\rm src} \sim \mathcal{N}^{\mathcal{S}}(3.22, 0.53, 0.55)$ *
Source position	$x_{\rm src}, y_{\rm src} \sim \mathcal{U}(-\theta_{\rm E}, \theta_{\rm E})$
Lens galaxy	
Elliptical power-law mass	
Lens centre ($''$)	$x_{\rm lens}, y_{\rm lens} \equiv (0, 0)$
Einstein radius ($''$)	$\theta_{\rm E} \sim \mathcal{N}^{\mathcal{S}}(5.45, 0.14, 0.63)$ *
Power-law slope	$\gamma_{\rm lens} \sim \mathcal{N}(2.0, 0.1)$
Axis ratio	$q_{\rm lens} \sim \mathcal{N}(0.7, 0.15)$
Orientation angle (rad)	$\phi_{\rm lens} \sim \mathcal{U}(-\pi/2, \pi/2)$
Environment	
External shear modulus	$\gamma_{\rm ext} \sim \mathcal{U}(0, 0.05)$
Orientation angle (rad)	$\phi_{\rm ext} \sim \mathcal{U}(-\pi/2, \pi/2)$
Light curve	
Stretch	$x_1 \sim \mathcal{N}^{\mathcal{S}}(-8.24, 1.23, 1.67)$
Colour	$c \sim \mathcal{N}^{\mathcal{S}}(2.48, -0.089, 0.12)$
Absolute magnitude	$M_{\rm abs} \sim \mathcal{N}(M_0, 0.12)$
	$M_0 = 5\log_{10}(H_0/74.03) - 19.24$
Milky Way extinction	$E(B-V) \sim \mathcal{U}(0, 0.2)$

Table 6.2 – **Parameter distributions for lensed systems.** The distribution of input parameters employed in the image simulation pipeline to generate the training, validation, and test data sets. $\mathcal{N}(\mu, \sigma)$ corresponds to a normal distribution with mean μ and standard deviation σ, $\mathcal{N}^{\mathcal{S}}(a, \mu, \sigma)$ denotes a skewed normal distribution with skewness parameter a, while $\mathcal{U}(x, y)$ implies a uniform distribution with bounds x and y.
* The skewed normal distributions for $z_{\rm lens}$, $z_{\rm src}$ and $\theta_{\rm E}$ are only approximations. The parameters are drawn from a joint distribution that is depicted in Figure 6.2.

description of microlensing into our simulation, since microlensing is expected to have a significant contribution to the time delay estimates for lensed supernovae.

To ensure that our final results are not dominated by correlations between model parameters, we assume that most of our parameters are independent. The only exception to this are the Einstein radius $\theta_{\rm E}$, lens redshift $z_{\rm lens}$ and source redshift $z_{\rm src}$, which we sample from the joint probability distribution obtained from a Monte Carlo simulation of gravitationally lensed type Ia supernova generated in Wojtak et al. (2019). The simulation assumes a population of

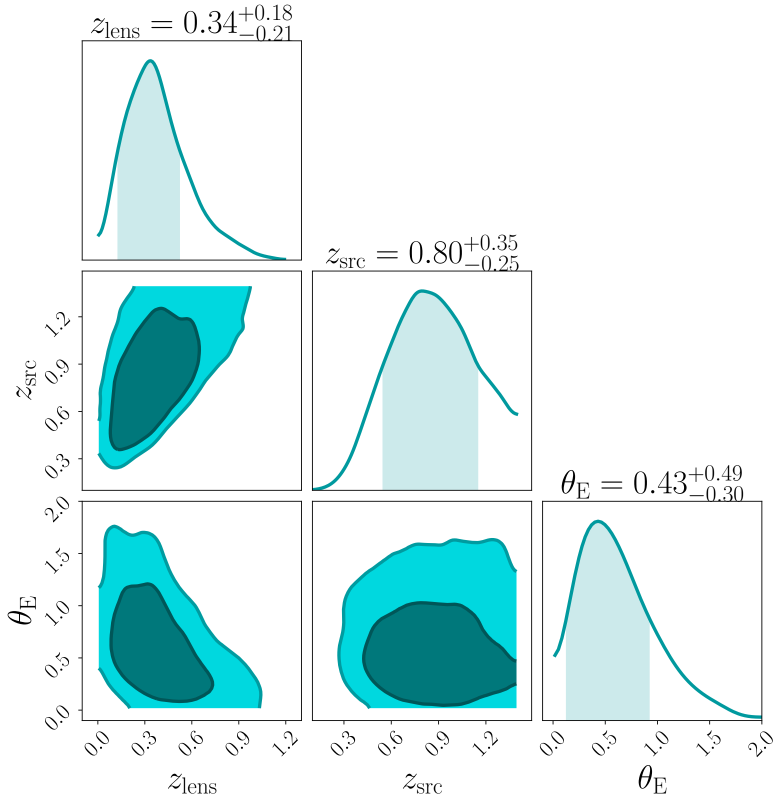


FIGURE 6.2 – The joint distribution of the lens redshift (z_{lens}), source redshift (z_{src}) and Einstein radius (θ_{E}) used to simulate the lensed supernova systems. The z_{lens}, z_{src} and θ_{E} combinations correspond to galaxy-source configurations where strong lensing occurs.

lens galaxies with the velocity dispersion function derived from the Sloan Digital Sky Survey observations (Choi et al., 2007) and a model of the volumetric rate of type Ia supernovae fitted to recent measurements of the type Ia supernova rate as a function of redshift (S. A. Rodney et al., 2014). The mass distribution is given by a singular isothermal ellipsoid model (cf. eq. (1.80) and Kormann et al., 1994) with the same distributions of ellipticity and the external convergence as given in Table 6.2. The lensing properties are computed using the *glafic* code[2] (Oguri, 2010). The supernova sample contains all strongly lensed observable cases for which at least one image is detectable by LSST in the i-band. Although the simulation does not exactly match the lens model used to generate images, drawing z_{lens}, z_{src} and θ_{E} from the precomputed simulation is sufficient to narrow down all available galaxy-source configurations to those where strong lensing occurs. We impose an additional upper limit on the source redshift of $z_{\mathrm{src}} < 1.4$ to ensure that the supernovae are not redshifted out of the i-filter. The resulting combinations of z_{lens}, z_{src} and θ_{E} values are depicted in Figure 6.2. For each configuration from the simulation, we obtain the final lensed supernova parameters by drawing random positions of the source and by sampling the light curve and remaining lens parameters from the distributions given in

[2] https://www.slac.stanford.edu/~oguri/glafic/

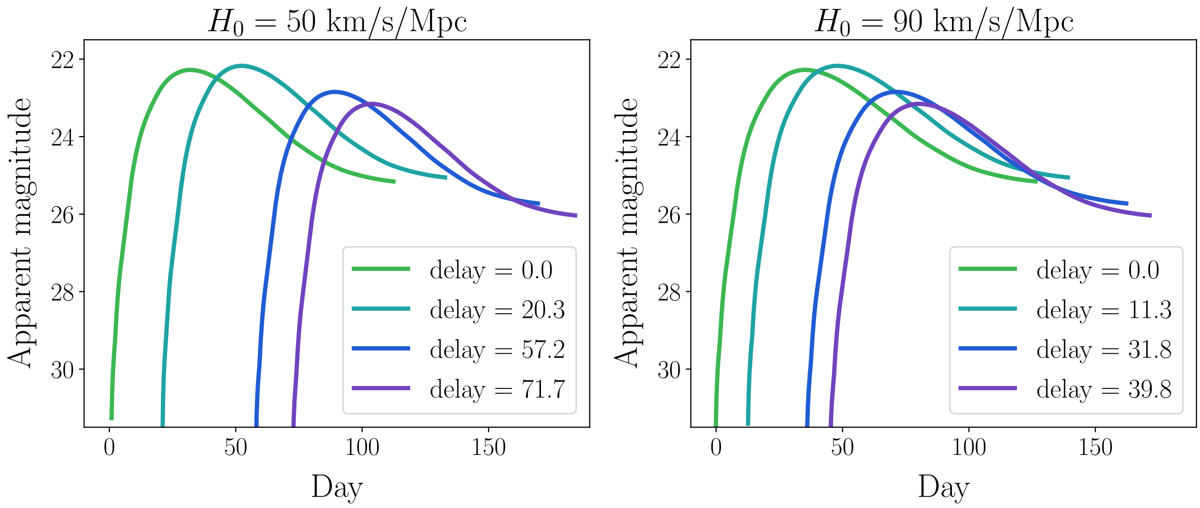


FIGURE 6.3 – Simulated observer-frame i-band light curves that have been used to generate the time-series images for the quad system in Figure 6.1, thereby demonstrating that a higher Hubble constant produces shorter time delays for identical lens systems. The lens parameters for this system are $z_{\text{lens}} = 0.3$, $z_{\text{src}} = 0.6$ and $\theta_{\text{E}} = 1.2$.

Table 6.2 until we find a system that is detectable by LSST.

6.2.2 Supernova light curves

We model the supernovae as point sources, using synthetic light curves in the observer frame for their variability. In this work, we only consider type Ia supernovae, since their characteristic light curves are easy to model, their standard candle nature offers an advantage for breaking the mass-sheet degeneracy, and they constitute a significant fraction ($\sim 26\%$) of the predicted LSST lensed supernova population (Wojtak et al., 2019).

The light curves are simulated using SNCosmo[3](Barbary et al., 2016) and its in-built parametric light curve model SALT2 (Guy et al., 2007), which takes as input an amplitude parameter x_0, stretch parameter x_1, and a colour parameter c. We sample the x_1 and c parameters from asymmetric Gaussian distributions that have been derived by D. Scolnic and Kessler (2016) for the Supernova Legacy Survey (Guy et al., 2010), the Sloan Digital Sky Survey (Sako et al., 2018), Pan-STARRS1 (Rest et al., 2014), and several low-redshift surveys. We employ the Tripp formula(Tripp, 1998) to calculate the absolute B-band peak magnitude M_{B} that a type Ia supernova, based on its stretch and colour parameters, is expected to have:

$$M_{\text{B}} = -\alpha x_1 + \beta c + \mathcal{N}(M_0, 0.12) \tag{6.3}$$

$$M_0 = 5\log_{10}\left(\frac{H_0}{H_{0,\text{fid}}}\right) + M_{\text{fid}}, \tag{6.4}$$

[3] https://sncosmo.readthedocs.io/en/stable/

where M_0 is the expected absolute magnitude of a supernova with $x_1 = c = 0$ in a Universe with Hubble constant H_0. The coefficients $\alpha = 0.14$ and $\beta = 3.1$ (D. Scolnic & Kessler, 2016) specify the correlation of absolute magnitude with the stretch and colour parameters, respectively. As reference fiducial values, we use the calibration from the *Supernovae and H_0 for the Equation of State of dark energy* project (SHoES) (Riess et al., 2021); $H_{0,\text{fid}} = 74.03$ km s^{-1} Mpc^{-1} and $M_{\text{fid}} = -19.24$. For clarity, let us emphasise here that by applying the cosmology correction in eq. (6.4), our resulting M_B values are independent of the SHoES calibration. Subsequently, the resulting absolute magnitude values for each supernova are used as input for SNCosmo to generate the corresponding light curves.

For our proof of concept, we assume that the footprint covered by LSST has a low dust extinction of $E(B-V) < 0.2$, corresponding to a region away from the galactic plane. However, the precise location of the LSST footprint has not yet been finalised and will likely include several regions with $E(B-V) > 0.2$ (R. L. Jones et al., 2020), which might make observations of lensed supernovae more challenging because the dust can obscure part of the supernova signal. In order to model the dust in the supernova light curves, we adopt a Milky Way dust extinction model (Fitzpatrick, 1999) with optical total-to-selective extinction ratio $R_\text{V} = 3.1$. The adopted distribution for Milky Way dust extinction and other input parameters to the light curve generation routine are provided in Table 6.2. Figure 6.3 displays the observer-frame light curves corresponding to the quad system depicted in Figure 6.1, which is simulated using two different values of the Hubble constant.

After the light curves have been generated, they are used to acquire the apparent magnitude (including K-corrections) of the lensed point sources. The brightness of each image follows the variability of the light curve, with a correction for the magnification and time delays computed by LENSTRONOMY. Finally, we transform the apparent magnitude to data counts per second using the magnitude zero-point of the instrument, which, when multiplied with the exposure time, yields the amplitude in the desired units for LENSTRONOMY. The approach described here does not constitute the first framework for generating time-series images; the software packages DEEPLENSTRONOMY (Morgan et al., 2021) and Mock Lenses in Time (MOLET Vernardos, 2021) were published during our study and offer an alternative for simulating time-varying sources.

6.2.3 LSST time-series images

LSST is a wide-field astronomical survey to be conducted at the Vera Rubin Observatory and is scheduled to start full operations in 2023. The survey will take multi-colour *ugrizy* images and cover 18,000 square degrees of the sky in a ten-year period. Due to its depth and sky coverage, LSST is currently the most promising transient survey to observe gravitationally lensed supernovae, with predicted rates of several hundreds a year (Goldstein & Nugent, 2017; Goldstein et al., 2019; Oguri & Marshall, 2010; Wojtak et al., 2019). Discoveries of new transients are realised with the *difference imaging* technique, in which a historic reference image of the sky is subtracted from nightly observations to detect any residual flux. In order to emulate the LSST output, we generate our data as difference images, without contributions from static sources, such as the lens and host galaxy of the supernova. Therefore, only the light properties of the supernova are of importance for our simulation pipeline. Our method assumes a perfect

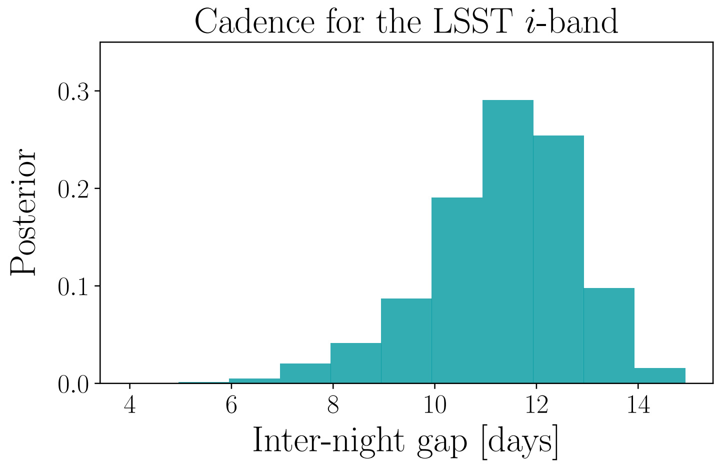


FIGURE 6.4 – Predicted distribution of inter-night gaps for the `baseline2018a` observing strategy for the LSST Wide, Fast, Deep survey in the i-band. Between each observation and the next, we draw the duration of the inter-night gap from this distribution, leading to an irregular cadence.

subtraction of the lens and the host galaxy, while in reality, image subtraction artefacts often arise due to varying seeing conditions.

For this project, we limit ourselves to images taken in the i-band, which is sufficiently red to clearly detect high-redshift supernovae, while also being assigned more frequent visits than the z or y-band. The precise cadence of LSST will have a significant impact on the number of detected lensed supernovae, with high sampling frequencies and high cumulative season length leading to a more favourable scenario (Huber et al., 2019). However, as of date, the final cadence strategy has not yet been decided upon. For our simulated time series, we use predictions from the `OpSim` scheduler[4] for the Wide, Fast, Deep (WFD) survey following the `baseline2018a` observing strategy. Figure 6.4 shows the predicted distribution of inter-night gaps for the LSST i-band, which has a median of 11.0 days between observations. As a consequence of the irregular cadence and varying duration of the lensed transients, the simulated time series have different time intervals between observations, as well as a different total number of observations.

The Vera Rubin Observatory will have a median seeing, or full width at half maximum of the point spread function (PSF), of 0.75 arsec in the i-band. Since the exact seeing for each night depends on the weather conditions on the site in Cerro Pachón in Chile, we draw a different PSF value for each observation from the predicted seeing distribution from the LSST observation simulator (T. E. Collett, 2015; Connolly et al., 2010). Table 6.2 features the PSF distribution adopted for the image simulation procedure, which is obtained by fitting a skewed normal distribution to the predicted seeing values.

The noise levels in LSST images originate from Poissonian count statistics and background noise, where the latter consists of read noise (10 e^- per exposure) and sky background (20.48

[4] https://cadence-hackathon.readthedocs.io/en/latest/current_runs.html

mag/arcsecond^2). An example of a simulated time series of images under LSST-like settings, yet with a different, fixed cadence, is displayed in Figure 6.1.

6.2.4 Predicted lensed supernova rates

The predicted lensed type Ia supernova rate for LSST, when both the magnification method (Goldstein & Nugent, 2017), which looks for objects that appear significantly brighter than expected, and the image multiplicity method (Oguri & Marshall, 2010) are considered, is 89 events per year (Wojtak et al., 2019). However, lensed supernovae discovered via the image multiplicity method are the only ones able to provide a measurement of the time delay between images, which is crucial for cosmological inference. The latter discovery channel is forecasted to detect around 44 events per year in the i-band. In our work, we take the aforementioned prediction from Wojtak et al. (2019) obtained with the image multiplicity method as a starting point, following the detection criteria proposed in Oguri and Marshall (2010):

- The maximum separation between images $\theta_{\max}$ should be between $0.5''$ and $4.0''$. A lower $\theta_{\max}$ implies that the images cannot be resolved with typical seeing conditions, while a higher $\theta_{\max}$ would correspond to an object lensed by a galaxy cluster rather than a single massive galaxy, as is the focus of this work.
- For doubly imaged supernovae (doubles), the flux ratio between the images must exceed 0.1 to avoid problems with the dynamic range of the telescope.
- Both images of doubles should be brighter than the detection limit of LSST (23.9 magnitudes). For quadruply imaged supernovae (quads), at least three images should be detected.

The predicted 44 events per year (of which ∼34 are doubles and ∼9 are quads) by Wojtak et al. (2019) include systems with very short time delays of the order of several days. However, due to the median cadence of LSST of 11 days, it will be highly challenging to obtain accurate time delay measurements for such systems. Therefore, we impose an additional criterion in our study:

- The maximum time delay of the lensed supernova should exceed 2 days, but have a duration below 150 days. We set the lower limit rather low so as to avoid cutting down our sample size too severely. The upper limit stems from the fact that higher time delays will fall outside the range of our simulated observations.

6.2.5 H_0 inference pipeline

We employ a 3D CNN, as described in more detail in Section 2.2.1, to process the time-series images, in combination with a simulation-based inference (SBI) framework, as outlined in Section 2.3.2, to infer robust uncertainties on the network predictions. For our specific problem of inferring the Hubble constant, the first step involves the generation of the time-series images of an ensemble of lensed supernovae via our image simulation pipeline. The latter can be considered as our physical model, $\mathcal{F}(\boldsymbol{\phi}) : \boldsymbol{\theta} \rightarrow \boldsymbol{D}$, which generates the data $\boldsymbol{D}$ from a set of model parameters $\boldsymbol{\theta}$ and some initial conditions $\boldsymbol{\phi}$. In our set-up, $\boldsymbol{\theta} \equiv \{H_0\}$ corresponds to the present-day value of the Hubble constant, while $\boldsymbol{\phi}$ encompasses all the relevant input parameters to the simulator, such as the lens and source redshift, source positions and the mass profile of the lens. We generate a training set that follows a uniform distribution in the Hubble constant: $H_0 \sim \mathcal{U}(20, 100)$. The broad range of H_0 values is adopted to ensure that we do not leak any additional cosmological information to the network. Subsequently, we train our spatio-temporal network, $\mathbb{NN}(\boldsymbol{\omega}, \boldsymbol{\eta}) : \boldsymbol{D} \to \tilde{\boldsymbol{d}}$, to yield a set of network predicted summaries $\tilde{\boldsymbol{d}}$ from the input data $\boldsymbol{D}$.

The majority of lens systems produce either doubly imaged supernovae or quadruply imaged ones, depending on the position of the source behind the lens galaxy. The resulting time-series images for doubles and quads vary significantly in their properties; most notably, they have a different number of images and, therefore, a different number of time delays between the images. For this reason, we employ two separate neural networks to process input data from doubles and from quads, such that the networks can specialise in their respective lensing configurations. For doubles, the time-delay distance $D_{\Delta t}$ can be determined analytically from estimates of the lens model (θ_E, e_1, e_2, γ_lens), the image positions ($\boldsymbol{x}_\mathrm{im}$, $\boldsymbol{y}_\mathrm{im}$), the source position ($x_\mathrm{src}$, y_src) and the time delay (Δt) between the images, via eq. (1.74). Since quads have three independent time delays (Δt_1, Δt_2 and Δt_3) between their four images, the computation of the time-delay distance is less trivial. In our study, we opt for an approach where the neural network predicts the relevant lensing parameters, after which we use only the information from the maximum time delay Δt_max between the first and the last image to calculate the time-delay distance analytically. We find that this method yields superior results to direct predictions by the neural network of the time-delay distance, even though this approach discards additional constraints from the time delays between the other image combinations. As such, the choice of network summaries for doubles and quads corresponds to $\tilde{\boldsymbol{d}} \equiv \{\Delta t_\mathrm{max}, \theta_\mathrm{E}, \gamma, e_1, e_2, x_\mathrm{src}, y_\mathrm{src}, \boldsymbol{x}_\mathrm{im}, \boldsymbol{y}_\mathrm{im}\}$, which are subsequently converted analytically into an estimate of the time-delay distance.

The final step of the pipeline is to convert $D_{\Delta t}$ analytically into H_0 via eq. (1.76) and a cosmological model to relate angular diameter distances to the Hubble constant. We adopt a standard ΛCDM model with $\Omega_\mathrm{m} = 0.3$ and $\Omega_\mathrm{k} = 0.0$, which has also been employed in the image simulation procedure. Additionally, this step requires information about the source and the lens redshifts, for which we supply the ground truth values of z_lens and z_src. We thereby assume that spectroscopic redshift measurements of the lens and the host galaxy will be carried out after the supernova has faded away. The redshifts are only provided in this final step and not fed as input to the CNNs, ensuring no leakage of cosmological information into the final results. We subsequently characterise the joint probability density function (PDF) of the predicted and ground truth H_0 values, $\mathcal{P}(H_0^\mathrm{pred}, H_0^\mathrm{true})$, using a Gaussian kernel density estimator (KDE). The 2D PDF for a test set of five thousand doubly lensed supernova systems is depicted in Figure 6.5. Finally, for any observed time series of images $\boldsymbol{D}_\mathrm{obs}$, we obtain the approximate posterior by slicing the joint PDF at the predicted H_0 value corresponding to the network predicted summaries, $\mathbb{NN}(\boldsymbol{\omega}, \boldsymbol{\eta}) : \boldsymbol{D}_\mathrm{obs} \to \tilde{\boldsymbol{d}}_\mathrm{obs}$, as follows:

$$\mathcal{P}(H_0|\tilde{\boldsymbol{d}}_\mathrm{obs}) \approx \mathcal{P}(H_0|\boldsymbol{D}_\mathrm{obs}, \boldsymbol{\omega}, \boldsymbol{\eta}). \tag{6.5}$$

Figure 6.6 illustrates a schematic overview of our pipeline to infer a single H_0 estimate from the time series of lensed supernovae images from doubles and quads.

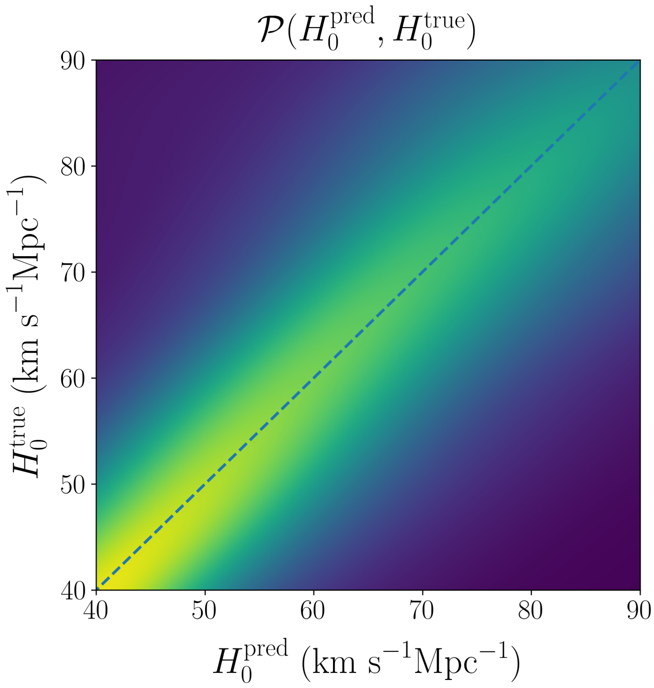


FIGURE 6.5 – Joint 2D PDF of predicted and true H_0 values, computed using a test set of five thousand doubly imaged supernovae. We obtain the approximate posterior $\mathcal{P}(H_0^{\rm pred}, H_0^{\rm true})$ for a particular lens system by making a vertical slice at the predicted value of H_0.

6.2.6 Network architecture and training

We implement our CNNs using the KERAS framework (Chollet et al., 2015) via a TENSORFLOW backend (Abadi et al., 2016). To train the neural networks, we use the ADAM (Kingma & Ba, 2014) optimiser with a learning rate of $\eta = 10^{-4}$, and the mean squared error loss function designed for regression problems. The batch size is set to 100 and we adopt the early stopping technique to avoid overfitting. We designate a random 25% of our original training data set as a validation set and terminate the training process when the validation loss shows no improvement for 10 consecutive epochs.

Our training set consists of 45 thousand lensed supernova systems, each generated with a different value of the Hubble constant. The test set used to construct the posterior distributions via SBI contains five thousand lens systems. A second test set of five thousand systems is used to create the posterior calibration plot as detailed in Section 6.3.3 and illustrated in Figure 6.12. Finally, we simulate a realistic evaluation set of 40 lens systems per year, all with a fixed Hubble constant of $H_0 = 70.0$ km s^{-1}Mpc^{-1}, to assess whether our CNN pipeline is able to recover this ground truth value from the ensemble of lens systems and to verify the constraining power of our inference machinery.

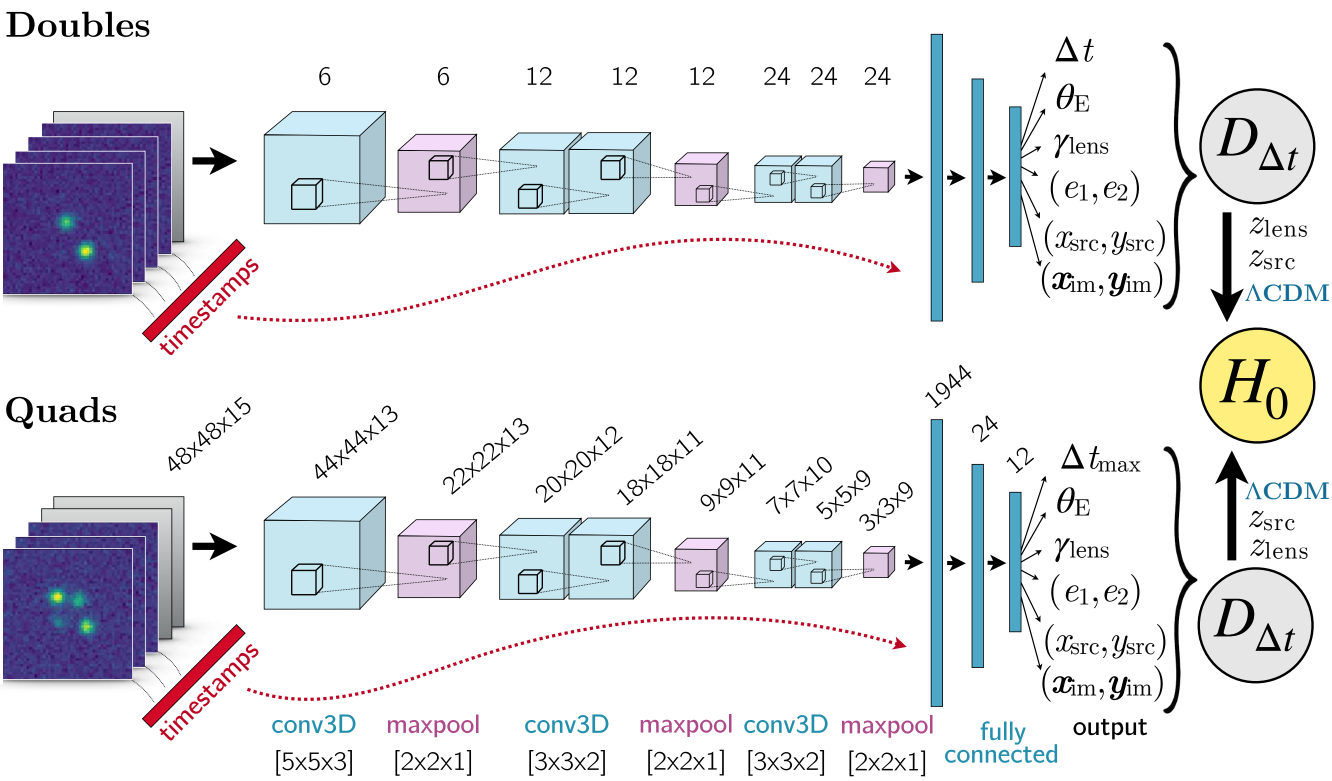


FIGURE 6.6 – Schematic representation of the pipeline designed to infer the cosmic expansion rate from simulated images of gravitationally lensed supernovae. The input data consists of time-series images, supplemented with empty filler layers to obtain a dimension of 48×48×15 for each system, and a vector containing the timestamps of the observations. The timestamp vector is concatenated with the feature maps after they have been flattened and fed to the fully connected network. A spatio-temporal neural network acts as a 3D convolutional feature extractor to convert the input data into estimates of the time delay and lens parameters, which are analytically converted into an estimate of the time-delay distance. In the final step of the pipeline, the time-delay distances are combined with the ground truth values of the lens and source redshift, under the assumption of a ΛCDM model, to calculate H_0. We employ an identical neural network architecture for doubles and quads. The number of feature maps is indicated in the first row, the resulting dimensions of the images and feature maps in the second one, and the kernel size of the convolutional and maxpooling layers in the final row.

Before the input images are fed to the neural network, each pixel is transformed according to $\log(a_i + D_i)$, where D_i denotes the collection of pixels corresponding to a specific lens system i, and a_i is the minimum point for the log-transformation, calculated for each lens system as $1 - \min(D_i)$. This ensures that the log-transformation is only applied to positive values and not to the negative pixels that originate from the observational noise. The log-transformation is included such that the pixels from the brightest image at the peak of the light curve do not outshine the rest of the supernova time evolution.

As depicted in Figure 6.6, the architecture of the CNNs for both the doubles and the quads consists of three blocks of convolutional layers, whereby the first block consists of one convolutional layer that employs a kernel size of $5 \times 5 \times 3$, and the second and third block both have two convolutional layers with kernel sizes of $3 \times 3 \times 2$. Every layer consists of several kernels, amounting to 6, 12, and 24 kernels for the first, second and third block, respectively. Following each block of convolutions, a $2 \times 2 \times 1$ maxpooling kernel is used to reduce the size of the feature maps. After the convolutional and maxpooling layers, the feature maps are flattened to a 2D vector and passed through two fully-connected layers with 24 and 12 neurons to eventually yield the output summary $\tilde{\boldsymbol{d}}$. We use the non-linear rectified linear unit (ReLU; cf. eq. (2.11)) as activation function throughout the whole network, except for the output layers, where we use a softplus activation function for output parameters that should be strictly positive (i.e. Δt, θ_E, γ), and linear activations for the remaining parameters. The resulting network has a relatively low complexity, with $\sim 40,000$ trainable parameters.

Although a CNN requires a pre-defined input size for all input images, in our case the number of observations differs per system. In order to circumvent this problem, we add filler layers consisting only of zeros to fill up the missing observations. We impose a maximum number of 15 observations, discarding images beyond this limit. Alongside the time-series image input, the neural network also needs information regarding the exact timestamps of the observations, in order to account for the irregular cadence. We provide this in the form of a vector containing the timestamps for each observation, which we concatenate with the feature maps after they have been flattened.

6.3 RESULTS

FIGURE 6.7 displays the neural network predictions for doubles and quads in terms of the time delay, source and image positions, and the lens parameters. The estimates of the aforementioned parameters are subsequently employed to calculate the time-delay distance and, ultimately, the Hubble constant. From the scatter plots, it can be seen that doubly imaged lens systems provide better estimations of the time delay and the image positions, while quads display tighter predictions for the lens model parameters θ_{E} and e_1. These results match our expectations, since the four images of quads provide useful additional information to constrain the lens model, while concurrently rendering it more challenging to infer time delays and image positions as the four images are more prone to overlap. The network is not able to provide any predictive power on γ_{lens}, neither for quads nor for doubles, and instead,it simply returns the underlying prior distribution. This is not a surprising finding, since the power-law slope is generally constrained with information from the pixels in the Einstein ring corresponding to the host galaxy (Park et al., 2021), which are not included in our difference images.

6.3.1 Individual parameter contributions to residual scatter in H_0

After converting the estimates of the parameters shown in Figure 6.7 into a final prediction of the Hubble constant, we quantify the individual contribution from each parameter to the residual scatter in order to get insight into the dominant sources of uncertainty on H_0.

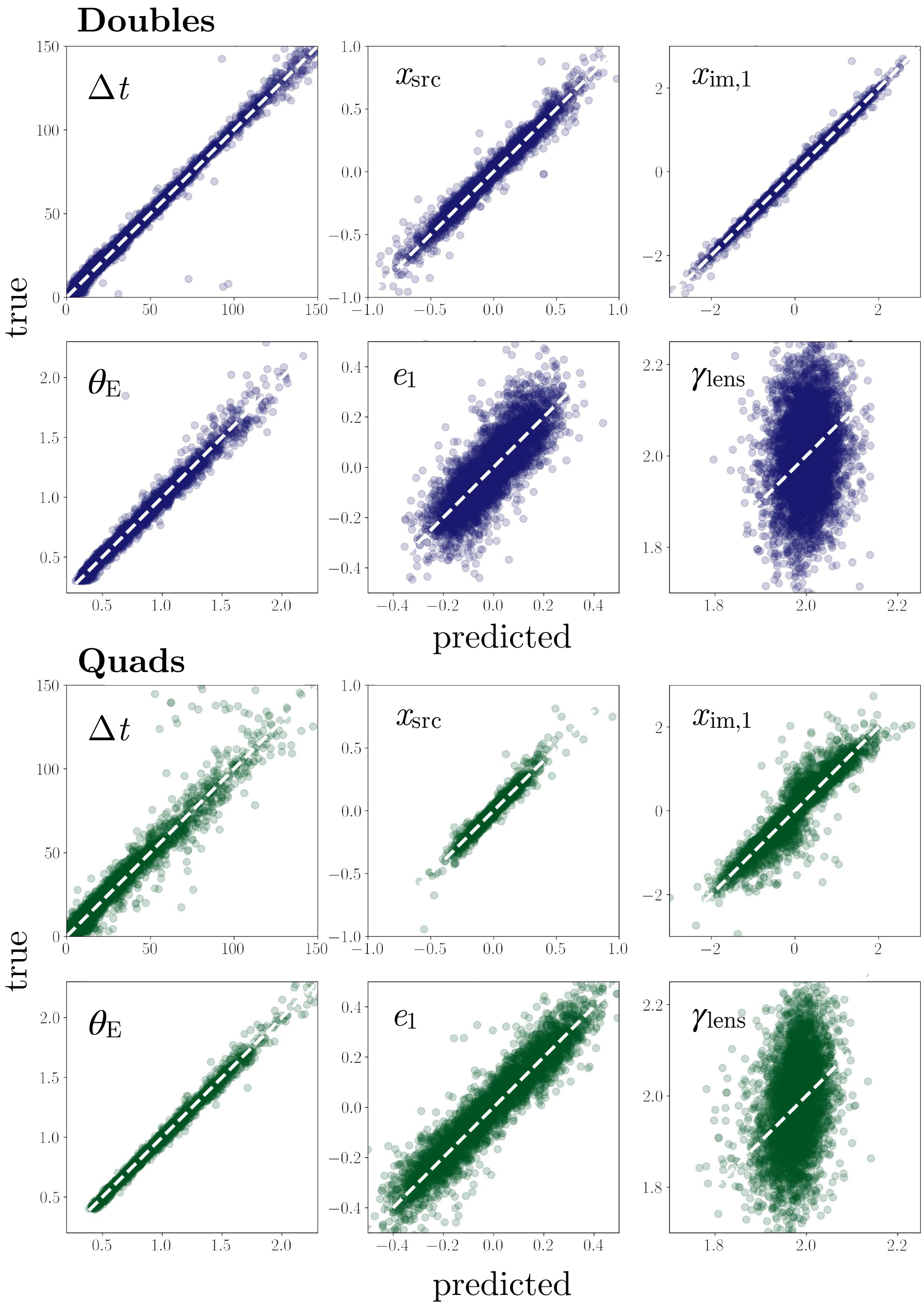


FIGURE 6.7 – Scatter plots of the network predictions for the parameters of interest for doubles and quads compared to their ground truth values. The white dashed lines indicate perfect predictions. The results for y_{src}, e_2, and the remaining image coordinates are similar to those for x_{src}, e_1, and $x_{\mathrm{im},1}$ The network predictions for doubles are more accurate in terms of the time delay and the image positions, while those for quads show tighter constraints on the lens model parameters θ_{E} and e_1.

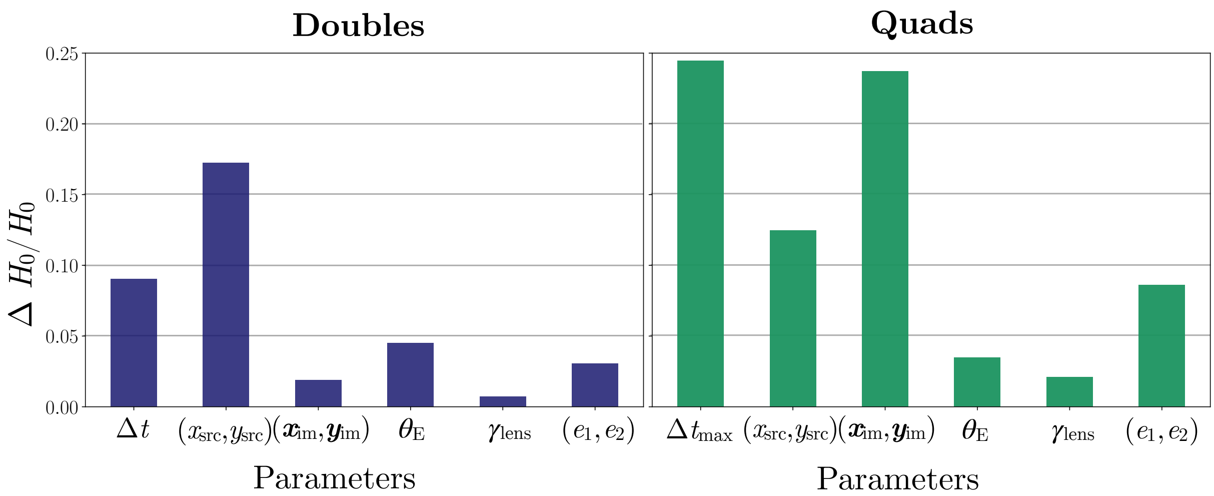


FIGURE 6.8 – Contributions from individual parameters to the total residual scatter in H_0 for doubly and quadruply imaged lens systems. The source position (x_{src}, y_{src}) and the time delay Δt are responsible for the largest uncertainty for doubles, while the image positions $(\boldsymbol{x}_{im}, \boldsymbol{y}_{im})$ and maximum time delay Δt_{max} induce the largest scatter for quads. The Einstein radius θ_E, ellipticity (e_1, e_2), and the power-law slope γ_{lens} contribute the least to the residual scatter.

We perform this analysis by computing H_0 from the *estimated* value of a certain parameter X, in combination with the *ground truth* values of the remaining parameters. In this way, the resulting scatter can only be attributed to parameter X. Figure 6.8 summarises the residual scatter contributions for each of the parameters of interest. For doubles, the source position and time delay are responsible for the largest contributions, while they yield tight predictions in the scatter plots of Figure 6.7. This demonstrates that the uncertainties in (x_{src}, y_{src}) and Δt have a much stronger influence on the final H_0 estimate than the uncertainty on any of the other parameters. The residual scatter in H_0 predictions from quads is mainly due to the image positions and the time delay. Let us highlight as well that the large spread in the power-law slope predictions does not interfere too critically with the final Hubble constant estimate, since γ_{lens} has the lowest contribution to the residual scatter.

Next, we investigate the residual scatter in H_0 predictions as a function of maximum time delay between the supernova images, depicted in Figure 6.9. As seen from the resulting trends, both doubles and quads display a high residual scatter for low time delay values, even after we discarded all systems with $\Delta t_{max} < 2$ days (cf. Section 6.2.4). This is in line with our expectations, since the median LSST i-band cadence of 11 days does not allow for a precise Δt measurement for short time delays. For quads, the scatter in Δt can be seen to increase again for very high time delay values, this effect can most likely be attributed to our cut off in the input image size. We imposed a maximum of 15 observations, which means that time-series images of systems with very long time delays will be based on incomplete light curves, reducing the amount of information available to the neural network. Doubly imaged lens systems appear to be less susceptible to this effect, possibly because for Δt predictions for two images, it is sufficient to only see the beginning of the light curve.

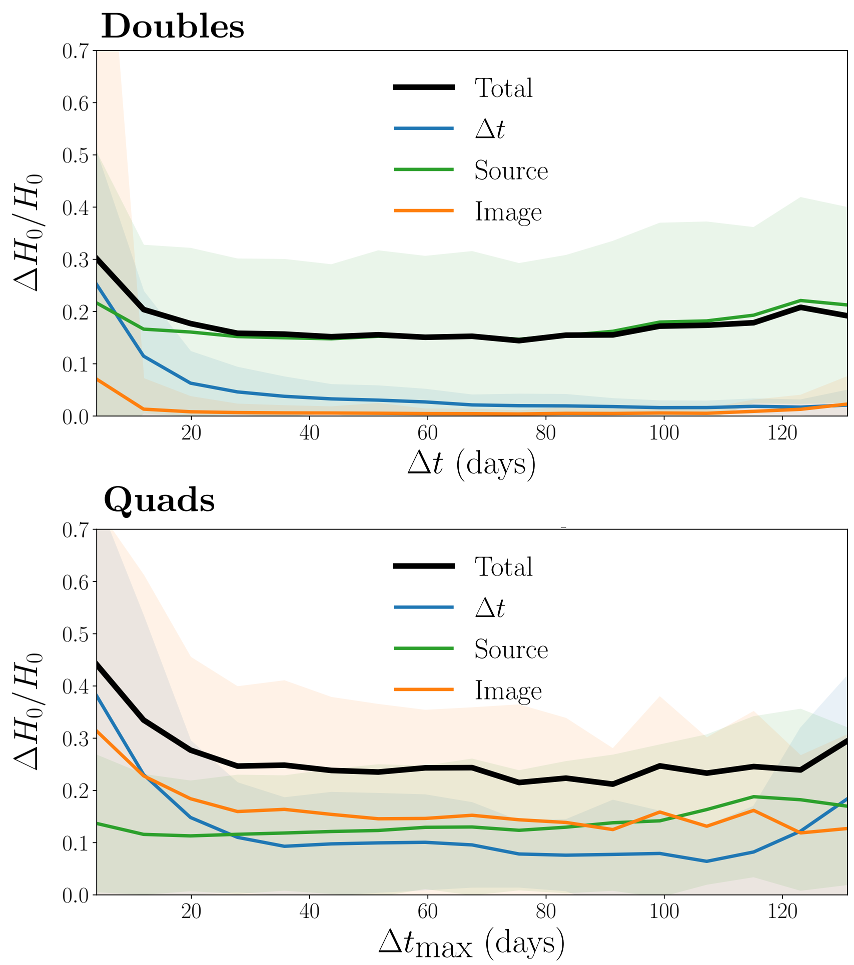


FIGURE 6.9 – The residual scatter in H_0, defined as $\Delta H_0/H_0 = |H_0^{\mathrm{true}} - H_0^{\mathrm{pred}}| \,/\, H_0^{\mathrm{true}}$, as a function of maximum time delay between supernova images. The coloured lines depict the individual contributions to the scatter from the time delay, source and image positions, which are the dominant sources of uncertainty on H_0, with the shaded regions corresponding to the standard deviation of the residual scatter. Low time delays correspond to a higher uncertainty on Δt and hence in H_0, since the LSST median cadence of 11 days is too low to attain sufficient precision on the time delay estimates. Extreme outliers with $H_0^{\mathrm{pred}} < 0$ or $H_0^{\mathrm{pred}} > 140$ have been removed from the analysis to avoid them completely dominating the results.

A transition can be seen to occur between low time delays, where Δt is the dominant source of uncertainty, and larger time delay values, where contributions from the source position (for doubles) or the image position (for quads) dominate the scatter. Let us emphasise that the majority of the lensed supernova systems have rather low time delays; around 50% of the systems have $\Delta t_{\max} < 25$ days. Therefore, it is important to find methods to decrease the uncertainty on Δt in order to improve the final H_0 predictions.

6.3.2 Joint H_0 inference

After incorporating the cuts in the time-delay distance as described in Section 6.2.4 into the predicted LSST lensed supernova rate from Wojtak et al. (2019), we forecast that around 40 type Ia supernovae per year will be discovered with the image multiplicity method that are suitable for our H_0 inference pipeline. Of this sample, ~31 are expected to be doubles and ~9 quads. These rates are used to construct a realistic evaluation set, with a ground truth value of $H_0 = 70$ km s^{-1}Mpc^{-1}.

Following these numbers, an evaluation set of 310 doubles and 90 quads is employed to represent the ten-year duration that LSST will be operational. For each lensed supernova in the sample, we determine an H_0 posterior through our pipeline outlined in Figure 6.6 in combination with the SBI framework. When the individual posteriors from the evaluation sample are combined into a joint PDF, we report a final estimate of $H_0 = 70.2^{+0.8}_{-0.9}$ km s^{-1}Mpc^{-1}, as presented in Figure 6.10. The combined posterior constitutes a 1.2% measurement of H_0, which is consistent with the ground truth and shows no detectable bias. However, the precision and accuracy of the measurement is completely driven by the doubles in the sample. The quads display a bias towards lower H_0 values, possibly due to our underlying prior of $H_0 \sim \mathcal{U}(20, 200)$. Seemingly, our current approach for employing quads, i.e. only considering the time delay constraints between the first and the last image, is not sufficient for them to contribute meaningfully to cosmological inference. Alternative methods that optimally exploit the information content of all four images might render them more useful, although the expected sample size of quads remains small.

Figure 6.11 displays the predicted joint H_0 posterior after one year of LSST observations, corresponding to a 4.3% unbiased measurement of the Hubble constant. Again, the majority of the constraining power emanates from doubly imaged supernovae.

6.3.3 Posterior validation

If the assigned posteriors are a realistic representation of the underlying empirical distribution of H_0 values, x% of their probability volume should contain the true value x% of the time. In order to verify this, we perform a calibration test as described in recent studies (Ho et al., 2020; Wagner-Carena et al., 2021). Briefly, for a predictive percentile p and for each lens system, we integrate our model posterior until the resulting area equals p and obtain the associated value of $H_{0,p}$. We then compute the fraction of lens systems in our test set with H_0 smaller than $H_{0,p}$, which corresponds to the empirical percentile $\hat{p}$. If our posteriors are perfectly calibrated, all the predictive percentiles will be identical to the empirical percentiles. For a posterior biased toward low H_0 predictions, we have $\hat{p} < p$, while for a posterior biased toward high H_0 values, we have $\hat{p} > p$ for the whole percentile range. A posterior that underestimates uncertainties has $\hat{p} > p$ for $[0, 50\%]$ and $\hat{p} < p$ for $[50, 100\%]$, and vice versa for overestimation. Figure 6.12 displays the resulting calibration curves for our model posteriors from both double and quadruple lens systems. The curves demonstrate a nearly perfect calibration, thus justifying the use of posteriors inferred via our SBI approach throughout this work.

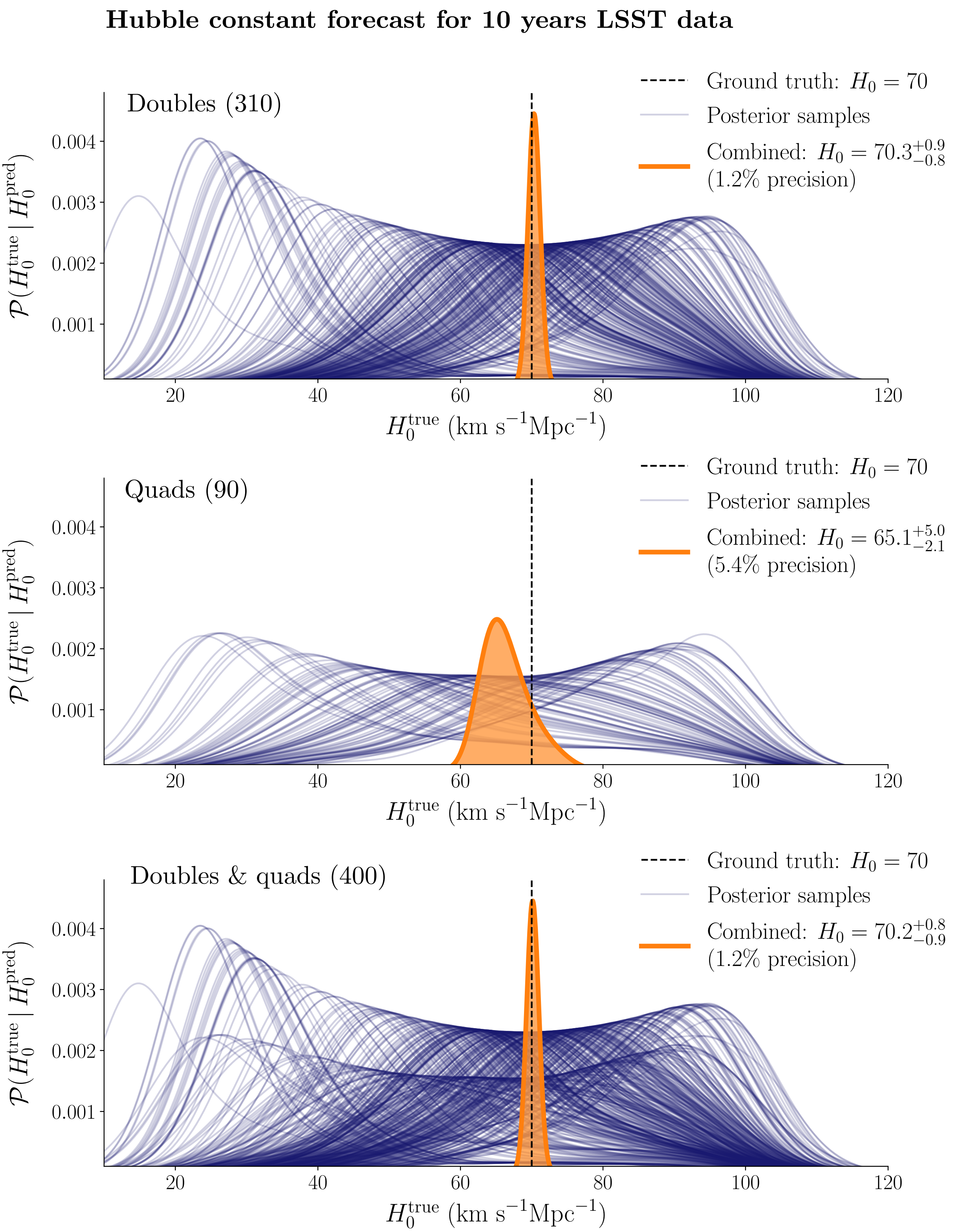


FIGURE 6.10 – H_0 posteriors inferred with SBI for simulations of 310 doubles and 90 quads, which constitute a realistic lensed supernova sample corresponding to 10 years of LSST operations. The sample was generated with a ground truth of $H_0 = 70$ km s^{-1}Mpc^{-1}. By combining the posteriors in a joint analysis, a 1.2% unbiased estimate of the Hubble constant is obtained, whereby the constraining power is completely driven by doubly imaged systems.

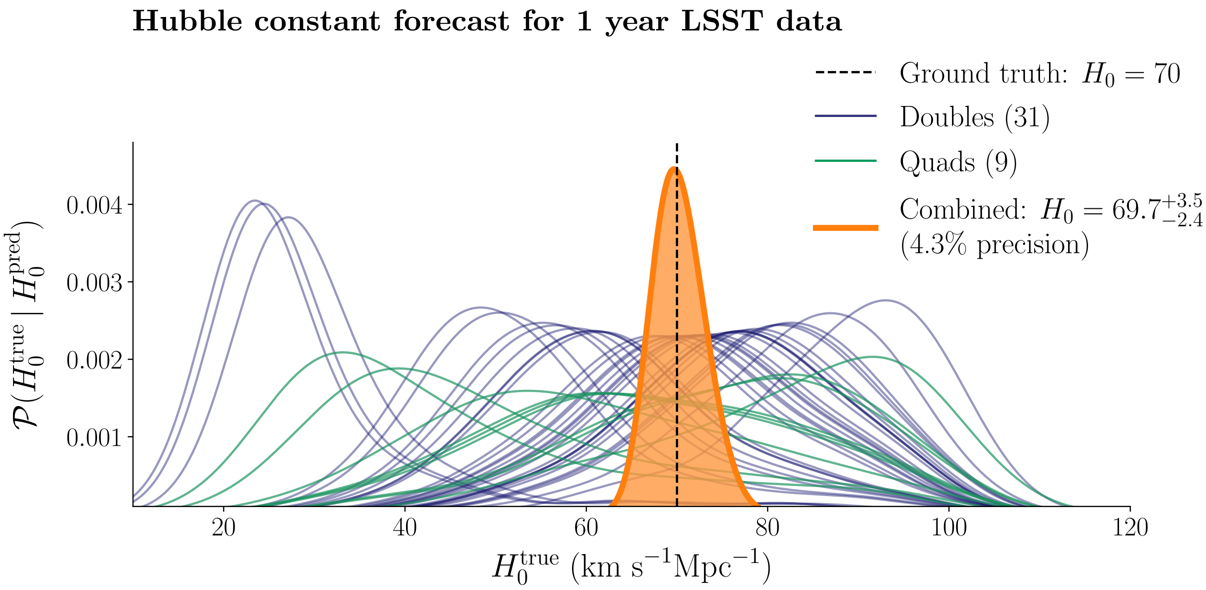


FIGURE 6.11 – H_0 posteriors inferred with SBI for a simulated sample of 31 doubles and 9 quads, corresponding to the expected lensed type Ia supernova yield from LSST after 1 year of observations. The sample was simulated with a ground truth of $H_0 = 70$ km s^{-1}Mpc^{-1}. The joint posterior provides a 4.3% unbiased measurement of the Hubble constant.

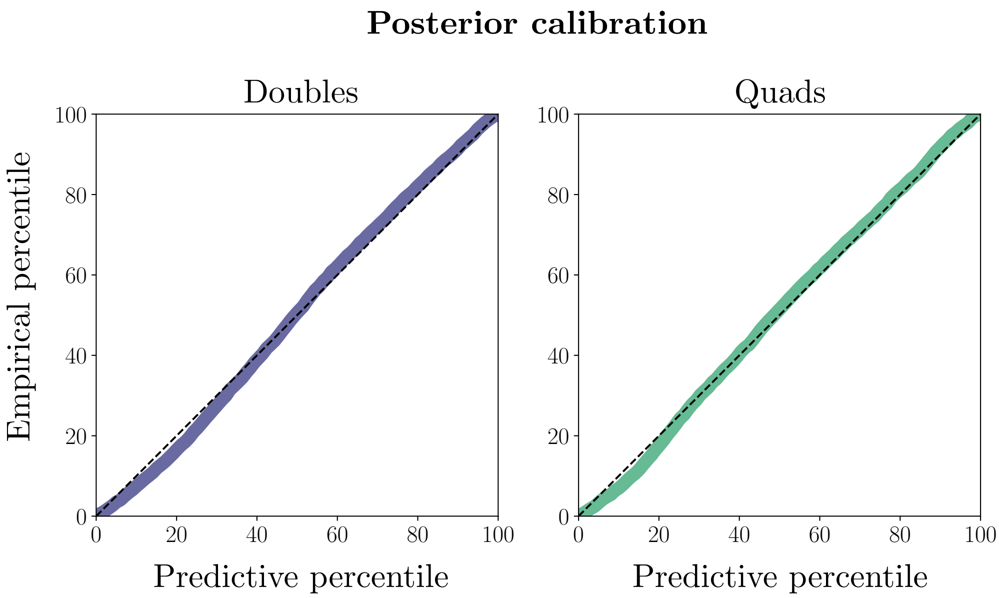


FIGURE 6.12 – Posterior calibration plots (or quantile-quantile plots) indicating how well the SBI-inferred posteriors correspond to the empirical distribution of true H_0 values. The dashed lines represent perfect calibration and the coloured lines depict the calibration of the inferred posterior PDFs. This plot illustrates the overall statistical consistency of our SBI-inferred posteriors, thereby justifying their use throughout our work.

6.4 DISCUSSION AND CONCLUSIONS

We have designed a spatio-temporal neural network, implemented within a simulation-based inference framework, to infer H_0 posteriors from time series of lensed supernova images. Specifically, our spatio-temporal model is tailored for the LSST context, and hence, we use realistic LSST-like simulated images and predicted lensed supernova rates in our study. After discarding the objects with time delays $\Delta t < 2$ days and $\Delta t > 150$ days from the predicted rates of Wojtak et al. (2019), we forecast to find ~ 40 lensed type Ia supernovae per year that are suitable for cosmological inference, of which ~ 31 are expected to be doubly imaged and ~ 9 quadruply imaged. Combining the posteriors from this sample, we obtain a 4.3% unbiased estimate of the Hubble constant corresponding to a time period of one year. After 10 years of LSST operations, the precision in H_0 from lensed supernovae in LSST is predicted to be 1.2%, whereby both the accuracy and precision are completely driven by doubles. In our results, quads are biased towards a lower value of H_0 and entail marginal constraining power. We find that quads are not as well-suited for cosmological inference as doubles, because of their high residual scatter in time delay and image position measurements, and their much smaller expected sample size. The approach we adopted in this study to use quads, namely calculating $D_{\Delta t}$ from the time delay between the first and the last image, may not be the optimal one, and alternative avenues that employ the full information content of the four images should be explored in further investigations.

The main contributions to the residual scatter in H_0 for doubles are the source position and the time delay measurement, indicating that a higher cadence has the potential to improve the results significantly. As a proof of concept, we limited ourselves to i-band data, but our network architecture can easily be modified to have multiple channels as input, each corresponding to a different band. Including the r, z and y bands will considerably improve the cadence, and consequently, the precision on the time delay estimates and the Hubble constant. Additionally, we only considered type Ia supernovae, which will constitute $\sim 26\%$ of the lensed supernova population, but including other types of supernovae such as IIns, which are predicted to make up $\sim 61\%$, would give us a much larger sample size.

In our current work, we neglect the effects of microlensing from substructures in the lens galaxy. However, microlensing effects could alter the light curves in a substantial way, especially for lensed supernovae as they typically have smaller angular separations from the lens galaxy than lensed quasars (Dobler & Keeton, 2006; Huber et al., 2019). Since microlensing can perturb light curves independently in each image, it can lead to additional systematic errors in time delays of order 4% (Goldstein et al., 2018). To quantify the microlensing effect in a realistic way, we are presently working together with Simon Huber to develop an approximate description of microlensing contributions to the supernovae light curves. This description will be based on the approach followed in Goldstein et al. (2018), Huber, Suyu, Ghoshdastidar, et al. (2021), and Huber, Suyu, Noebauer, Chan, et al. (2021), wherein a microlensing magnification map is combined with radiative transfer models of the expanding supernova photosphere.

The PEMD + external shear model adopted in our study to simulate the lensed supernovae generally provides an adequate description of stellar kinematics (Koopmans et al., 2009) and

X-ray observations (Humphrey & Buote, 2010) in the local Universe, while using a low number of degrees of freedom. However, it has recently been shown (Birrer et al., 2020; Kochanek, 2020; Sonnenfeld, 2018) that it can lead to over-constrained mass profiles and can possibly bias the estimates of the radial density profile, and consequently, H_0. As further investigation for this work, more complex lens models or non-parametric approaches to describe the lens potential can be considered.

An additional avenue to make our simulated difference images more realistic is by performing a manual subtraction of the reference image, instead of assuming a perfect subtraction of the lens and host galaxy. This would include some image subtraction artefacts into our simulated images, thereby making them more similar to real transient survey images. Training the network on such a sample will render it more robust to image subtraction artefacts in the real data.

In order to simulate the lensed supernova images and to convert the final estimates of the time-delay distance into the Hubble constant, a standard ΛCDM model with $\Omega_\mathrm{m} = 0.3$ and $\Omega_\mathrm{k} = 0.0$ is adopted throughout our study. This makes our results dependent on the choice of cosmological model. An alternative would be to generate the data while assuming a distribution of Ω_m and Ω_k values, ideally covering a wide range of values to account for the variety of possible underlying cosmologies. Subsequently, a cosmology-independent method could be used instead of the Friedmann equation to convert angular diameter distances into H_0, such as the polynomial parametrisations employed in Chapters 3 and 4.

We conclude by emphasising that although LSST will not be ideal for single lensed supernova systems and will require follow-up observations for high precision time delay constraints, this work underlines the usefulness of pure LSST data in a joint population analysis. Our pipeline enables the use of lensed supernovae as cosmological probes in a way that is complementary to existing H_0 measurements, without the requirement of challenging follow-up observations. Additional complexity, such as the topics discussed above, can be easily incorporated into our framework to quantify their effects on the final precision in H_0. An exciting application of our pipeline is to combine it with the spatio-temporal engine for finding gravitationally lensed supernovae, as presented in our recent work (Kodi Ramanah, Arendse, et al., 2021), which reaches a classification accuracy of around 99% for LSST data. After identifying the lensed supernovae, their time-series images can be fed to the second spatio-temporal network presented here to obtain the corresponding H_0 posteriors.

AUTHOR CONTRIBUTIONS

NA led the project, built the image time-series and supernova light curve simulation pipeline, designed and optimised the neural network architecture, implemented the simulation-based inference framework, performed the analysis, produced the figures, wrote the manuscript. DKR proposed the main idea, validated the image simulation pipeline, assisted with the network optimisation and use of the Infinity Cluster, contributed to writing-up the manuscript. RW proposed the main idea, validated the image simulation pipeline and contributed to writing-up the manuscript.

ACKNOWLEDGEMENTS

We thank Ji Won Park for useful suggestions regarding lens mass model inferences and Alexander Bauer for inspiring discussions and insights. This work was supported by a VILLUM FONDEN Investigator grant (project number 16599). DKR is a DARK fellow supported by a Semper Ardens grant from the Carlsberg Foundation (reference CF15-0384). This work made use of the Infinity Cluster hosted by Institut d'Astrophysique de Paris and an HPC facility funded by a grant from VILLUM FONDEN (project number 16599).

III

Galaxy cluster masses

CHAPTER 7

Simulation-based inference of galaxy cluster masses

This chapter is based on the following article:
"Simulation-based inference of dynamical galaxy cluster masses with 3D convolutional neural networks"

Published in Monthly Notices of the Royal Astronomical Society (MNRAS), Volume 501, Issue 3, pp. 4080-4091 (2021).

Authors: Doogesh Kodi Ramanah, Radosław J. Wojtak, Nikki Arendse

Abstract

We present a simulation-based inference framework using a convolutional neural network to infer dynamical masses of galaxy clusters from their observed 3D projected phase-space distribution, which consists of the projected galaxy positions in the sky and their line-of-sight velocities. By formulating the mass estimation problem within this simulation-based inference framework, we are able to quantify the uncertainties on the inferred masses in a straightforward and robust way. We generate a realistic mock catalogue emulating the Sloan Digital Sky Survey (SDSS) Legacy spectroscopic observations (the main galaxy sample) for redshifts $z \lesssim 0.09$ and explicitly illustrate the challenges posed by interloper (non-member) galaxies for cluster mass estimation from actual observations. Our approach constitutes the first optimal machine learning-based exploitation of the information content of the full 3D projected phase-space distribution, including both the virialised and infall cluster regions, for the inference of dynamical cluster masses. We also present, for the first time, the application of a simulation-based

inference machinery to obtain dynamical masses of around 800 galaxy clusters found in the SDSS Legacy Survey, and show that the resulting mass estimates are consistent with mass measurements from the literature.

7.1 INTRODUCTION

Galaxy clusters are formed by the collapse of high density regions in the early Universe, and they are important to study the formation and evolution of large-scale cosmic structures. The cluster abundance as a function of mass and its evolution are sensitive to the amplitude of density perturbations and to the properties of dark matter and dark energy. Galaxy clusters can therefore provide competitive cosmological constraints that are complementary to other cosmological probes. As future surveys, such as the Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al., 2016), the Vera C. Rubin Observatory (Ivezic et al., 2008), Euclid (Racca et al., 2016) and eROSITA (Merloni et al., 2012), will provide unprecedented volumes of data extending to high redshifts, the accuracy and precision of cluster mass estimation techniques will become crucial. With the ever increasing scale of state-of-the-art cosmological simulations (e.g. Ishiyama et al., 2020; Villaescusa-Navarro et al., 2020) providing considerable volumes of training data, along with the limitations of traditional techniques, the use of machine learning (ML) algorithms to infer cluster masses has become an increasingly attractive and viable option (e.g. Armitage et al., 2019; Calderon & Berlind, 2019; Ho et al., 2020; Ho et al., 2019; Kodi Ramanah, Wojtak, Ansari, et al., 2020; Ntampaka et al., 2015; Ntampaka et al., 2016; Sutherland et al., 2012; Yan et al., 2020). These models are typically trained on a large simulated data set, such that the algorithm learns the connection between the observables and cluster masses. Once optimised, they can subsequently be used to predict masses for unseen data, provided that the simulations used for training are sufficiently accurate to replicate the characteristics of the galaxy survey of interest with high fidelity (Cohn & Battaglia, 2020).

For observations probing galaxy kinematics in galaxy clusters, ML methods offer a promising alternative to traditional methods of cluster mass estimation which are usually based on scaling relations, the virial theorem or the Jeans equation, and are limited by several assumptions, primarily involving dynamical equilibrium and spherical symmetry, as briefly reviewed in Kodi Ramanah, Wojtak, Ansari, et al. (2020). Recently, convolutional neural networks (CNNs), by virtue of their sensitivity to visual features, have been applied by Ho et al. (2019) to obtain accurate dynamical mass estimates of galaxy clusters in spectroscopic surveys. The network inputs are images generated by a kernel density estimator from the 2D projected phase-space distributions defined by the cluster-centric projected distance and line-of-sight velocities of galaxies observed in the fields of clusters. The challenge with ML methods is often to not only produce single point estimates, but also a reliable estimate of the associated uncertainties. The most recent attempts to approach the problem of uncertainty estimation used normalising flows (Kodi Ramanah, Wojtak, Ansari, et al., 2020) to infer the conditional probability distribution of the dynamical cluster masses and approximate Bayesian inference to assign prior distributions to the neural network weights (Ho et al., 2020). Despite the increasing popularity of ML-based methods, the classical techniques of cluster mass estimation still currently prevail over the applications to observational data. Nevertheless, they are likely to be superseded by their ML counterparts for future applications involving next-generation surveys.

The primary challenge intrinsic to galaxy cluster mass estimation is posed by interlopers. These are galaxies that are not gravitationally bound to the cluster, but that are located along the line of sight and have similar line-of-sight velocities to the cluster. Distinguishing interlopers from member galaxies is a problematic task, because redshift surveys can only provide information about the positions and velocities of objects along the line of sight, and not perpendicular to it. Finding an effective way of reducing contamination from interlopers, with the limited information available from surveys, is essential to improve the accuracy of galaxy cluster mass estimates.

In this work, we propose to work at the level of 3D projected phase-space distribution, characterised by the sky projected galaxy positions and their line-of-sight velocities, instead of the standard 2D phase-space, to alleviate the interloper contamination and improve the precision of cluster mass estimation, as motivated by the following arguments. Cluster members are distributed more symmetrically around the cluster centre, while interlopers can clump in any place. Moreover, 2D phase-space density is averaged over the position angle, such that the information on any axially asymmetric localisation of interlopers is lost, rendering it more difficult for the algorithm to differentiate between interlopers and cluster members. In contrast, 3D phase-space density retrieves the information encoded in the position angle and is, therefore, expected to provide a better separation between cluster members and interlopers. Moreover, dynamical substructures have been shown to result in an artificial overestimation of cluster masses (Old et al., 2018; Tucker et al., 2018). These substructures may also induce an asymmetry within the boundary of dark matter halos, such that the 3D phase-space density will more adequately account for the presence of substructures and help to mitigate this bias. To optimise the information from the 3D dynamical phase-space distribution, we make use of 3D convolutional kernels, naturally designed to extract spatial features, in neural networks. Compared to the previous studies, we also adopt larger apertures than the virial sphere. This allows us to include the cluster infall zone as an extra constraining power in the estimation of dynamical masses. The observed infall patterns around galaxy clusters have long been used to measure cluster mass profiles at large distances (Diaferio, 1999; Diaferio & Geller, 1997; M. Falco et al., 2014; Rines et al., 2003).

We opt for a simulation-based inference approach to quantify the uncertainties on the neural network predictions. Simulation-based inference (e.g. Cranmer et al., 2019, and references therein), often referred to as *likelihood-free inference*, encompasses a class of statistical inference methods where simulations are used to estimate the posterior distributions of the parameters of interest conditional on data, without any prior knowledge or assumption of the likelihood distribution. Simulation-based inference has emerged as a viable alternative to perform Bayesian inference under complex generative physical models using only simulations. This framework allows all physical effects encoded in forward simulations to be properly accounted for in the inference pipeline, without having recourse to inadequate or misguided likelihood assumptions. As such, simulation-based inference, and variants thereof, have recently garnered significant interest for cosmological data analysis (e.g. Akeret et al., 2015; Alsing et al., 2019; Alsing &

Wandelt, 2019; Alsing et al., 2018; Charnock et al., 2018; Jennings & Madigan, 2017; Leclercq, 2018; Lintusaari et al., 2017; Wang et al., 2020).

In essence, we present a simulation-based inference framework for the estimation of the dynamical mass of galaxy clusters with 3D convolutional neural networks. The approach presented here is complementary to our previous neural flow (NF) mass estimator (Kodi Ramanah, Wojtak, Ansari, et al., 2020, hereafter NF2020) in various aspects. This is primarily a conceptually different framework of uncertainty estimation using neural networks. The simulation-based inference machinery, as presented here, allows the inference of the approximate posterior distribution of cluster masses given their 3D projected phase-space distribution, using an ensemble of simulated clusters and a neural network designed to extract summary statistics. In contrast, the NF mass estimator is a neural density estimator, where the cluster mass inference problem is formulated within a conditional density estimation framework. The two methods also differ in network architecture and dimensionality of their respective inputs. The approach presented here employs 3D convolutional kernels to fully exploit the information encoded in the 3D phase-space distribution of galaxy clusters, while the NF mass estimator relies on fully connected layers, i.e. multilayer perceptrons, and works at the level of the compressed 2D phase-space dynamics.

The remainder of this chapter is organised as follows. Section 7.2 provides an overview of the 3D dynamical phase-space distribution in terms of the key observables used for training the neural network. We also outline the mock generation procedure for cluster catalogues emulating the features of the actual SDSS data set and the preprocessing steps involved in the preparation of the training and test sets. We then describe the simulation-based inference approach utilised in this work, the network architecture and the training procedure in Section 7.3. We subsequently validate and demonstrate the performance of the optimised model on the test cluster catalogue in Section 7.4 and follow up by inferring cluster masses from the actual SDSS catalogue in Section 7.5. Finally, we provide a summary of the main aspects and findings of our work in Section 7.6, and highlight potential future investigations with cosmological applications.

7.2 DYNAMICAL PHASE-SPACE DISTRIBUTION

We outline the general problem of cluster mass estimation by first introducing the dynamical phase-space distribution. We then describe the generation of the mock SDSS catalogue which will be used to train and evaluate the performance of the neural network in future sections.

7.2.1 Galaxy cluster observables

The definition of cluster's halo mass adopted throughout this work is M_{200c}, corresponding to the mass contained in a sphere with mean density equal to 200 times the critical density of the Universe at the halo's redshift. We obtain an estimate of the mass by employing the full projected phase-space distribution of galaxy clusters. This consists of the positions of each member galaxy projected onto the (x, y) plane of the sky, denoted as $(x_{\text{proj}}, y_{\text{proj}})$, as well as their

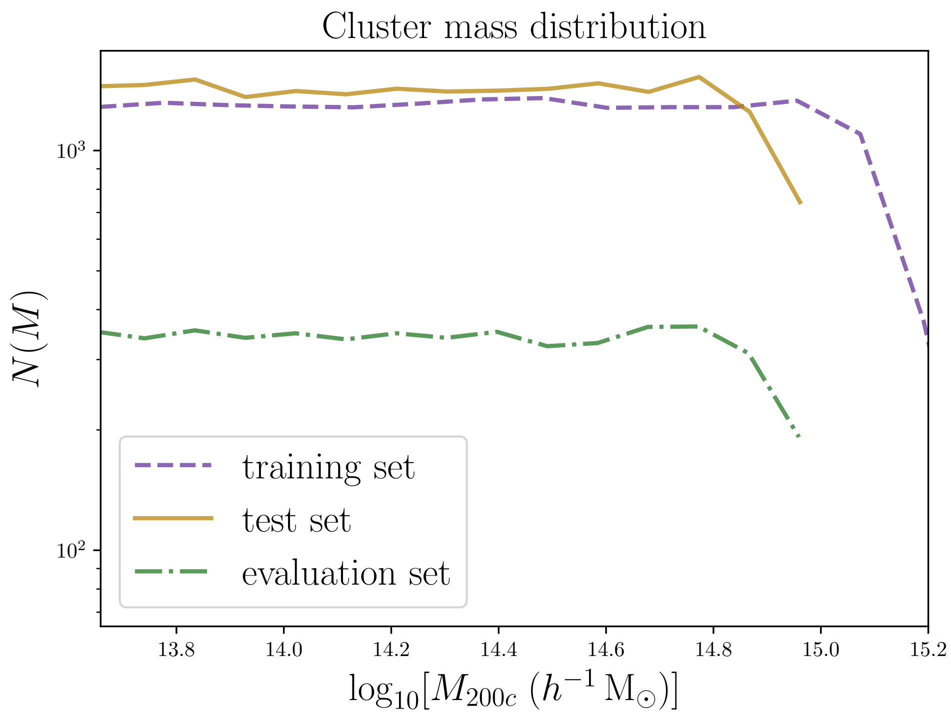


FIGURE 7.1 – Cluster mass distribution, i.e. variation of number of clusters with logarithmic mass for the training, test and evaluation sets extracted from the mock SDSS catalogue. To ensure we do not induce any cosmological information or selection bias while training the neural network, we upsample the relatively scarce high-mass clusters using independent lines of sight, thereby resulting in an approximately flat mass distribution for the training set.

separate line-of-sight velocities, v_{los}, as provided by redshift surveys. In this work, instead of computing the projected radial distance from the cluster centre as $R_{\mathrm{proj}} = (x_{\mathrm{proj}}^2 + y_{\mathrm{proj}}^2)^{1/2}$ as is typically done, we exploit the information from x_{proj} and y_{proj} separately. This should make our model more sensitive to interlopers and substructures, which are often located asymmetrically around the cluster centre. We adopt units of h^{-1}Mpc for x_{proj} and y_{proj} throughout this work.

7.2.2 Mock cluster catalogues

We generate mock observations of galaxy clusters using publicly available galaxy catalogues derived from the MultiDark simulations (Klypin et al., 2016).[1] Among the three different semi-analytic galaxy formation models applied to the simulation (Knebe et al., 2018), we opted for Semi-Analytic Galaxies (SAG), which includes the most complete implementation of modelling orphan galaxies and, therefore, produces the most realistic distribution of galaxies in the cluster cores (Cora, 2006; Cora et al., 2018). For more details regarding implementations of the star formation and feedback processes in SAG as well as a comparison to the remaining two semi-

[1] http://skiesanduniverses.org

analytic models, i.e. GALACTICUS (Benson, 2012) and the Semi-Analytic Galaxy Evolution (SAGE) model (Croton et al., 2006), we refer the interested reader to Knebe et al. (2018). The galaxy catalogues from SAG contain the positions and absolute magnitudes in the SDSS filters at all snapshots of the simulation. The background dark matter simulation (MDPL2) was run for the Planck ΛCDM cosmological model (Planck Collaboration et al., 2014). The simulation box has a size of $1\,h^{-1}$ Gpc and a mass resolution of $1.51 \times 10^{9}\,h^{-1}\,\mathrm{M}_{\odot}$.

We select galaxy clusters as massive dark matter halos found in the halo catalogues produced by the ROCKSTAR halo finder (Behroozi et al., 2013). For every halo, we construct its cluster's projected phase-space diagram by drawing a line of sight and computing the corresponding projections of the galaxy positions and velocities onto the plane of the sky and the line of sight, respectively. All phase-space coordinates are calculated relative to the central galaxy assigned to the main cluster halo and the observed velocities include the Hubble flow with respect to the cluster centre. The final projected phase-space diagrams are generated by applying the following cuts: ± 2200 km s^{-1} in line-of-sight velocities v_{los} and $\pm 4\,h^{-1}$Mpc in proper distances x_{proj} and y_{proj}.

Aiming at generating mock data which resemble the main spectroscopic galaxy sample of the SDSS Legacy Survey (Strauss et al., 2002), we adopt a flux limit of 18.0 magnitude in r-band. The flux limit is 0.2 magnitude lower than the actual SDSS limiting magnitude in order to compensate the slightly lower counts of galaxies in simulated clusters than in the SDSS ones (see Knebe et al., 2018). The apparent magnitudes of all galaxies in the field of each galaxy cluster are computed by assigning each simulated cluster an observer located at comoving distance randomly drawn from a uniform distribution within a 3D ball. The maximum comoving distance is $250\,h^{-1}$ Mpc, for which galaxy cluster detection in the SDSS main galaxy sample is complete down to a cluster mass of $\sim 10^{14.0} h^{-1}\mathrm{M}_{\odot}$ (Abdullah, Klypin, et al., 2020).

Our mock SDSS galaxy catalogue is generated assuming completeness of spectroscopic observations down to the assumed flux limit. This is an idealised assumption because the actual completeness of the SDSS decreases in high-density regions due to the physical limit on the minimum distance between SDSS fibres. However, since our CNN mass estimator operates on smoothed kernel density estimator density maps, we expect that downsampling due to incompleteness of SDSS spectroscopic observations should not have a noticeable impact on the final mass estimates. The insensitivity of CNN mass estimators based on smoothed density maps to stochastic downsampling was shown in Ho et al. (2019) and Kodi Ramanah, Wojtak, Ansari, et al. (2020). This test can be repeated for a density-dependent incompleteness resembling the SDSS selection for spectroscopic observations. Considering an extreme case when cluster data are missing spectroscopic velocities inside cluster cores subtending 55 arc secs, which is the minimum distance between SDSS fibres (Strauss et al., 2002), we find that, for a sample of SDSS-like clusters, our CNN trained on complete mocks yields mass estimates only 0.017 dex lower in average. Since the SDSS is more complete than this extreme example, primarily by virtue of an optimised tiling, we conclude that a realistic bias is even smaller and currently negligible compared to the precision of our mass estimator.

Keeping in mind possible future applications of our dynamical mass estimator for cosmological inference with the cluster abundance, it is instructive to generate a galaxy cluster sample for which the distribution of cluster masses is independent of cosmological model through the mass function. An optimal solution is to consider a set of clusters with a flat distribution in log mass within a possibly wide range of dynamical cluster masses. Aiming at generating a sample with $\sim 10^4$ galaxy clusters, we downsample the actual mass function below halo mass $M_{200c} \approx 10^{14.3} h^{-1} \mathrm{M}_\odot$ and generate up to 25 projections per cluster at higher masses. In order to minimise correlations between projected phase-space diagrams derived from the same cluster, we use a set of directions (up to 25 lines of sight) found by maximising angular separations between every two closest sight lines. The adopted maximum number of sight lines per cluster is not sufficient to keep a flat distribution at the high-mass end, i.e. $\log_{10} M_{200c} \gtrsim 14.9$ (cf. Figure 7.1). This, however, can hardly be improved because further increase of upsampling would introduce strong correlations between phase-space diagrams generated from the same galaxy cluster. The final sample contains 4.3×10^4 galaxy clusters with a minimum halo mass of $10^{13.7} h^{-1} \mathrm{M}_\odot$.

The overdensity threshold used in the halo mass definition depends on redshift. This leads to a well-known non-physical evolution of halo masses which reflects merely the redshift dependence of the critical density (Diemer et al., 2013). Since phase-space diagrams do not provide any information on cluster redshifts required to adjust the overdensity threshold, mass estimates from neural networks may be consequently affected by an additional noise. For a wide redshift range, the noise may be sufficiently large so that it would be necessary to supplement each phase-space diagram with the information on cluster redshift setting the corresponding overdensity threshold. However, for our mock data spanning a relatively narrow redshift range $z \leq 0.085$, the expected uncertainty due to the lack of information on cluster redshifts amounts to only 0.006 dex, which is significantly lower than the level of precision obtained in our work and similar studies, i.e. ~ 0.1 dex.

We extract the training set, with a flat mass distribution, containing around seventeen thousand clusters by randomly drawing from the mock catalogue. The corresponding validation set, used for early stopping when optimising the neural network, is designated as 10% of the training set, such that it contains ~ 1700 clusters and only ~ 15500 clusters are utilised during training. The test set consisting of twenty thousand clusters is obtained by randomly sampling from the remaining clusters in the mock catalogue. The remaining ~ 5000 clusters in the catalogue then constitute an evaluation set. The test set is used in the simulation-based inference framework (cf. Section 2.3.2), whilst the purpose of the evaluation set is to assess the performance of the network (cf. Section 7.4.2). The mass distributions of the non-overlapping training, test and evaluation sets are depicted in Figure 7.1.

7.2.3 Kernel density estimator

Before the observables (v_{los}, x_{proj}, y_{proj}) are provided as input to the ML model, they are first preprocessed with a kernel density estimator (KDE) to create a smooth PDF mapping in the 3D phase-space distribution. This is done in order to obtain similar-sized arrays as inputs for all clusters, which contain different numbers of member galaxies, as well as to create a visual input (image) for the convolutional neural network. An in-depth review of kernel density estimation is provided in Diggle and Gratton (1984), Sheather (2004), and Wand and Jones (1994).

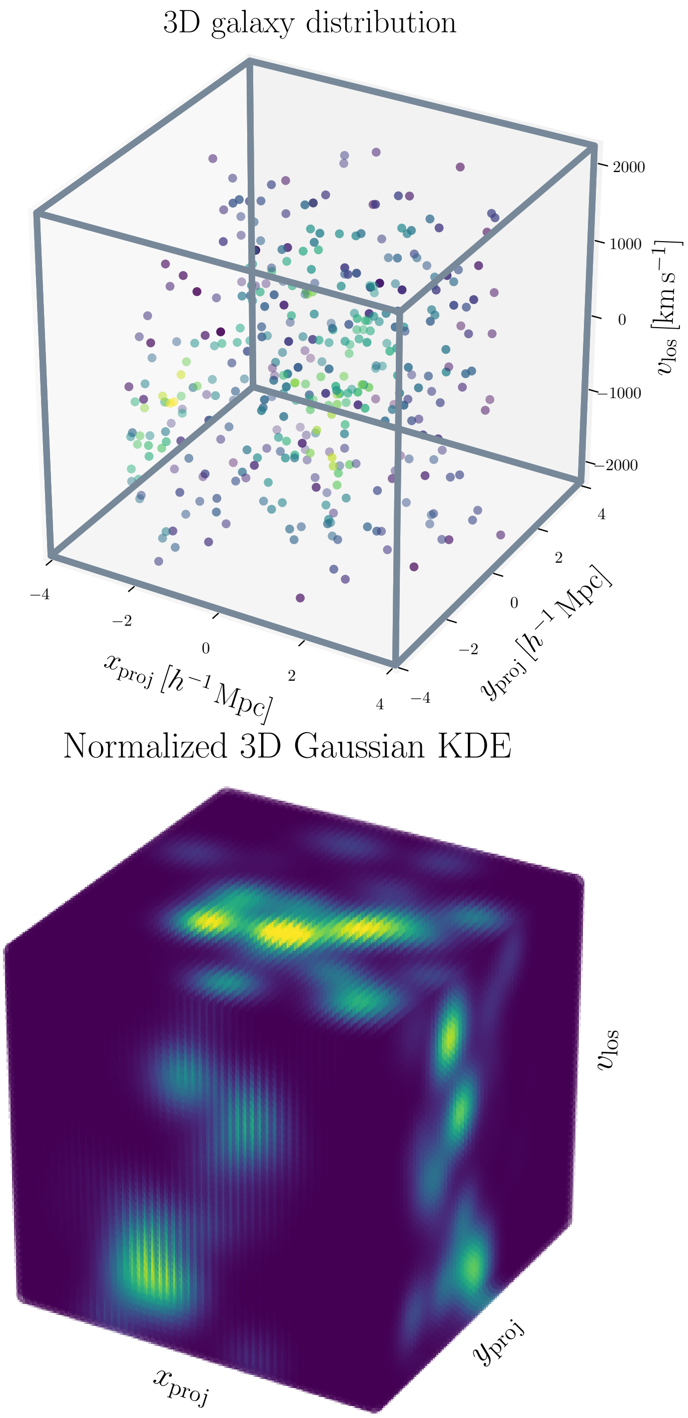


Figure 7.2 – Simulated 3D phase-space distribution of galaxies observed in an example cluster of galaxies, represented via its 3D galaxy distribution (*top panel*) and 3D Gaussian KDE (*bottom panel*), consisting of projected galaxy positions in the sky, $x_{\rm proj}$ and $y_{\rm proj}$, and line-of-sight velocities, $v_{\rm los}$, in dynamical phase space. The KDE representation serves as inputs to our 3D convolutional neural network.

Let our complete set of n observables $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$, in which each variable $\boldsymbol{X}_i$ is given by $(v_{\text{los}}, x_{\text{proj}}, y_{\text{proj}})$, be drawn from an unknown distribution with density f. The density evaluated at a point $\boldsymbol{x} = (v_{\text{los}}, x_{\text{proj}}, y_{\text{proj}})$ can be approximated as

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n|\mathbf{H}|^{1/2}} \sum_{i=1}^{n} K\Big[\mathbf{H}^{-1/2}(\boldsymbol{x} - \boldsymbol{X}_i)\Big], \tag{7.1}$$

where K is the kernel function and $\mathbf{H}$ is a 3×3 bandwidth matrix. The KDE sums up the density contributions from the collection of data points $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ at the evaluation point $\boldsymbol{x}$. Data points close to $\boldsymbol{x}$ contribute significantly to the total density, while data points further away from $\boldsymbol{x}$ have only a relatively small contribution. The shape of the density contributions is determined by the kernel function, and their size and orientation are dictated by the bandwidth matrix. In this work, we use a 3-dimensional Gaussian kernel given by

$$K(\boldsymbol{u}) = (2\pi)^{-3/2} \, |\mathbf{H}|^{-1/2} \, \exp\Big(-\tfrac{1}{2} \, \boldsymbol{u}^{\top} \, \mathbf{H}^{-1} \, \boldsymbol{u}\Big), \tag{7.2}$$

with $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{X}_i$. For the bandwidth matrix, a scaling factor κ is multiplied with the covariance matrix of the data, $\mathbf{H} = \kappa\Sigma$. The scaling factor should be sufficiently small to encapsulate even the more subtle features of the data and small-scale signal expected for low-mass clusters, but large enough that the ML model is robust to changes in galaxy number count, and can easily interpolate between the data sets of discrete and quite scarcely distributed points. We performed some numerical experiments with three distinct scaling factors, $\kappa = \{0.15, 0.175, 0.20\}$, and found only a marginal influence on the model predictions. We, therefore, opted for the intermediate value of $\kappa = 0.175$ in our study.

An example of a 3D KDE representation that serves as input to the neural network for one particular cluster is illustrated in Figure 7.2. For all clusters considered in this work, unless otherwise stated, the extents of the observables are as follows: $v_{\text{los}} \in [-2200, 2200]\,\text{km}\,\text{s}^{-1}$, $x_{\text{proj}} \in [-4.0, 4.0]\,h^{-1}$Mpc and $y_{\text{proj}} \in [-4.0, 4.0]\,h^{-1}$Mpc, with 50 voxels along each axis, resulting in 3D slices of dimension 50^3. Concerning the choice of the maximum extent for x_{proj} and y_{proj}, we performed an optimisation procedure for different sizes of $\{1.6, 4.0, 6.0\}\,h^{-1}$Mpc and opted for $4.0\,h^{-1}$Mpc, which resulted in a noticeable improvement in precision of our mass estimator relative to a size of $1.6\,h^{-1}$Mpc and virtually no loss of constraining power relative to a size of $6.0\,h^{-1}$Mpc.

7.3 SIMULATION-BASED INFERENCE WITH NEURAL NETWORKS

In order to optimally extract information from the 3D phase-space distribution, we employ a 3D CNN (as described in more detail in Section 2.2.1), within a simulation-based inference (SBI) framework (cf. Section 2.3.2) to robustly quantify the uncertainties on the model predictions.

For the particular mass inference problem studied here, the SBI approach entails the generation of an ensemble of galaxy clusters using a physical model or simulator. We employ

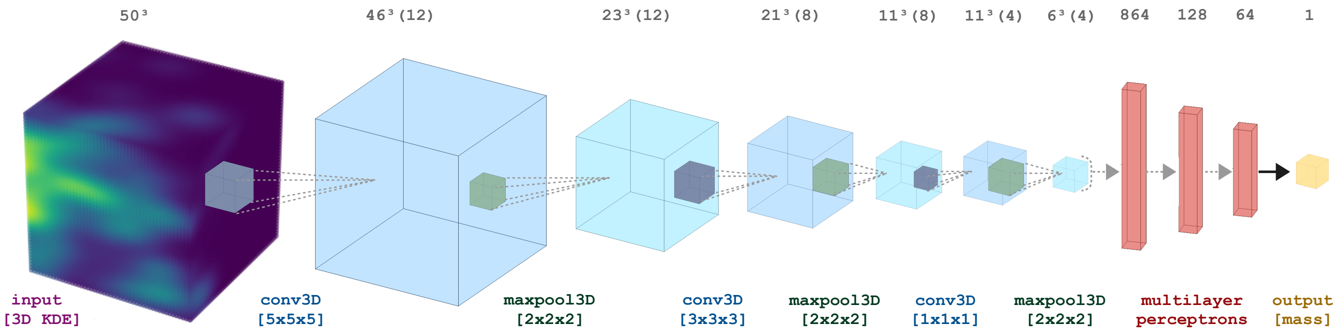


FIGURE 7.3 – Schematic representation of our $\mathrm{CNN_{3D}}$ architecture to predict cluster masses from their 3D dynamical phase-space distributions. The dimensions of the input 3D slice and those of the subsequent slices, resulting from the convolutional and maxpooling operations, are indicated in the top row, with the number of feature maps per layer given in parentheses. The respective kernel sizes of the latter operations are given in the bottom row, with single strides employed and without use of padding. The CNN extracts the informative spatial features from the 3D phase-space distribution and gradually compresses the high-dimensional space to a single scalar which corresponds to the dynamical cluster mass.

some physical priors, in terms of a flat distribution in dynamical mass (as motivated by decoupling from the cosmological model imprinted in the halo mass function) and uniform spatial distribution, in the cluster generation procedure (cf. Section 7.2.2). Given that our ultimate objective is the inference of cluster masses from the SDSS catalogue, we generate a realistic set of SDSS-like clusters. By feeding the 3D phase-space distributions of this set of generated clusters to our trained neural network, $\mathbb{NN}(\boldsymbol{\omega},\boldsymbol{\eta}) : \boldsymbol{D} \to \tilde{d}$, we obtain a corresponding set of predicted summaries $\tilde{d}$, which, by design, correspond to the cluster masses. This allows us to characterise the joint probability distribution of data (via the compressed summaries) and parameters, $\mathcal{P}(\tilde{d}, M)$, via a kernel (or neural) density estimator. In this work, we make use of a Gaussian KDE (cf. Figure 7.4 in Section 7.4) as our density estimator. By slicing this joint distribution at any observed data fed to the network, $\mathbb{NN}(\boldsymbol{\omega},\boldsymbol{\eta}) : \boldsymbol{D}_{\mathrm{obs}} \to \tilde{d}_{\mathrm{obs}}$, we obtain the approximate posterior as follows:

$$\mathcal{P}(M|\tilde{d}_{\mathrm{obs}}) \approx \mathcal{P}(M|\boldsymbol{D}_{\mathrm{obs}}, \boldsymbol{\omega}, \boldsymbol{\eta}). \tag{7.3}$$

7.3.1 Neural network architecture

The underlying objective of our SBI framework is to infer the posterior of the dynamical mass of a galaxy cluster, given its 3D phase-space distribution characterised by the projected sky positions and the line-of-sight velocities, i.e. $\mathcal{P}(M|\{\boldsymbol{x}_{\mathrm{proj}}, \boldsymbol{y}_{\mathrm{proj}}, \boldsymbol{v}_{\mathrm{los}}\})$. The neural network takes as input a 3D slice $\widetilde{\boldsymbol{\mathcal{D}}}$, which is a 3D array of the Gaussian KDE applied to the phase-space distribution, as described in Section 7.2, with an example illustrated in Figure 7.2. The training data set, therefore, consists of pairs of $\{M, \widetilde{\boldsymbol{\mathcal{D}}}\}$.

A schematic of our 3D CNN (hereafter $\mathrm{CNN_{3D}}$) architecture is depicted in Figure 7.3. The network extracts spatial features from the input 3D phase-space distribution by performing convolutions with a kernel of size $5 \times 5 \times 5$. We employ several such kernels in one layer to probe different aspects of the input 3D slice, yielding a set of feature maps which are subsequently fed to a maxpooling layer for the purpose of dimensionality reduction. We use a $2 \times 2 \times 2$ maxpooling kernel to reduce the slice size by a factor of two. We adopt single strides and no padding for both operations. By repeatedly alternating between these two types of layers, we can reduce the initial 3D distribution to a compact representation of features. At this point, the resulting 3D slice may be flattened to a vector, with this vectorised set of features fed to the final layers which consist of fully connected layers of neurons, i.e. multilayer perceptrons. Finally, the output layer yields the dynamical cluster mass, as desired. We encode ReLU (Nair & Hinton, 2010) activation functions (as defined in eq. (2.11)) in the convolutional layers, and linear activations in the final fully connected layers. We highlight the relatively low complexity of the network architecture with $\sim 10^5$ trainable weights.

7.3.2 Training methodology

We train our $\mathrm{CNN}_{3\mathrm{D}}$ model as a regression over the logarithmic cluster mass by minimising a mean squared error loss function with respect to the network weights. The model and training routine are implemented using the KERAS library (Chollet et al., 2015) via a TENSORFLOW backend (Abadi et al., 2016). We make use of the *Adam* (Kingma & Ba, 2014) optimizer, with a learning rate of $\eta = 10^{-4}$ and first and second moment exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, respectively. The batch size is set to 100. We train the neural network for around 50 epochs, requiring around 10 minutes on an NVIDIA V100 Tensor Core GPU. In order to prevent any overfitting, we adopt the standard regularisation technique of early stopping in our training routine. For this purpose, 25% of the original training data set is kept as a separate validation set, with both the training and validation losses monitored during training. We opt for an early stopping criterion of 5 epochs, such that training is halted when the validation loss no longer shows any improvement for 5 consecutive epochs, and the optimised weights of the previously saved best fit model are restored.

7.4 VALIDATION AND PERFORMANCE

7.4.1 Uncertainty estimation

WE NOW assess the performance of our optimised $\mathrm{CNN}_{3\mathrm{D}}$ model on the evaluation set. As part of the SBI procedure, we first compute the joint 2D probability density function (PDF), $\mathcal{P}(\tilde{d}, M)$, of the summary statistics $\tilde{d}$ extracted by the neural network and the parameters M obtained using the test set containing around twenty thousand clusters (cf. Section 7.2.2). Recall that the neural summary statistics in this case are, by design, taken to be point predictions of masses by the $\mathrm{CNN}_{3\mathrm{D}}$, while the parameters correspond to the ground truth masses. We make use of a bivariate Gaussian KDE, with a bandwidth scaling of $\kappa = 0.20$, to obtain the 2D PDF depicted in Figure 7.4. This involves the application of eqs. (7.1) and (7.2), where now $\boldsymbol{x} = (\tilde{d}, M)$ corresponds to a given evaluation point in this 2D parameter space and $\boldsymbol{X}_i$ describes the collection of pairs of $\{\tilde{d}, M\}$ data points, with $\mathbf{H}$ being a 2×2 bandwidth matrix. To infer the posterior PDFs of the dynamical masses of the clusters in the evaluation set, we first obtain

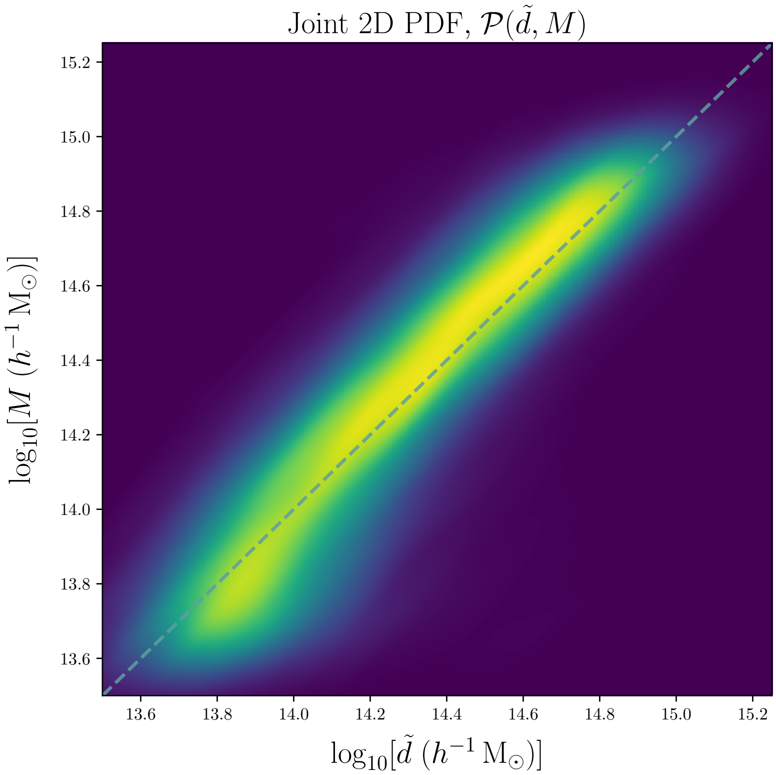


FIGURE 7.4 – Joint 2D PDF of network predicted summaries $\tilde{d}$ and parameters M, i.e. $\mathcal{P}(\tilde{d}, M)$, obtained using a bivariate Gaussian KDE with a bandwidth scaling factor of 0.20. Recall that $\tilde{d}$, by design, corresponds to the point dynamical mass estimates from our $\mathrm{CNN_{3D}}$ model. A test set consisting of twenty thousand clusters is used to compute this 2D PDF. This is representative of the prediction scatter with respect to the ground truth masses and is employed in our simulation-based inference framework to compute and assign uncertainties associated to point masses predicted by our $\mathrm{CNN_{3D}}$. We obtain the approximate posterior, $\mathcal{P}(M|\{\boldsymbol{x}_{\mathrm{proj}}, \boldsymbol{y}_{\mathrm{proj}}, \boldsymbol{v}_{\mathrm{los}}\}, \boldsymbol{\theta}, \boldsymbol{\alpha})$, given a set of network weights $\boldsymbol{\theta}$ and hyperparameters $\boldsymbol{\alpha}$, for a particular cluster by making a vertical slice at the neural network predicted value of $\tilde{d}$.

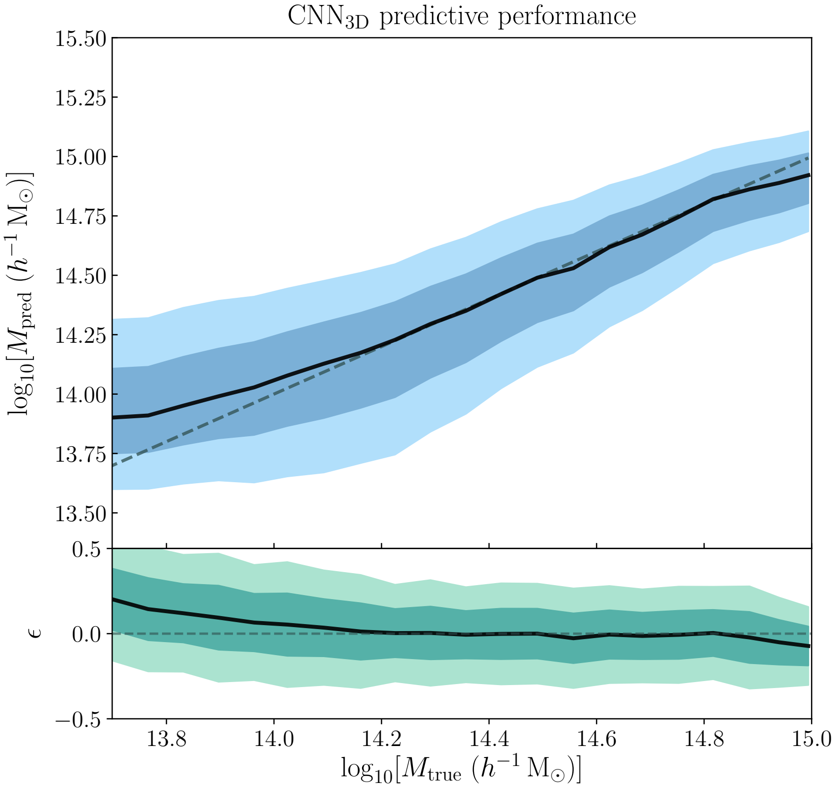


FIGURE 7.5 – Predictive performance of our CNN$_{3D}$ model. *Top panel:* CNN$_{3D}$ predictions against ground truth, depicting the mean prediction (solid line) and the predicted confidence intervals (shaded 1σ and 2σ regions) of the posterior probability density as a function of logarithmic bins of $M_{\rm true}$ for ~ 5000 galaxy clusters from the evaluation set. The simulation-based inference approach, as expected, yields larger uncertainties for low-mass clusters. *Bottom panel:* Distribution of residual scatter as a function of the logarithmic true cluster mass. The solid line corresponds to the mean logarithmic residual scatter, $\epsilon \equiv \log_{10}(M_{\rm true}/M_{\rm pred})$, if we consider only the maximum likelihood predictions (i.e. point mass estimates) from our CNN$_{3D}$, in logarithmic bins of $M_{\rm true}$. The shaded bands depict the log-normal scatter (1σ and 2σ regions) about the mean residuals. The CNN$_{3D}$ tends to overestimate masses of poor clusters below $\log[M_{\rm true}(h^{-1}{\rm M}_\odot)] \approx 14.0$ dex. The correspondingly larger uncertainties for clusters in this mass regime demonstrate the reliability of the simulation-based inference framework to provide uncertainties that are not underestimated.

the network point predictions using the trained $\mathrm{CNN_{3D}}$ model. We subsequently vertically slice the joint PDF from Figure 7.4 at the point estimates to infer the approximate posterior PDFs for the mass of each cluster. From the posteriors, we quantify the 1σ uncertainties by integrating the 68% probability volume, such that the upper and lower 1σ uncertainty limits may be asymmetrical.

7.4.2 Performance evaluation

Using the inferred posterior mass PDFs for the clusters in the evaluation set, we evaluate the performance of our $\mathrm{CNN_{3D}}$ on the realistic mock catalogue by plotting our model predictions against the ground truth masses of the ~ 5000 clusters from the evaluation set in the top panel of Figure 7.5. We bin the model predictions in logarithmic mass intervals with the mean prediction and confidence intervals (1σ and 2σ regions) of the posterior probability density depicted via the solid line and shaded regions, respectively. The top panel shows the efficacy of our $\mathrm{CNN_{3D}}$ model to recover the ground truth masses of the clusters from the evaluation set within the 1σ uncertainty limit. The bottom panel displays the distribution of residuals, $\epsilon \equiv \log_{10}(M_{\mathrm{true}}/M_{\mathrm{pred}})$, in the $\mathrm{CNN_{3D}}$ point predictions relative to the ground truth, as a function of the logarithmic cluster mass, with the solid line indicating the mean residual scatter and the shaded bands corresponding to the 1σ and 2σ regions. The $\mathrm{CNN_{3D}}$ predictions have a mean residual and log-normal scatter of $\langle\epsilon\rangle = 0.04$ dex and $\sigma_\epsilon = 0.16$ dex.

From the bottom panel of Figure 7.5, we observe the tendency of the $\mathrm{CNN_{3D}}$ to overestimate the masses for clusters with masses below $\log[M_{\mathrm{true}}(h^{-1}\mathrm{M}_\odot)] \approx 14.0$ dex. This may primarily be attributed to the realistic effects included in our mock catalogue as detailed in Section 7.4.3 below. Nevertheless, this relatively high residual scatter due to the overprediction of cluster mass is properly accounted for in the network predicted uncertainties, with the lower 1σ and 2σ limits being larger than the upper limits. This demonstrates the capacity of the SBI framework to yield reliable uncertainties that are not underestimated. Conversely, the network slightly underestimates the masses for the most massive clusters above $\log[M_{\mathrm{true}}(h^{-1}\mathrm{M}_\odot)] \approx 14.9$ dex. There is a two-fold plausible explanation for this effect. First, the selection cuts (cf. Section 7.2.2) to produce the 3D phase-space diagrams may not be sufficiently large to capture all the galaxy members of the massive clusters, resulting in incomplete cluster samples. The second explanation is related to possible mean-reversion edge effects, as also reported by Ho et al. (2020) and Ho et al. (2019), Ntampaka et al. (2016), whereby the model predictions of cluster masses at the edge of the mass range considered here are biased towards the average. In general, this systematic bias is related to a neural network's tendency to be more adept at interpolation than extrapolation. To mitigate such biases, we would require more training clusters beyond the edges of the mass regime of the training set, i.e. $\log[M_{\mathrm{true}}(h^{-1}\mathrm{M}_\odot)] < 13.7$ dex and $\log[M_{\mathrm{true}}(h^{-1}\mathrm{M}_\odot)] > 15.0$ dex. Note that this mean-reversion effect may also be partially responsible for the overprediction of cluster masses in the low-mass regime.

For the validation of posterior recovery within the SBI framework, we perform a similar statistical test to recent ML studies (e.g. Ho et al., 2020; Perreault Levasseur et al., 2017; Wagner-Carena et al., 2021). This test is based on model calibration, whereby a posterior containing x% of the probability volume should contain the ground truth within this specific volume x% of

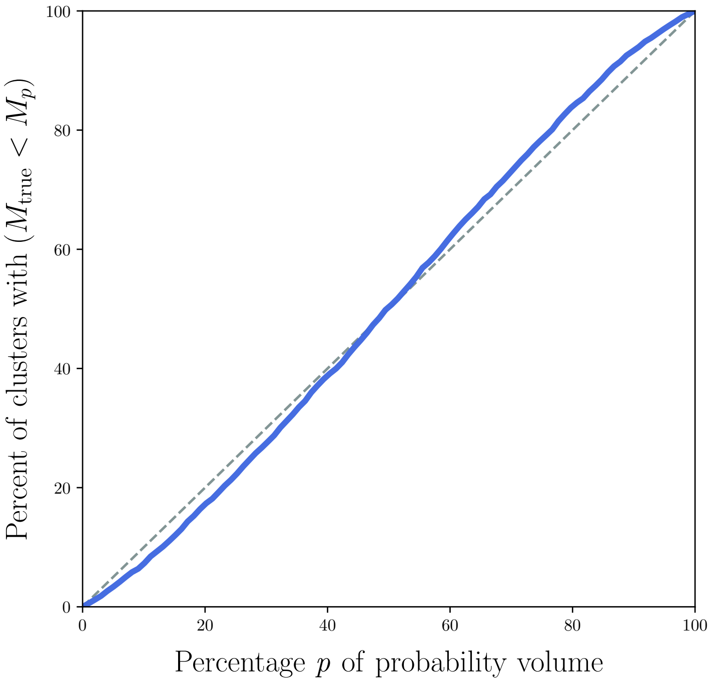


FIGURE 7.6 – Validation of posterior recovery within the SBI framework via a posterior calibration (or quantile-quantile) plot. The dashed line indicates perfect or ideal calibration, with the solid blue line depicting the calibration of the inferred posterior PDFs. The latter illustrates the overall statistical consistency of the inferred posteriors, displaying near ideal calibration, thereby avoiding both under and overconfidence.

the time. As such, for our given problem with one-dimensional mass posteriors, this test boils down to the computation of coverage probabilities, defined as the fraction of test samples where the ground truth lies within a particular confidence interval. Note that this posterior validation test does not assume Gaussian nature of PDFs and holds for any arbitrary PDFs. A clear and intuitive description of this test on a toy model is provided in the appendix of Wagner-Carena et al. (2021). The posterior calibration plot for the ~ 5000 clusters in the evaluation set, averaged over all clusters, is depicted in Figure 7.6. The diagonal dashed line implies ideal calibration, as a result of a perfect match between the number of test samples and the percentage of probability volume. The calibration plot of the inferred posterior PDFs is depicted via the solid blue line, illustrating the near ideal calibration and overall statistical consistency of the inferred posteriors. The absence of any significant deviations from the ideal calibration reference line implies that the SBI framework does not exhibit any particular strong under or overconfidence in assigning uncertainties to the CNN_{3D} model predictions.

7.4.3 Visualisation of interloper contamination

Our mock catalogue contains a realistic level of contamination by interloper galaxies, as expected from the actual SDSS observations. In this section, we explicitly highlight how the presence of these spurious galaxies renders the mass estimation extremely challenging.

The interloper contamination principally induces a bias (overestimation) in the neural network predictions which is more significant for the low-mass clusters, substantiating the relatively larger residual scatter for clusters with masses smaller than ~ 14.0 dex, as depicted in the bottom panel of Figure 7.5. This outcome is caused by several low-mass clusters, for which the $\mathrm{CNN_{3D}}$ systematically and significantly overpredicts the mass. To illustrate that interlopers constitute the underlying cause of these inaccurate mass estimates, we compute the contamination per cluster as the mass ratio of interloper clusters to the original cluster. A cluster is considered to be an interloper cluster when it is more massive than the original cluster, is located within a distance of $R_{\mathrm{proj}} = (x_{\mathrm{proj}}^2 + y_{\mathrm{proj}}^2)^{1/2} = 4\,h^{-1}$ Mpc and has $\Delta v_{\mathrm{los}} < 2200$ km s^{-1}. An additional factor that exacerbates this problem and renders the task of the $\mathrm{CNN_{3D}}$ more convoluted is the distance between the interloper and original clusters in 3D phase space. The closer the two clusters are together, the more difficult it is to tell them apart. The relative phase-space distance between the two clusters, denoted by c_1 and c_2, respectively, is computed as

$$d(c_1, c_2) = \sqrt{\left(\frac{\Delta x_{\mathrm{proj}}}{R_{200c,\mathrm{av}}}\right)^2 + \left(\frac{\Delta y_{\mathrm{proj}}}{R_{200c,\mathrm{av}}}\right)^2 + \left(\frac{\Delta v_{\mathrm{los}}}{v_{200c,\mathrm{av}}}\right)^2}, \tag{7.4}$$
$$\text{with } R_{200c} = \left(\frac{3M_{200c}}{4\pi 200\rho_c}\right)^{1/3} \text{ and } v_{200c} = \sqrt{\frac{GM_{200c}}{R_{200c}}},$$

where $R_{200c,\mathrm{av}} = \frac{1}{2}(R_{200c}^{c_1} + R_{200c}^{c_2})$ and $\Delta x_{\mathrm{proj}} = x_{\mathrm{proj}}^{c_1} - x_{\mathrm{proj}}^{c_2}$, with Δy_{proj}, Δv_{los} and $v_{200c,\mathrm{av}}$ analogously defined.

Figure 7.7 provides a stark illustration of the overwhelming interloper contamination inherent to the individual clusters from the test set, with the colour bar corresponding to the degree of interloper contamination and the marker size corresponding to the inverse of the distance. As can be seen, the clusters with the largest overestimation of their dynamical masses are also the ones whose phase-space diagrams are highly contaminated by interlopers in the form of independent clusters. For some of these clusters, the interloper cluster is around 20 times more massive than the original cluster, which renders the mass estimation extremely challenging. Observational data, such as the SDSS catalogue used in this work, would be similarly plagued by interloper contamination. While the performance of our $\mathrm{CNN_{3D}}$ model with a mean residual and log-normal scatter of $\langle\epsilon\rangle = 0.04$ dex and $\sigma_\epsilon = 0.16$ dex is not as impressive as the recent ML techniques at first glance, this is purely due to the more realistic mock catalogue employed here.

7.4.4 Information gain with higher dimensionality

In an attempt to illustrate the gain in information by exploiting the full 3D phase-space distribution, we also train our $\mathrm{CNN_{3D}}$ on the mock catalogue from Ho et al. (2019) and compare the performance of our network to their 1D and 2D counterparts in terms of the logarithmic residual scatter in Figure 7.8. $\mathrm{CNN_{1D}}$ infers cluster masses solely from the univariate distribution of line-of-sight velocities, i.e. $\{v_{\mathrm{los}}\}$, while $\mathrm{CNN_{2D}}$ additionally takes as input the

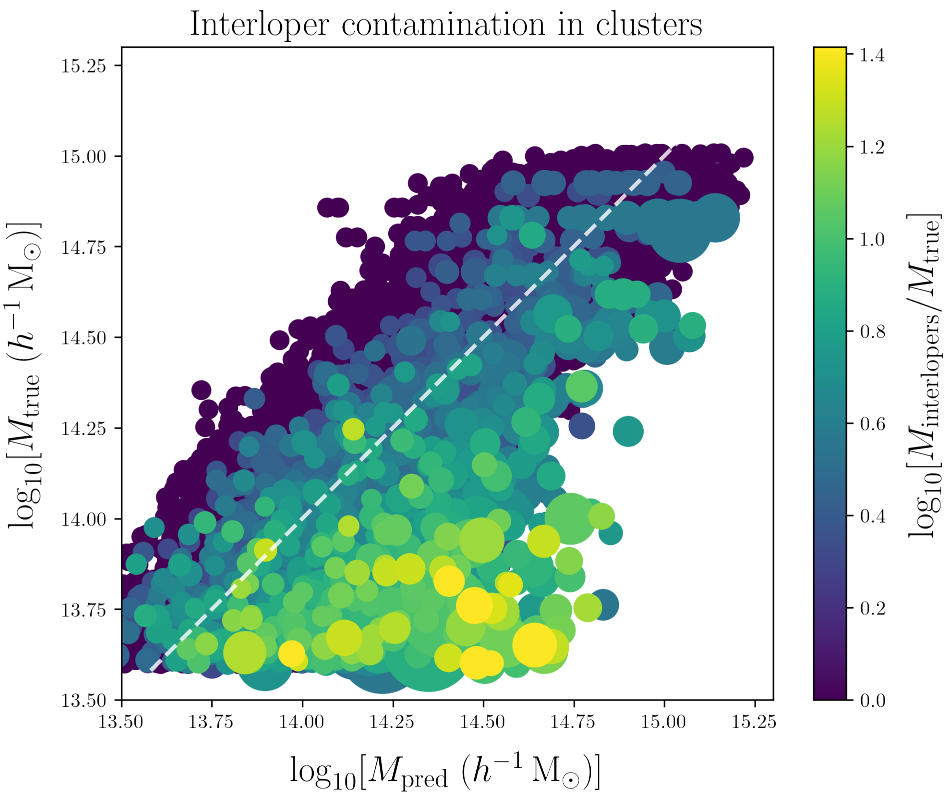


FIGURE 7.7 – Effect of clustered interlopers on the $\mathrm{CNN_{3D}}$ mass predictions. Each individual mass measurement is coloured according to the relative mass of interloper clusters contaminating the phase-space diagram of the main cluster. In addition, the size of the markers indicates the inverse distance between the interloper and original clusters in the projected phase space, as defined by eq. (7.4). Interlopers residing in relatively massive clusters which overlap closely with the original cluster in the projected phase space can hardly be distinguished from the original cluster members, giving rise to a substantial mass overestimation.

sky-projected radial positions given by $R_{\mathrm{proj}} = (x^2_{\mathrm{proj}} + y^2_{\mathrm{proj}})^{1/2}$, such that it relies on the joint distribution of $\{R_{\mathrm{proj}}, v_{\mathrm{los}}\}$. For the sake of comparison, we use similar phase-space cuts to Ho et al. (2019) for our $\mathrm{CNN_{3D}}$ model, i.e. $v_{\mathrm{los}} \in [-2200, 2200]\,\mathrm{km\,s^{-1}}$, $x_{\mathrm{proj}} \in [-1.6, 1.6]\,h^{-1}$Mpc and $y_{\mathrm{proj}} \in [-1.6, 1.6]\,h^{-1}$Mpc. Note that only the point (maximum *a posteriori*) estimates from our approach are used to produce the comparison plot displayed in Figure 7.8. As expected, we find that the precision of the mass estimator, as indicated by the log-normal residual scatter (shaded 1σ and 2σ regions), improves progressively with further information, thereby justifying the development and application of our $\mathrm{CNN_{3D}}$ model in this work.

As in our previous work (NF2020), we quantify the precision of cluster mass estimation in terms of the total scatter about the best-fit power-law relation between the ground truth and predicted cluster masses. Adopting the approach employed in Wojtak et al. (2018), we express the total scatter σ into a richness-dependent component given by σ_N and a richness-independent

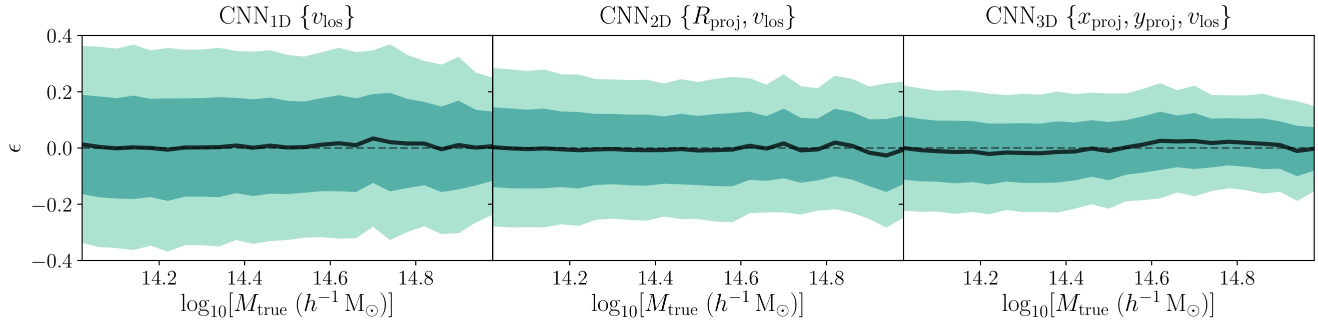


FIGURE 7.8 – Comparison of the performance of three CNN models with distinct input dimensionality. Performance is quantified in terms of the residual scatter, $\epsilon \equiv \log_{10}(M_{\mathrm{true}}/M_{\mathrm{pred}})$, in the CNN point predictions relative to the ground truth. The mean residual scatter is depicted via solid dark lines, with the shaded bands corresponding to the log-normal scatter (1σ and 2σ regions). The CNNs are trained with progressively larger dimensionality of the phase-space distribution of the same mock cluster catalogue from Ho et al. (2019). In all cases, the adopted maximum projected distance from the cluster centre is $1.6\,h^{-1}$Mpc. The results illustrate the gain in constraining power when the information content of the full 3D phase-space distribution of galaxies is exploited instead of relying merely on the velocity dispersion as in the 1D case displayed in the left panel. The $\mathrm{CNN_{1D}}$ and $\mathrm{CNN_{2D}}$ results are reproduced from Ho et al. (2019).

part denoted by σ_0, as follows:

$$\sigma^2 = \sigma_N^2 (N_{\mathrm{mem}}/100)^{-1} + \sigma_0^2, \tag{7.5}$$

where N_{mem} indicates the number galaxies within R_{200c} of the cluster's host dark matter halo. We determine the values of σ_N and σ_0 by fitting the above equation to the logarithmic residuals in the cluster mass predictions for the test set. The best fit model recovers the measured scatter with a fully satisfactory precision of 5 per cent. We carry out this procedure for the three CNN models and also include the results of the neural flow mass estimator (NF2020). Note that the same mock cluster catalogue from Ho et al. (2019) was used for all the methods. The recent ML techniques, illustrated in Figure 7.9, all outperform the traditional cluster mass estimators (cf. Figure 6 in NF2020) extensively tested in the Galaxy Cluster Mass Comparison Project (Old et al., 2015), which are not shown for the sake of clarity. For our $\mathrm{CNN_{3D}}$ model, we find $\sigma_N = 0.04$ dex and $\sigma_0 = 0.08$ dex, with the richness-dependent error smaller by a factor of two relative to $3/(\sqrt{2N_{\mathrm{mem}}}\ln 10) = 0.09$ dex expected for the mass estimation based solely on the scaling relation with the velocity dispersion. As expected from Figure 7.8, Figure 7.9 shows a progressive improvement in precision going from $\mathrm{CNN_{1D}}$ to $\mathrm{CNN_{3D}}$ due to the gain in constraining power when exploiting the full information from the 3D phase-space distribution of galaxies rather than relying only on the velocity dispersion. In general, the CNN mass estimators are less sensitive to cluster richness than the neural flow model. Figures 7.8 and 7.9, therefore, present an adequate depiction of the network performance and a fair comparison with recent ML methods, demonstrating the precision of our $\mathrm{CNN_{3D}}$ model.

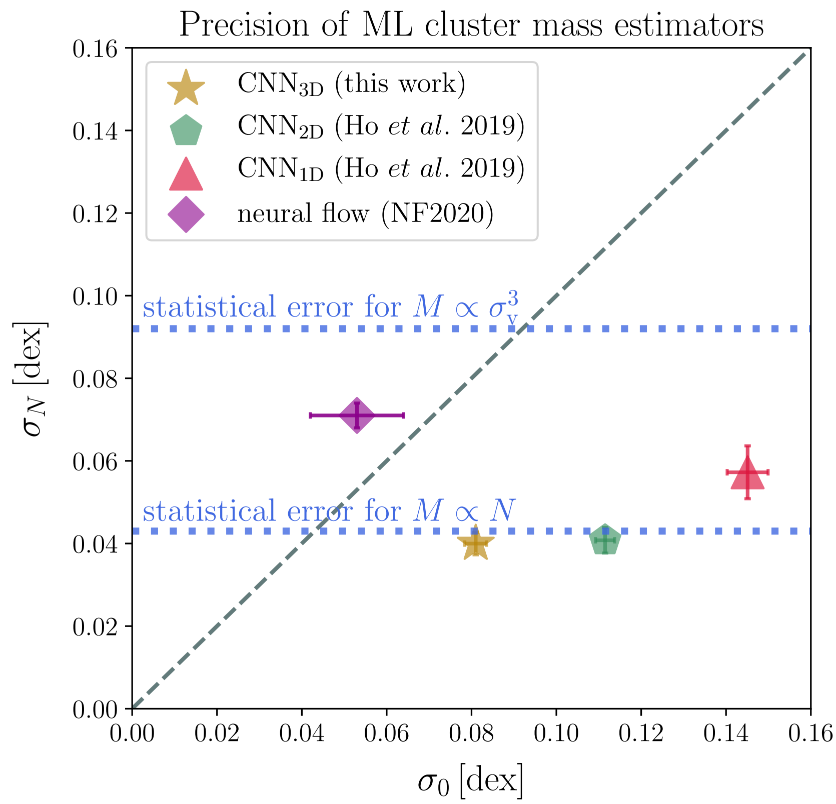


FIGURE 7.9 – Comparison of the precision of three recently proposed ML cluster mass estimators, along with the CNN$_{3D}$ model from this work, all based on mock observations from Ho et al. (2019) with the maximum projected distance from the cluster centre of $1.6\,h^{-1}$Mpc. We quantify the precision in terms of richness-dependent error σ_N and richness-independent systematic error σ_0 (cf. eq. (7.5)). In accordance with Figure 7.8, this shows the progressive improvement in the precision of CNN models with increasing input dimensionality. The horizontal dotted lines indicate the two characteristic levels of Poisson-like scatter for velocity dispersion-based and richness-based methods. Our CNN$_{3D}$ model is also less sensitive to the cluster richness than the neural flow mass estimator (NF2020).

7.5 APPLICATION TO SDSS CATALOGUE

We now apply the trained neural network to redshift data from the SDSS catalogue to infer the dynamical masses of the galaxy clusters and use the bivariate KDE (cf. Figure 7.4) to derive their corresponding uncertainties. We subsequently perform a detailed comparison of the inferred dynamical masses to recent measurements from the literature.

We use the publicly available *GalWeight* catalogue containing galaxy clusters found in the main galaxy sample of the SDSS with the GALWEIGHT algorithm (Abdullah, Wilson, Klypin, et al., 2020).[2] We select galaxy clusters at comoving distances shorter than the upper limit assumed in the mock data, i.e. $250\,h^{-1}$ Mpc, and we discard clusters whose observational cones given by the maximum physical distance of x_{proj} and y_{proj} are not included in the footprint of the SDSS main galaxy sample. The resulting sample consists of 801 galaxy clusters. For each cluster, we find velocities and positions of all galaxies from the main spectroscopic SDSS sample in its field. Angular positions were converted into physical distances assuming the deceleration parameter $q_0 = -0.55$ consistent with the Planck cosmology (Planck Collaboration, Ade, Aghanim, Arnaud, et al., 2016), although the impact of cosmological model in the adopted redshift range ($z \leq 0.085$) is negligible. We use the same cuts in the projected phase space coordinates as for the mock observations. We also adopt cluster centres and redshifts from the *GalWeight* catalogue which set them at the peak of a smoothed galaxy density in the projected phase space. The cluster catalogue also provides the measurements of dynamical cluster masses based on the virial theorem with the surface term computed for NFW density profile extrapolated beyond the virial sphere. The mass estimations account for cluster membership by a special scheme of assigning weights to all galaxies observed in the phase-space diagram. The scheme was devised using mock data generated from cosmological simulations (Abdullah et al., 2018).

The galaxy clusters from the catalogue are subjected to an initial preprocessing step similar to the preparation of the training set. We compute the 3D Gaussian KDE of their respective phase-space distributions, as outlined in Section 7.2.3, with the resulting 3D slices subsequently provided as inputs to our $\mathrm{CNN_{3D}}$ model. The point estimates are then fed to the SBI pipeline to obtain their respective uncertainties, resulting in the inferred dynamical masses for the SDSS clusters. To compare our predictions with the recent results from Abdullah, Wilson, Klypin, et al. (2020), we compute the 1D PDF of the difference between the two sets of predictions, normalised by the combined uncertainties, i.e. $\log_{10}(M_{\mathrm{pred}}/M_{\mathrm{GalWeight}})/\sigma_{\mathrm{com}}$, where $\sigma_{\mathrm{com}} \equiv (\sigma^2_{\mathrm{pred}} + \sigma^2_{\mathrm{GalWeight}})^{1/2}$ and $M_{\mathrm{GalWeight}}$ is the cluster mass estimate from the *GalWeight* galaxy cluster catalogue (Abdullah, Wilson, Klypin, et al., 2020) with associated uncertainty $\sigma_{\mathrm{GalWeight}}$. We compute the 1D PDF by binning this mass contrast, with the resulting distribution illustrated in Figure 7.10. The latter distribution has a mean and standard deviation of $\mu = -0.02$ and $\sigma = 1.05$, respectively, which approximately corresponds to a normalised Gaussian distribution. This highlights the overall consistency of our mass predictions with those from Abdullah, Wilson, Klypin, et al. (2020), with the absence of kurtosis implying a negligible bias or error underestimation/overestimation with respect to the former literature estimates, which would otherwise render the 1D PDF leptokurtic or platykurtic.

[2] https://mohamed-elhashash-94.webself.net/galwcat/

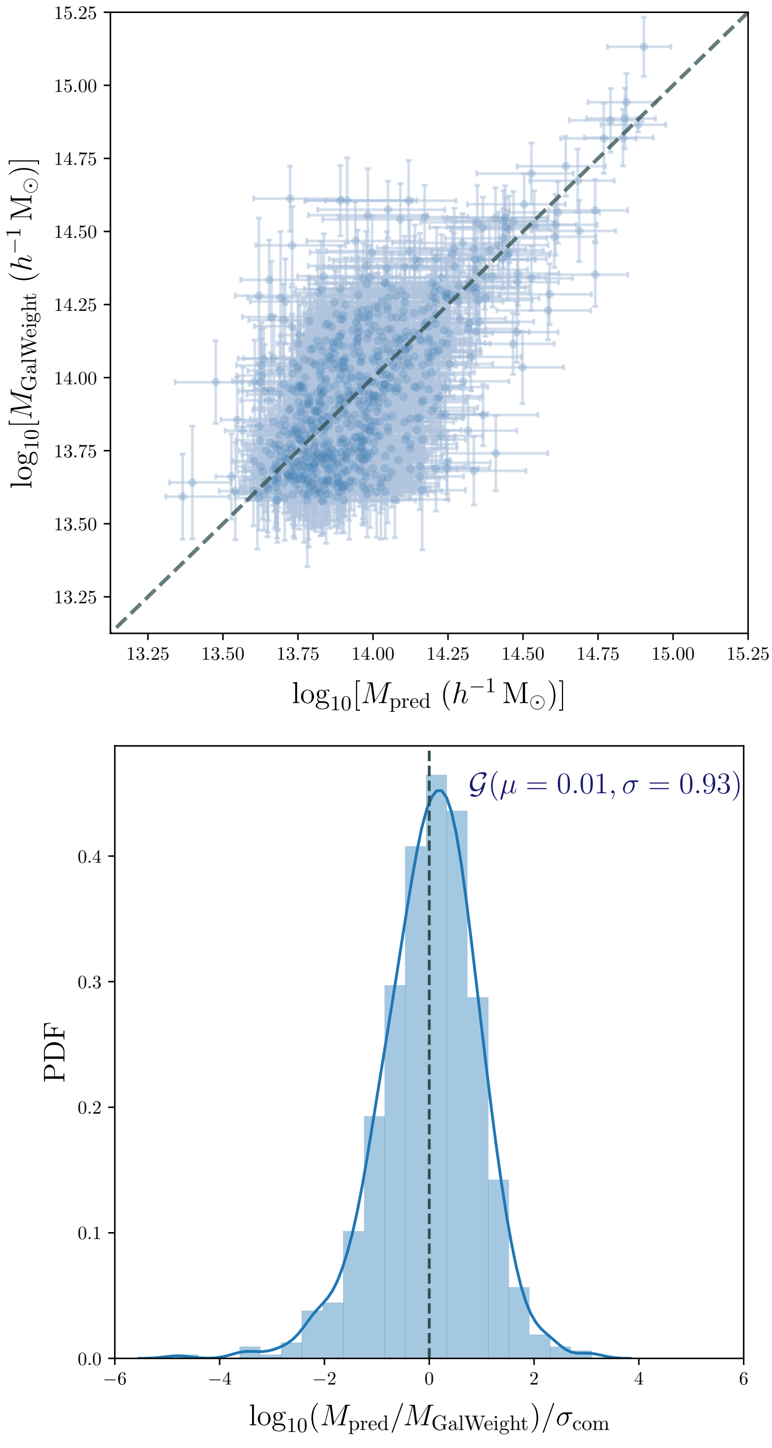


FIGURE 7.10 – Comparison of our SDSS cluster mass predictions with the recent estimates from the *GalWeight* galaxy cluster catalogue (Abdullah, Wilson, Klypin, et al., 2020), illustrated via a qualitative visual depiction (*top panel*) and a 1D PDF (*bottom panel*) of the difference between the predictions, normalised by the corresponding uncertainties in our predictions. The resulting distribution is approximately characterised by a normalised Gaussian distribution, quantitatively indicating the overall consistency between our cluster mass predictions and those from Abdullah, Wilson, Klypin, et al. (2020).

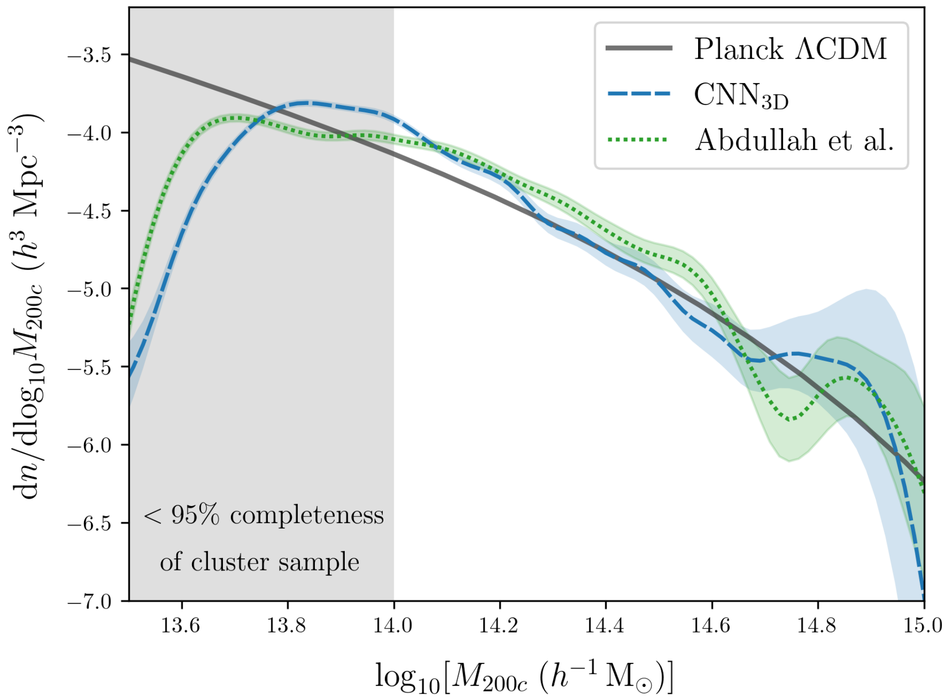


FIGURE 7.11 – Cluster mass function derived from the dynamical mass measurements obtained for a sample of 760 SDSS galaxy clusters at redshifts $z \leq 0.085$ using the new cluster mass estimation method devised in this work ($\mathrm{CNN_{3D}}$). The result is compared to the corresponding cluster mass function computed for alternative mass estimates from Abdullah, Wilson, Klypin, et al. (2020) based on the virial theorem and the theoretical halo mass function as predicted for Planck ΛCDM cosmology. The lines show the measured mass function obtained with a kernel density estimator, while the shaded bands indicate 1σ confidence interval from Poisson errors. The shaded grey region corresponds to the approximate mass range where the cluster sample is incomplete.

7.5.1 Cluster mass function

This section constitutes non-peer reviewed supplementary material that is not included in the published version.

Figure 7.11 shows the cluster mass function derived from our measurements of dynamical masses and those derived by Abdullah, Wilson, Klypin, et al. (2020). In both cases, we used the same sample of 760 galaxy clusters found in the SDSS footprint reduced by the perimeter area containing clusters affected by incompleteness due to proximity of the survey's edge. The total area of the reduced SDSS footprint is 6670 square degrees. Since the tests of the *GalWeight* cluster finder based on both mock and real SDSS observations (Abdullah, Klypin, et al., 2020; Abdullah, Wilson, Klypin, et al., 2020) ensure that the sample of rich clusters detected in the SDSS data is complete up to the maximum comoving distance adopted in our study ($z \leq 0.085$), we compute the mass function assuming a constant selection function (weights equal to 1 for all clusters). The same tests also show that a minimum cluster mass for which the cluster finder is more than 95 per cent complete within the considered comoving volume is approximately $10^{14.0}h^{-1}\mathrm{M}_{\odot}$. We indicate the corresponding mass incompleteness range with a shaded area.

The cluster mass functions computed from the two mass estimators are fully consistent. It is also readily apparent that both estimates of the cluster mass function recover the halo mass function of the Planck cosmological model (Planck Collaboration et al., 2014) with a universal fitting function from Tinker et al. (2008) down to the approximate mass limit of the cluster sample completeness, i.e. $\sim 10^{14.0}h^{-1}\mathrm{M}_{\odot}$.

7.6 CONCLUSIONS AND OUTLOOK

We have presented a simulation-based inference framework, based on 3D convolutional feature extractors, to infer the galaxy cluster masses from their 3D dynamical phase-space distributions, which consist of the projected positions in the sky and the galaxy line-of-sight velocities, i.e. $\{x_{\mathrm{proj}}, y_{\mathrm{proj}}, v_{\mathrm{los}}\}$. The simulation-based inference framework allows us to quantify the uncertainties on the inferred masses in a straightforward and robust way. By optimally exploiting the information content of the full projected phase-space distribution, the network yields dynamical cluster mass estimates with precision comparable to the best existing traditional methods. As such, this fast and robust tool is a novel and complementary addition to the state-of-the-art machine learning techniques in the cluster mass estimation toolbox.

We train our $\mathrm{CNN}_{\mathrm{3D}}$ model using a realistic mock cluster catalogue emulating the properties of the actual SDSS catalogue. Once optimised on the training set, we use our $\mathrm{CNN}_{\mathrm{3D}}$ model within a simulation-based inference framework to infer the dynamical masses of a set of SDSS clusters and their associated uncertainties for the first time using a machine learning-based cluster mass estimator, and we obtain results consistent with mass estimates from the literature. The primary advantage of simulation-based inference, as employed in this work, is that it yields accurate and statistically consistent uncertainties. If the neural network used to perform the feature extraction to derive summary statistics is sub-optimal, the uncertainties will only be inflated, thereby obviating overconfident posteriors. Moreover, we clearly illustrate the difficulties related to the presence of interlopers close to the cluster centre when dealing with actual observations. In practice, there exists no effective solution to this predicament inherent to the mass estimation problem. As a consequence, our simulation-based inference framework yields correspondingly larger uncertainties for such problematic clusters. This conservative approach ensures that the uncertainties of highly contaminated clusters are not underestimated.

The design of our network architecture, based on the use of 3D convolutional kernels, is justified by the gain in constraining power with progressively larger dimensionality of the input phase-space distribution, as substantiated by smaller log-normal residual scatter and improved precision (cf. Figures 7.8 and 7.9, respectively). Compared to our recently proposed neural flow mass estimator (Kodi Ramanah, Wojtak, Ansari, et al., 2020), our $\mathrm{CNN}_{\mathrm{3D}}$ model is more robust to the size of galaxy samples with spectroscopic redshifts, i.e. galaxy selection effects. The former method employs normalising flows, implemented via a stack of multilayer perceptrons, to predict the posterior cluster mass PDFs from 2D phase-space distributions $\{R_{\mathrm{proj}}, v_{\mathrm{los}}\}$, thereby deriving uncertainties in a conceptually distinct approach.

The performance of our CNN_{3D} mass estimator, along with that of our recent neural flow mass estimator and the variational inference approach by Ho et al. (2020), provides exciting avenues to infer cosmological constraints from the SDSS catalogue using cluster abundances (Abdullah, Klypin, et al., 2020). These three novel machine learning algorithms yield robust and reliable cluster masses with complementary ways of deriving uncertainties and, therefore, may be utilised to constrain the cluster mass function to complement standard approaches based on traditional mass estimators.

DATA AVAILABILITY

The source code repository, containing JUPYTER tutorial notebooks and the mock cluster catalogue, is available at https://github.com/doogesh/SBI_dynamical_mass_estimator.

AUTHOR CONTRIBUTIONS

DKR led the project, produced the KDE representation of the input data, designed and optimised the neural network architecture, implemented the SBI framework, evaluated the network performance, produced most of the figures, and wrote the paper. RW proposed the main idea, produced the mock catalogue, constructed the cluster mass function, validated the results, and contributed to writing-up the manuscript. NA quantified the interloper contamination, constructed the cluster mass function, investigated properties of the KDE, quantified the 1σ uncertainty regions of the asymmetric posterior distributions, produced Figures 7.7 and 7.11, assisted with the network optimisation and contributed to writing-up the manuscript.

ACKNOWLEDGEMENTS

We express our gratitude to the reviewer whose detailed feedback helped to improve various aspects of our analysis. We acknowledge Tom Charnock for providing us with the motivation to make use of simulation-based inference with neural networks. We also thank Matthew Ho for his insights pertaining to the use of KDE for preprocessing the phase-space distribution and for his constructive feedback on our work. We are grateful to Guilhem Lavaux and Jens Jasche for interesting discussions related to Bayesian posterior validation. We also express our appreciation to Gary Mamon for his perceptive comments on our manuscript. DKR is a DARK fellow supported by a Semper Ardens grant from the Carlsberg Foundation (reference CF15-0384). This work was supported by a VILLUM FONDEN Investigator grant (project number 16599). This work has made use of the Horizon and Henon Clusters hosted by Institut d'Astrophysique de Paris. This work has been done within the activities of the Domaine d'Intérêt Majeur (DIM) Astrophysique et Conditions d'Apparition de la Vie (ACAV), and received financial support from Région Ile-de-France. This work made use of an HPC facility funded by a grant from VILLUM FONDEN (project number 16599).

CHAPTER 8

Conclusions & outlook

The aim of this thesis was to address tensions and open questions in cosmology from both an observational and theoretical perspective, aided by machine learning techniques. Altogether, the main findings and contributions of the doctoral research to the field are:

- A strong (5σ) tension arises in terms of the Hubble constant and sound horizon when distance calibrations from SHoES and HoLiCOW in combination with Baryon Acoustic Oscillations are compared to observations from the CMB. This combined tension cannot be solved in a satisfying way by modifications of the standard ΛCDM model, neither by early-time nor by late-time extensions. This might tentatively point in the direction of systematics as a solution to the prevailing tension, or perhaps to the need for a combination of several models. These findings, which are described in Chapter 4, underline the requirement for new, independent measurements of the cosmic expansion rate.
- Gravitationally lensed supernovae are a promising probe for such new measurements. Our deep learning pipeline presented in Chapter 5 exploits the full spatial and temporal information of time-series images and provides a way to distinguish lensed supernovae from unlensed ones in transient surveys.
- Simulated LSST time-series images corresponding to a realistic sample of type Ia supernovae can yield a competitive joint measurement of the Hubble constant, even without the addition of follow-up observations. Chapter 6 presents a spatio-temporal neural network capable of performing such an inference. The dominant sources of uncertainty for this measurement are the time delays between the lensed images, and the supernova source position (for doubles) or the image positions (for quads).
- Machine learning-based galaxy cluster mass estimators are able to construct a cluster

mass function that is consistent with mass estimates from the literature and with the predicted halo mass function from *Planck* ΛCDM cosmology. We have demonstrated this in Chapter 7 by means of a convolutional neural network that, for the first time, exploits the full 3D projected phase-space distribution of galaxy clusters.

Prospects for upcoming galaxy surveys

The frameworks developed in this thesis will become even more useful when applied to data from the next-generation surveys.

The spectroscopic galaxy survey DESI will provide a dense sampling of BAO measurements up to a redshift of $z = 3$. Consequently, this will allow for a much tighter measurement of the sound horizon and Hubble constant, which can be used to test new cosmological models in more stringent ways. Additionally, the data will yield considerably better constraints on the cosmic curvature from low-redshift observations.

Future wide-field surveys to be conducted at the Vera C. Rubin Observatory, the Nancy Grace Roman Space Telescope, and *Euclid* will supply an unparalleled sample of type Ia supernovae to constrain the shape of the expansion history to much higher precision. Moreover, they are predicted to discover orders of magnitudes more gravitationally lensed supernovae than the number of presently known objects. With the unprecedented volumes of data that will be produced each night, our automated deep learning pipelines for lensed supernovae detection and H_0 inference will be essential tools to extract insights and optimise the scientific returns from the data.

However, the high-precision data from upcoming missions will also pose new challenges. It will be of paramount importance to ensure that the models used to interpolate the expansion history do not introduce any biases, not even at a sub-percent level. The cosmographic models adopted in Chapters 3 and 4 are adequate for the current lower-precision data, but will fall short when it comes to modelling and interpreting the high-redshift and high-precision data from next-generation instruments. Therefore, we need to develop improved cosmology-independent methods, possibly aided by advances in the field of machine learning, such as Gaussian Processes, genetic algorithms and symbolic regression.

Prospects for lensed supernovae

In order to optimise the lensed supernova detection pipeline further, several features can be included in the simulated training set to make it even more similar to real observations. Firstly, instead of generating images where only the supernova is visible, we could compute more realistic difference images by subtracting a reference image of the lens and host galaxy from an image containing the lens galaxy, host galaxy and supernovae. This procedure would lead to some image subtraction artefacts that are also visible in the real data. Additionally, we could add a third class of objects that consists of active galactic nuclei, since these sources can potentially mimic lensed supernovae.

Currently, we are in the process of testing our lensed supernovae detection pipeline on real data from the Young Supernova Experiment and the Zwicky Transient Factory. Although the predicted lensed supernovae rate in the aforementioned transient surveys is at most a few objects per year, it will be useful to develop insights into the possible sources of contamination and false positives. In this way, we will be fully prepared when the data sets from LSST become available, which should constitute a much higher lensed supernovae rate.

When employing lensed supernovae for cosmological inference, it is of even greater importance to ensure that the training set is as similar as possible to real observations. Any physical features that are neglected in the simulation can lead to an underestimation of the uncertainties on the parameter of interest. For lensed supernovae, microlensing constitutes a particular challenge. Microlensing contributions are generally unpredictable, time-varying, and different for each supernova image, which can introduce additional uncertainties in the time-delay measurements. In an attempt to overcome this issue, we have established a new collaboration with Simon Huber, whose work describes microlensing in a detailed manner through magnification maps and radiative transfer simulations. Our aim is to convert this elaborate microlensing description into an accurate statistical description that can be easily implemented in simulations, so as to robustly quantify the influence of microlensing on the Hubble constant inference.

Another point of consideration is the mass model adopted for the lens galaxy. In our study, we employ a PEMD model where the Einstein radius, ellipticity and power-law slope are free parameters, in combination with external shear. However, Birrer et al. (2020) show that the assumed shape of this model partly breaks the mass-sheet degeneracy. In other words, the PEMD + shear model is not flexible enough to be be completely driven by the data, and an additional parameter should be added to the PEMD model to account for the mass-sheet transformation. As an avenue for future work, we could introduce more free parameters into the PEMD model, adopt a different lens model altogether, or even consider a non-parametric approach to describe the lens potential.

An interesting result of our study is that doubles will give rise to the vast majority of constraining power on H_0. Compared to quads, their time delays and image positions can be determined more accurately by the neural network, and they are predicted to comprise the largest fraction of detected lensed supernovae. However, the use of quads in cosmological inference is not as straightforward as for doubles. The manner in which we employ them in our study, i.e. by only considering the maximum time delay between the first and the last image, may not be the optimal one. It will be interesting to consider alternative avenues for inference on quads with neural networks, ideally by including the full information content from all four images, to see if this can improve the accuracy and precision of their final H_0 measurement.

While our work, as a proof on concept, only considers lensed type Ia supernova observations in the i-band, a natural extension would be to include images from multiple bands. This will considerably improve the cadence and, consequently, the precision on the time delay estimates and the Hubble constant. Additionally, we would like to include other types of supernovae in the simulation in order to increase our sample size. Type IIn supernovae, in particular, are promising objects, since they are predicted to make up the largest fraction of the lensed supernovae population.

Prospects for dynamical cluster mass inference

With the machinery for accurate and robust cluster mass estimates in place, future work in this direction can take the exciting avenue of using the cluster mass function to infer constraints on the matter density and clustering amplitude, which are complementary to standard approaches based on traditional mass estimators. However, this line of work would require further study into the cluster selection function for our specific sample, which characterises the completeness of the sample. Without a good understanding of the selection function, it is difficult to quantify how representative the observed objects are of the underlying parent population, which can lead to biased estimates of cosmological parameters.

Another interesting avenue for further study is to compare our mass uncertainties inferred via the simulation-based inference framework to those derived with a neural flow mass estimator by Kodi Ramanah, Wojtak, Ansari, et al. (2020) and a variational inference approach by Ho et al. (2020). This could lead to new insights into simulations used to generate mock galaxy cluster catalogues, and into the different methods of uncertainty estimation for deep learning models.

IV

Appendices

APPENDIX A

Appendix

A.1 Planck compressed likelihood

Supporting material for Chapter 4.

Much of the constraining power of the CMB power spectrum can be compressed in three parameters: the physical density of baryons $\Omega_b h^2$, which determines relative heights of the peaks in the power spectrum, and two so-called shift parameters that describe two fundamental and directly measured angular scales related to the sound horizon and the Hubble horizon at decoupling. The shift parameters are defined by the following equations:

$$\mathcal{R} = \sqrt{\Omega_m}\frac{D_A(z_*)}{H_0^{-1}} \tag{A.1}$$

$$\theta_* = \frac{r_s(z_*)}{D_A(z_*)}, \tag{A.2}$$

where z_* is redshift of decoupling and D_A is the comoving angular diameter distance which for flat models considered in this work is given by

$$D_A = c\int_0^z \frac{dz}{H(z)} \tag{A.3}$$

$$H^2(z) = H_0^2[\Omega_m(1+z)^3 + \Omega_{DE}(z) + \Omega_\gamma(1+z)^4], \tag{A.4}$$

where Ω_γ denotes the density parameter of radiation, i.e. $\Omega_\gamma = 2.47\times10^{-5}h^{-2}$.

The comoving sound horizon is given by

$$r_s(z) = \frac{c}{\sqrt{3}}\int_z^\infty \frac{dz}{H(z)\sqrt{1+\frac{3\Omega_b}{4\Omega_\gamma}(1+z)^{-1}}}. \tag{A.5}$$

Here, an additional contribution to the energy density driving the expansion comes from relativistic neutrinos. The density parameter of relativistic neutrinos $\Omega_{\rm n}$ is given by

$$\Omega_{\rm n} = N_{\rm eff}\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\Omega_{\gamma}, \tag{A.6}$$

where $N_{\rm eff}$ is the effective number of neutrinos with $N_{\rm eff} = 3.046$ for the baseline model.

We compute redshift z_* of decoupling employing the following fitting formula (Hu & Sugiyama, 1996)

$$z_* = 1047[1 + 0.00124(\Omega_{\rm b}h^2)^{-0.738}][1 + g_1(\Omega_{\rm m}h^2)^{g_2}] \tag{A.7}$$
$$g_1 = 0.0783(\Omega_{\rm b}h^2)^{-0.238}[1 + 39.5(\Omega_{\rm b}h^2)^{0.763}]^{-1} \tag{A.8}$$
$$g_2 = 0.56[1 + 21.1(\Omega_{\rm b}h^2)^{1.81}] \tag{A.9}$$

The sound horizon imprinted in galaxy clustering and measured from BAO observations is fixed at the drag epoch when the baryons are released from the Compton drag of the photons. The corresponding drag redshift $z_{\rm d}$ can be calculated using the following fitting function (Hu & Sugiyama, 1996)

$$z_{\rm d} = 1345\frac{(\Omega_{\rm m}h^2)^{0.251}[1 + b_1(\Omega_{\rm b}h^2)^{b_2})]}{1 + 0.659(\Omega_{\rm m}h^2)^{0.828}} \tag{A.10}$$
$$b_1 = 0.313(\Omega_{\rm m}h^2)^{-0.419}[1 + 0.607(\Omega_{\rm m}h^2)^{0.674}] \tag{A.11}$$
$$b_2 = 0.238(\Omega_{\rm m}h^2)^{0.223}. \tag{A.12}$$

The compressed CMB likelihood is given by a three-dimensional Gaussian distribution in the three parameters mentioned above, i.e. $\Omega_{\rm n}h^2$, $\mathcal{R}$ and θ_*. We employ the mean values and the covariance matrix determined from publicly available MCMC models obtained for a flat ΛCDM model fitted to the *Planck* observations including the temperature, polarisation and lensing data (Planck Collaboration, Aghanim, et al., 2018): $(100\Omega_{\rm b}h^2, 100\theta_*, \mathcal{R}) = (2.237 \pm 0.015, 1.0411 \pm 0.00031, 1.74998 \pm 0.004)$ with the following correlation matrix

$$\begin{pmatrix} 1.00 & 0.34 & -0.63 \\ 0.34 & 1.00 & -0.46 \\ -0.63 & -0.46 & 1.00 \end{pmatrix} \tag{A.13}$$

The compressed likelihood recovers accurately the actual constraints obtained from the complete likelihood for a flat ΛCDM model (see Fig. A.1). Only a fine adjustment of the redshift scales in both fitting formulae ($\delta z/z \sim 10^{-3}$ smaller relative to the values adopted in Hu and Sugiyama (1996)) was applied in order to correct for a sub-percent bias in the mean values of relevant parameters. In general, both approximations used to compute z_* and $z_{\rm drag}$ are accurate to within 1 per cent in a wide range of the matter and baryon density parameters (Hu & Sugiyama, 1996).

For early-time extensions of the standard ΛCDM cosmology (such as a model with free $N_{\rm eff}$), the compressed likelihood turns out to be insufficient, leading to a family of models with a wide range of amplitudes of the first peak in the power spectrum. In order to circumvent

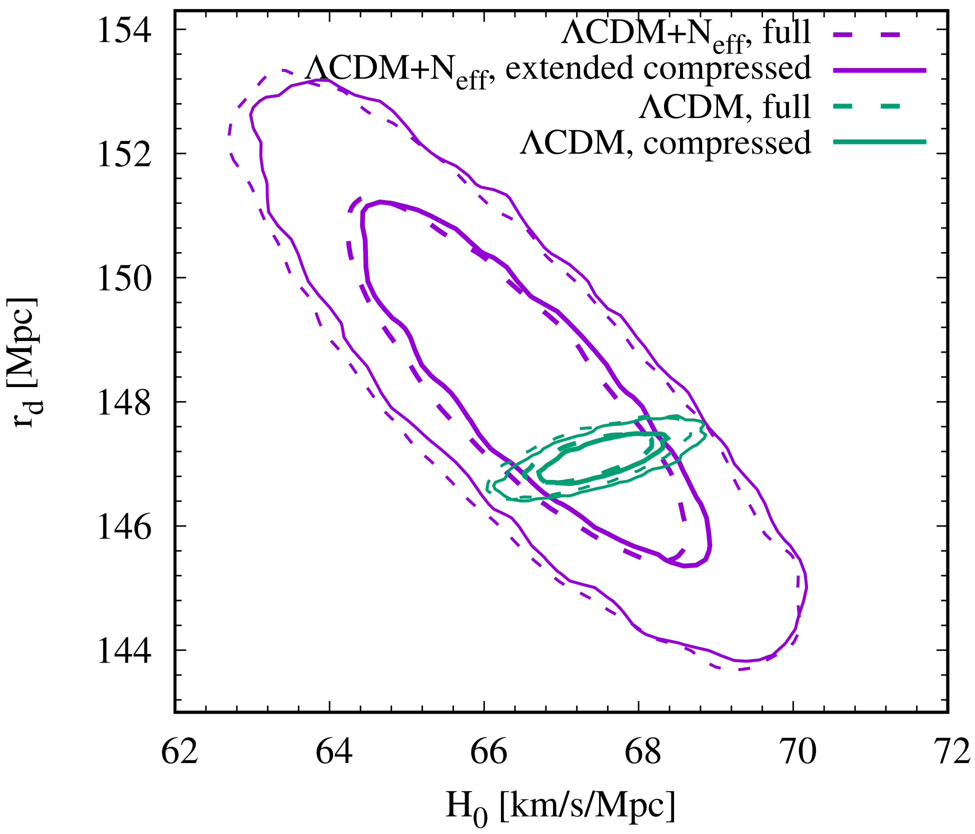


FIGURE A.1 – Comparison between constraints on r_d and H_0 from the full *Planck* likelihood (dashed lines) and the compressed likelihood (for post-recombination modifications of ΛCDM) or the extended compressed likelihood (for pre-recombination modifications of ΛCDM) used in this study (solid lines). The robustness test comprises two cases: the standard flat ΛCDM model and its extension with a free number of neutrinos.

this problem, we extend the compressed likelihood described above by accounting for the height of the first peak in the power spectrum as an additional constraint. Bearing in mind that the amplitude scales with $\Omega_{dm}h^2$, i.e. the physical density of dark matter, a simple extension relies on adding $\Omega_{dm}h^2$ as the fourth variable in the compressed likelihood function. Using *Planck* results for a ΛCDM model with a free effective number of neutrinos as a base early-time extension (inferred from the full temperature and polarisation data), we determine the mean values and the covariance matrix of the new four-parameter compressed likelihood, obtaining $(100\Omega_b h^2, 100\theta_*, \mathcal{R}, \Omega_{dm}h^2) = (2.225\pm0.0223, 1.0414\pm0.00054, 1.7529\pm0.0056, 0.1184\pm0.0029)$

and the following correlation matrix

$$\begin{pmatrix} 1.00 & -0.50 & -0.79 & 0.51 \\ -0.50 & 1.00 & 0.30 & -0.81 \\ -0.79 & 0.30 & 1.00 & -0.19 \\ 0.51 & -0.81 & -0.19 & 1.00 \end{pmatrix} \tag{A.14}$$

Fig. A.1 demonstrates that the extended compressed likelihood accurately recovers the actual constraints on $r_{\rm d}$ and H_0 from *Planck* for a model with a free effective number of neutrinos.

A.2 POLYNOMIAL PARAMETRISATIONS

Supporting material for Chapter 4.

THIS section gives more detailed information about the polynomial parametrisations used throughout this work.

A.2.1 Expansion formulas

Our first model is the simplest one and adopts a polynomial expansion of $H(z)$ in z.

$$H(z) = H_0\left[1 + b_1\, z + b_2\, z^2 + \mathcal{O}(z^3)\right], \tag{A.15}$$

where H_0 is the Hubble constant and the coefficient b_1 is related to the deceleration parameter q_0 through

$$b_1 = 1 + q_0. \tag{A.16}$$

In our second model, the luminosity distance D_L is expanded as a polynomial in $\log(1+z)$. [1]

$$\begin{aligned} x &= \log(1+z), \\ D_L(z) &= \frac{c\ln(10)}{H_0}\left[x + c_2 x^2 + c_3 x^3 + c_4 x^4 + \mathcal{O}(x^5)\right], \end{aligned} \tag{A.17}$$

where the coefficient c_2 is related to the deceleration parameter through the following relation:

$$c_2 = \frac{\ln(10)}{2}(2 - q_0). \tag{A.18}$$

This different parametrisation is chosen in order to avoid convergence problems with the Taylor expansion around zero, when employing data with redshifts $z > 1$. By introducing a new variable x that satisfies $x = 0$ when $z = 0$ and $x < 1$ when $z \to 2$ (where the upper limit of 2 is based on the highest lensed quasar redshift), the parametrisation is kept within the convergence radius of the Taylor expansion.

Our third model describes transverse comoving distances D_M by polynomials in $z/(1+z)$.

$$\begin{aligned} y &= \frac{z}{1+z}, \\ D_M(z) &= \frac{c}{H_0}\left[y + d_2 y^2 + d_3 y^3 + d_4 y^4 + \mathcal{O}(y^5)\right], \end{aligned} \tag{A.19}$$

where the coefficient d_2 is related to the deceleration parameter through

$$d_2 = \tfrac{1}{2}(1 - q_0). \tag{A.20}$$

This parametrisation is, similar to the one in Model 2, chosen to overcome convergence problems.

A.2.2 Truncation of the polynomials

An important thing to consider is at which order the Taylor expansions should be truncated. Higher orders of expansions can give better approximations to the shape of the data, but also

[1] Here, $\log(1+z)$ refers to the log base 10, and not to the natural logarithm.

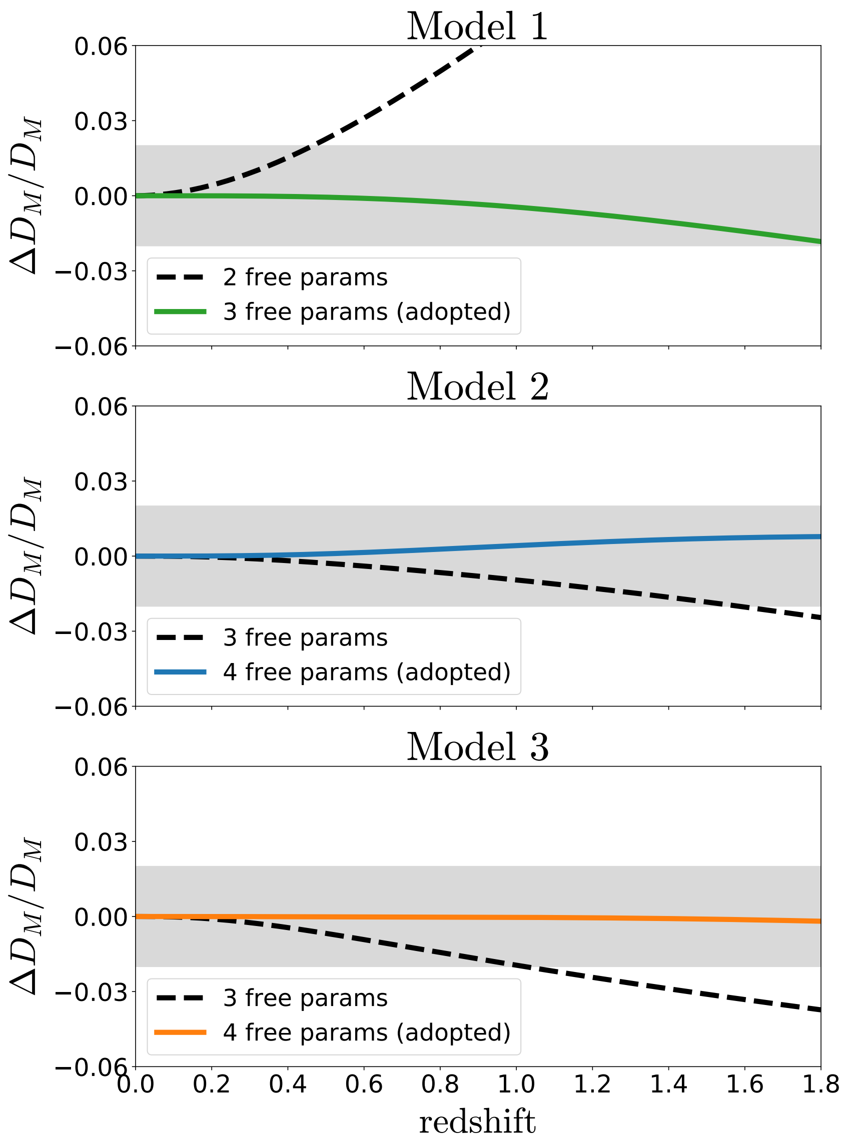


Figure A.2 – Relative differences between distances in a fiducial flat ΛCDM model and distances derived from Models 1-3 with free parameters matched to the kinematical coefficients of the fiducial model, $\Delta D_{\mathrm{M}}/D_{\mathrm{M}} = (D_{\mathrm{M,\,expansion}} - D_{\mathrm{M,\Lambda CDM}})/D_{\mathrm{M,\Lambda CDM}}$. The solid lines show the results satisfying the convergence criterion which sets the truncation of polynomials used in the adopted models in this study.

introduce more free parameters and therefore larger uncertainties. In order to determine the truncation of the polynomials as given in A.15, A.17 and A.19, we perform a convergence test to check that the models can accurately recover expansion history of a fiducial flat ΛCDM cosmological model in a redshift range of observational data used in our study, i.e. $z < 1.8$. The test relies on comparing distances from Models 1-3 to the actual distances in the fiducial model. Free parameters of the models are determined by matching coefficients of Taylor expanded Hubble parameter in Models 1-3 and the fiducial model. The latter yields well-known kinematical coefficients (Visser, 2004; Weinberg, 1972):

$$\begin{aligned} q_0 &= \tfrac{3}{2}\Omega_\mathrm{m} - 1, \\ j_0 &= 1, \\ s_0 &= 1 - \tfrac{9}{2}\Omega_\mathrm{m}. \end{aligned} \tag{A.21}$$

Since the errors that we obtain by combining calibrations of HoLiCOW and SHoES are around 2% (see Table 4.3), we require our models to be within a 2 % accuracy of ΛCDM distances in this test. The results can be seen in figure A.2 for $\Omega_m = 0.3$, where the shaded region corresponds to this imposed limit. It suffices to employ three free parameters (corresponding to a second order polynomial) for Model 1 and four free parameters (corresponding to a fourth order polynomial) for models 2 and 3 to satisfy the convergence condition. Since a further increase of the number of free parameters is disfavoured by the BIC obtained in fits with the actual late-time observations, these polynomial truncations are adopted in our study (see Table 4.1). The BIC score is calculated as

$$\mathrm{BIC} \; = \; \ln(N)k - 2\ln(\mathcal{L}_{m.a.p.}), \tag{A.22}$$

where N is the number of data points and k is the number of all free parameters in the cosmological fits.

A.2.3 Test with mock distance modulus data

As a final test for our polynomial parametrisation models, we investigate if any biases are introduced when we fit Models 1-3 to flat ΛCDM data. We transform the Pantheon SN data set to a mock data set, by replacing their binned distance modulus entries by the fiducial flat ΛCDM values (adopting $H_0 = 74$ km s^{-1} Mpc^{-1} and $\Omega_\mathrm{m} = 0.3$) at the same redshifts. For the errors associated with the distance moduli we keep the original Pantheon ones. By construction, best fit ΛCDM parameters are equal to their fiducial values, whereas relative shifts in best fit parameters obtained for non-ΛCDM models measure the corresponding biases. This test is similar to the one performed by Yang et al., 2019, in which they find that our Model 2 introduces an artificial bias. However, their mock data set is based on Pantheon data as well as high-redshift quasar and GRB data (with $z_\mathrm{max} = 6.7$), while in our work we only use sources below $z = 1.8$. Figure A.3 shows the best-fit values for the coefficients b_i, c_i and d_i of Models 1-3, obtained with MCMC, and their true values in a flat ΛCDM cosmology. As can be seen, they are in complete agreement with each other. In fact, the relative difference in H_0 between the fiducial value and those of Models 1-3 is 0.03%, 0.02% and 0.02%, respectively. This bias is about a hundred times smaller than the current precision achieved by SHoES and HoLiCOW data (which is around 2%). The bias in $q0$ is larger: 2.0%, 1.2% and 1.3% for Models 1-3, but still negligible compared to our obtained errors in q_0 (which are 10% at best).

This test demonstrates that if the underlying cosmology is flat ΛCDM, then our models will not introduce any significant biases in the Pantheon redshift range. The convergence test in the previous section also guarantees this. The bias that Yang et al., 2019 found in their model was a consequence of it not passing the convergence test over the complete redshift range of $z = 0-7$.

We repeat the test for the PEDE model and for a wCDM cosmology with $w = -1.2$. In both cases we assume $\Omega_{\rm m} = 0.3$. We find only a sub-percent bias in the best fit H_0 and a few-percent bias in q_0 where the actual values are given by

$$
\begin{aligned}
q_{0,\rm PEDE} &= \tfrac{3}{2}\Omega_{\rm m} - \frac{1-\Omega_{\rm m}}{2\ln(10)} - 1, \\
q_{0,w\rm CDM} &= \tfrac{1}{2} + \tfrac{3}{2}w(1-\Omega_{\rm m}).
\end{aligned}
\tag{A.23}
$$

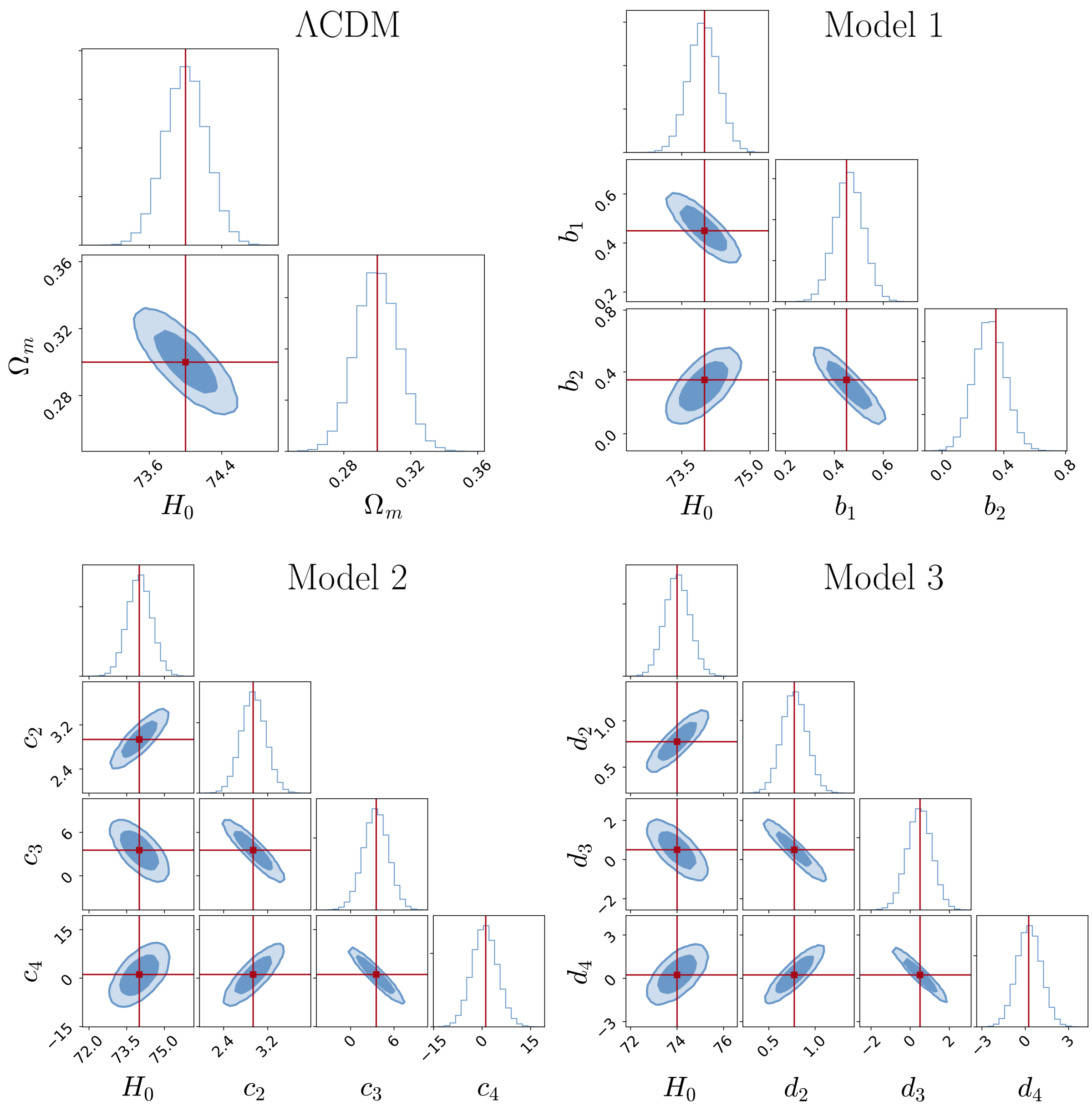


FIGURE A.3 – Best-fit values of flat ΛCDM and polynomial parametrisation Models 1-3 to mock data. The mock data is generated by replacing the Pantheon distance modulus points by their fiducial flat ΛCDM values. The red lines indicate the canonical ΛCDM values of Ω_m, H_0 and the expansion coefficients b_i, c_i and d_i.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv e-prints*.

Abdullah, M. H., Klypin, A., & Wilson, G. (2020). Cosmological Constraints on Ω_m and σ_8 from Cluster Abundances Using the GalWCat19 Optical-spectroscopic SDSS Catalog. *ApJ*, *901*(2), Article 90, 90. https://doi.org/10.3847/1538-4357/aba619

Abdullah, M. H., Wilson, G., & Klypin, A. (2018). GalWeight: A New and Effective Weighting Technique for Determining Galaxy Cluster and Group Membership. *ApJ*, *861*(1), Article 22, 22. https://doi.org/10.3847/1538-4357/aac5db

Abdullah, M. H., Wilson, G., Klypin, A., Old, L., Praton, E., & Ali, G. B. (2020). GalWeight Application: A Publicly Available Catalog of Dynamical Parameters of 1800 Galaxy Clusters from SDSS-DR13, (GalWCat19). *ApJS*, *246*(1), Article 2, 2. https://doi.org/10.3847/1538-4365/ab536e

Abell, G. O. (1958). The Distribution of Rich Clusters of Galaxies. *ApJS*, *3*, 211. https://doi.org/10.1086/190036

Addison, G. E., Watts, D. J., Bennett, C. L., Halpern, M., Hinshaw, G., & Weiland, J. L. (2018). Elucidating ΛCDM: Impact of Baryon Acoustic Oscillation Measurements on the Hubble Constant Discrepancy. *ApJ*, *853*(2), Article 119, 119. https://doi.org/10.3847/1538-4357/aaa1ed

Akeret, J., Refregier, A., Amara, A., Seehars, S., & Hasner, C. (2015). Approximate Bayesian computation for forward modeling in cosmology. *J. Cosmology Astropart. Phys.*, *2015*(8), Article 043, 043. https://doi.org/10.1088/1475-7516/2015/08/043

Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D., Blazek, J. A., Bolton, A. S., Brownstein, J. R., Burden, A., Chuang, C.-H., Comparat, J., Cuesta, A. J., Dawson, K. S., Eisenstein, D. J., Escoffier, S., Gil-Marín, H., Grieb, J. N., Hand, N., Ho, S., ... Zhao, G.-B. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *MNRAS*, *470*(3), 2617–2652. https://doi.org/10.1093/mnras/stx721

Albrecht, A., & Steinhardt, P. J. (1982). Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking. *Phys. Rev. Lett.*, *48*(17), 1220–1223. https://doi.org/10.1103/PhysRevLett.48.1220

Alcock, C., & Paczynski, B. (1979). An evolution free test for non-zero cosmological constant. *Nature*, *281*, 358. https://doi.org/10.1038/281358a0

Alestas, G., & Perivolaropoulos, L. (2021). Late-time approaches to the Hubble tension deforming H(z), worsen the growth tension. *MNRAS*, *504*(3), 3956–3962. https://doi.org/10.1093/mnras/stab1070

Alpher, R. A., & Herman, R. (1948). Evolution of the Universe. *Nature*, *162*(4124), 774–775. https://doi.org/10.1038/162774b0

Alsing, J., Charnock, T., Feeney, S., & Wandelt, B. (2019). Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, *488*(3), 4440–4458. https://doi.org/10.1093/mnras/stz1960

Alsing, J., & Wandelt, B. (2019). Nuisance hardened data compression for fast likelihood-free inference. *MNRAS*, *488*(4), 5093–5103. https://doi.org/10.1093/mnras/stz1900

Alsing, J., Wandelt, B., & Feeney, S. (2018). Massive optimal data compression and density estimation for scalable, likelihood-free inference in cosmology. *MNRAS*, *477*(3), 2874–2885. https://doi.org/10.1093/mnras/sty819

Amanullah, R., Goobar, A., Clément, B., Cuby, J. .-G., Dahle, H., Dahlén, T., Hjorth, J., Fabbro, S., Jönsson, J., Kneib, J. .-P., Lidman, C., Limousin, M., Milvang-Jensen, B., Mörtsell, E., Nordin, J., Paech, K., Richard, J., Riehm, T., Stanishev, V., & Watson, D. (2011). A Highly Magnified Supernova at z = 1.703 behind the Massive Galaxy Cluster A1689. *ApJ*, *742*(1), Article L7, L7. https://doi.org/10.1088/2041-8205/742/1/L7

Amon, A., Gruen, D., Troxel, M. A., MacCrann, N., Dodelson, S., Choi, A., Doux, C., Secco, L. F., Samuroff, S., Krause, E., Cordero, J., Myles, J., DeRose, J., Wechsler, R. H., Gatti, M., Navarro-Alsina, A., Bernstein, G. M., Jain, B., Blazek, J., . . . Weller, J. (2021). Dark Energy Survey Year 3 Results: Cosmology from Cosmic Shear and Robustness to Data Calibration. *arXiv e-prints*, Article arXiv:2105.13543, arXiv:2105.13543.

Anchordoqui, L. A., Di Valentino, E., Pan, S., & Yang, W. (2021). Dissecting the H_0 and S_8 tensions with Planck + BAO + supernova type Ia in multi-parameter cosmologies. *Journal of High Energy Astrophysics*, *32*, 28–64. https://doi.org/10.1016/j.jheap.2021.08.001

Andersen, P. (2018). *Candles that light up the Dark* (Doctoral dissertation).

Aragon-Calvo, M. A. (2019). Classifying the large-scale structure of the universe with deep neural networks. *MNRAS*, *484*(4), 5771–5784. https://doi.org/10.1093/mnras/stz393

Arendse, N. (2018). *Investigating the build-up of the Local Group galaxies with the CLUES simulation* (Doctoral dissertation).

Arendse, N., Agnello, A., & Wojtak, R. J. (2019). Low-redshift measurement of the sound horizon through gravitational time-delays. *A&A*, *632*, Article A91, A91. https://doi.org/10.1051/0004-6361/201935972

Arendse, N., Kodi Ramanah, D., & Wojtak, R. (2021). Inferring the Hubble constant from gravitationally lensed supernovae in LSST with spatio-temporal neural networks. *In prep.*

Arendse, N., Wojtak, R. J., Agnello, A., Chen, G. C. .-F., Fassnacht, C. D., Sluse, D., Hilbert, S., Millon, M., Bonvin, V., Wong, K. C., Courbin, F., Suyu, S. H., Birrer, S., Treu, T., & Koopmans, L. V. E. (2020). Cosmic dissonance: are new physics or systematics behind a short sound horizon? *A&A*, *639*, Article A57, A57. https://doi.org/10.1051/0004-6361/201936720

Armitage, T. J., Kay, S. T., & Barnes, D. J. (2019). An application of machine learning techniques to galaxy cluster mass estimation using the MACSIS simulations. *MNRAS*, *484*(2), 1526–1537. https://doi.org/10.1093/mnras/stz039

Arnouts, S., Schiminovich, D., Ilbert, O., Tresse, L., Milliard, B., Treyer, M., Bardelli, S., Budavari, T., Wyder, T. K., Zucca, E., Le Fèvre, O., Martin, D. C., Vettolani, G., Adami, C., Arnaboldi, M., Barlow, T., Bianchi, L., Bolzonella, M., Bottini, D., . . . Zanichelli, A. (2005). The GALEX VIMOS-VLT Deep Survey Measurement of the Evolution of the 1500 Å Luminosity Function. *ApJ*, *619*(1), L43–L46. https://doi.org/10.1086/426733

Ata, M., Baumgarten, F., Bautista, J., Beutler, F., Bizyaev, D., Blanton, M. R., Blazek, J. A., Bolton, A. S., Brinkmann, J., Brownstein, J. R., Burtin, E., Chuang, C.-H., Comparat, J., Dawson, K. S., de la Macorra, A., Du, W., du Mas des Bourboux, H., Eisenstein, D. J., Gil-Marín, H., . . . Zhu, F. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2. *MNRAS*, *473*(4), 4773–4794. https://doi.org/10.1093/mnras/stx2630

Avestruz, C., Li, N., Zhu, H., Lightman, M., Collett, T. E., & Luo, W. (2019). Automated Lensing Learner: Automated Strong Lensing Identification with a Computer Vision Technique. *ApJ*, *877*(1), Article 58, 58. https://doi.org/10.3847/1538-4357/ab16d9

Avni, Y., & Tananbaum, H. (1986). X-Ray Properties of Optically Selected QSOs. *ApJ*, *305*, 83. https://doi.org/10.1086/164230

Aylor, K., Joy, M., Knox, L., Millea, M., Raghunathan, S., & Kimmy Wu, W. L. (2019). Sounds Discordant: Classical Distance Ladder and ΛCDM-based Determinations of the Cosmological Sound Horizon. *ApJ*, *874*, Article 4, 4. https://doi.org/10.3847/1538-4357/ab0898

Barbary, K., Barclay, T., Biswas, R., Craig, M., Feindt, U., Friesen, B., Goldstein, D., Jha, S., Rodney, S., Sofiatti, C., Thomas, R. C., & Wood-Vasey, M. (2016). SNCosmo: Python library for supernova cosmology.

Barkana, R., & Loeb, A. (2001). In the beginning: the first sources of light and the reionization of the universe. *physrep*, *349*, 125–238. https://doi.org/10.1016/S0370-1573(01)00019-9

Barkana, R. (1998). Fast Calculation of a Family of Elliptical Mass Gravitational Lens Models. *ApJ*, *502*(2), 531–537. https://doi.org/10.1086/305950

Baron, D. (2019). Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, Article arXiv:1904.07248, arXiv:1904.07248.

Bassett, B., & Hlozek, R. (2010). Baryon acoustic oscillations. In P. Ruiz-Lapuente (Ed.), *Dark energy: Observational and theoretical approaches* (p. 246).

Bautista, J. E., Vargas-Magaña, M., Dawson, K. S., Percival, W. J., Brinkmann, J., Brownstein, J., Camacho, B., Comparat, J., Gil-Marín, H., Mueller, E.-M., Newman, J. A., Prakash, A., Ross, A. J., Schneider, D. P., Seo, H.-J., Tinker, J., Tojeiro, R., Zhai, Z., & Zhao, G.-B. (2018). The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations at Redshift of 0.72 with the DR14 Luminous Red Galaxy Sample. *ApJ*, *863*(1), Article 110, 110. https://doi.org/10.3847/1538-4357/aacea5

Bayes, M., & Price, M. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London Series I*, *53*, 370–418.

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. (2013). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *ApJ*, *762*, Article 109, 109. https://doi.org/10.1088/0004-637X/762/2/109

Bellm, E. C., Kulkarni, S. R., Graham, M. J., Dekany, R., Smith, R. M., Riddle, R., Masci, F. J., Helou, G., Prince, T. A., Adams, S. M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., ... Zolkower, J. (2019). The Zwicky Transient Facility: System Overview, Performance, and First Results. *PASP*, *131*(995), 018002. https://doi.org/10.1088/1538-3873/aaecbe

Benevento, G., Hu, W., & Raveri, M. (2020). Can late dark energy transitions raise the Hubble constant? *Phys. Rev. D*, *101*(10), Article 103517, 103517. https://doi.org/10.1103/PhysRevD.101.103517

Benson, A. J. (2012). G ALACTICUS: A semi-analytic model of galaxy formation. *New A*, *17*(2), 175–197. https://doi.org/10.1016/j.newast.2011.07.004

Berger, P., & Stein, G. (2019). A volumetric deep Convolutional Neural Network for simulation of mock dark matter halo catalogues. *MNRAS*, *482*, 2861–2871. https://doi.org/10.1093/mnras/sty2949

Bernal, J. L., Verde, L., & Riess, A. G. (2016). The trouble with H_0. *J. Cosmology Astropart. Phys.*, *10*, Article 019, 019. https://doi.org/10.1088/1475-7516/2016/10/019

Bernal, J. L., Verde, L., Jimenez, R., Kamionkowski, M., Valcin, D., & Wandelt, B. D. (2021). Trouble beyond H_0 and the new cosmic triangles. *Phys. Rev. D*, *103*(10), Article 103533, 103533. https://doi.org/10.1103/PhysRevD.103.103533

Bernardini, M., Mayer, L., Reed, D., & Feldmann, R. (2020). Predicting dark matter halo formation in N-body simulations with deep regression networks. *MNRAS*. https://doi.org/10.1093/mnras/staa1911

Betoule, M., Kessler, R., Guy, J., Mosher, J., Hardin, D., Biswas, R., Astier, P., El-Hage, P., Konig, M., Kuhlmann, S., Marriner, J., Pain, R., Regnault, N., Balland, C., Bassett, B. A., Brown, P. J., Campbell, H., Carlberg, R. G., Cellier-Holzem, F., ... Wheeler, C. J. (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *A&A*, *568*, Article A22, A22. https://doi.org/10.1051/0004-6361/201423413

Birrer, S., Shajib, A. J., Galan, A., Millon, M., Treu, T., Agnello, A., Auger, M., Chen, G. C. .-F., Christensen, L., Collett, T., Courbin, F., Fassnacht, C. D., Koopmans, L. V. E., Marshall, P. J., Park, J. .-W., Rusu, C. E., Sluse, D., Spiniello, C., Suyu, S. H., ... Van de Vyvere, L. (2020). TDCOSMO. IV. Hierarchical time-delay cosmography – joint inference of the

Hubble constant and galaxy density profiles. *A&A*, *643*, Article A165, A165. https://doi.org/10.1051/0004-6361/202038861

Birrer, S., Treu, T., Rusu, C. E., Bonvin, V., Fassnacht, C. D., Chan, J. H. H., Agnello, A., Shajib, A. J., Chen, G. C.-F., Auger, M., Courbin, F., Hilbert, S., Sluse, D., Suyu, S. H., Wong, K. C., Marshall, P., Lemaux, B. C., & Meylan, G. (2019). HoLiCOW - IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant. *MNRAS*, *484*, 4726–4753. https://doi.org/10.1093/mnras/stz200

Birrer, S., & Amara, A. (2018). lenstronomy: Multi-purpose gravitational lens modelling software package. *darkU*, *22*, 189–201. https://doi.org/10.1016/j.dark.2018.11.002

Birrer, S., Amara, A., & Refregier, A. (2016). The mass-sheet degeneracy and time-delay cosmography: analysis of the strong lens RXJ1131-1231. *J. Cosmology Astropart. Phys.*, *2016*(8), Article 020, 020. https://doi.org/10.1088/1475-7516/2016/08/020

Birrer, S., Dhawan, S., & Shajib, A. J. (2021). The Hubble constant from strongly lensed supernovae with standardizable magnifications. *arXiv e-prints*, Article arXiv:2107.12385, arXiv:2107.12385.

Blomqvist, M., du Mas des Bourboux, H., Busca, N. G., de Sainte Agathe, V., Rich, J., Balland, C., Bautista, J. E., Dawson, K., Font-Ribera, A., Guy, J., Le Goff, J.-M., Palanque-Delabrouille, N., Percival, W. J., Pérez-Ràfols, I., Pieri, M. M., Schneider, D. P., Slosar, A., & Yèche, C. (2019). Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14. *A&A*, *629*, Article A86, A86. https://doi.org/10.1051/0004-6361/201935641

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight Uncertainty in Neural Networks. *ArXiv e-prints*, Article arXiv:1505.05424.

Bocquet, S. (2015). *Galaxy Cluster Cosmology* (Doctoral dissertation).

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. (2006). The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies. *ApJ*, *638*(2), 703–724. https://doi.org/10.1086/498884

Bond, J. R., & Efstathiou, G. (1984). Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, *285*, L45–L48. https://doi.org/10.1086/184362

Bonvin, V., Millon, M., Chan, J. H. .-H., Courbin, F., Rusu, C. E., Sluse, D., Suyu, S. H., Wong, K. C., Fassnacht, C. D., Marshall, P. J., Treu, T., Buckley-Geer, E., Frieman, J., Hempel, A., Kim, S., Lachaume, R., Rabus, M., Chao, D. C. .-Y., Chijani, M., ... Meylan, G. (2019). COSMOGRAIL. XVIII. time delays of the quadruply lensed quasar WFI2033-4723. *A&A*, *629*, Article A97, A97. https://doi.org/10.1051/0004-6361/201935921

Borgani, S. (2008). Cosmology with Clusters of Galaxies. In M. Plionis, O. López-Cruz, & D. Hughes (Eds.), *A pan-chromatic view of clusters of galaxies and the large-scale structure* (p. 24). https://doi.org/10.1007/978-1-4020-6941-3_9

Breuval, L., Kervella, P., Anderson, R. I., Riess, A. G., Arenou, F., Trahin, B., Mérand, A., Gallenne, A., Gieren, W., Storm, J., Bono, G., Pietrzyński, G., Nardetto, N., Javanmardi, B., & Hocdé, V. (2020). The Milky Way Cepheid Leavitt law based on Gaia DR2 parallaxes of companion stars and host open cluster populations. *A&A*, *643*, Article A115, A115. https://doi.org/10.1051/0004-6361/202038633

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Busti, V. C., de la Cruz-Dombriz, Á., Dunsby, P. K. S., & Sáez-Gómez, D. (2015). Is cosmography a useful tool for testing cosmology? *Phys. Rev. D*, *92*(12), Article 123512, 123512. https://doi.org/10.1103/PhysRevD.92.123512

Calderon, V. F., & Berlind, A. A. (2019). Prediction of galaxy halo masses in SDSS DR7 via a machine learning approach. *MNRAS*, *490*(2), 2367–2379. https://doi.org/10.1093/mnras/stz2775

Camarena, D., & Marra, V. (2020a). A new method to build the (inverse) distance ladder. *MNRAS*, *495*(3), 2630–2644. https://doi.org/10.1093/mnras/staa770

Camarena, D., & Marra, V. (2020b). Local determination of the Hubble constant and the deceleration parameter. *Physical Review Research*, *2*(1), Article 013028, 013028. https://doi.org/10.1103/PhysRevResearch.2.013028

Camarena, D., & Marra, V. (2021). On the use of the local prior on the absolute magnitude of Type Ia supernovae in cosmological inference. *MNRAS*, *504*(4), 5164–5171. https://doi.org/10.1093/mnras/stab1200

Cañameras, R., Schuldt, S., Suyu, S. H., Taubenberger, S., Meinhardt, T., Leal-Taixé, L., Lemon, C., Rojas, K., & Savary, E. (2020). HOLISMOKES. II. Identifying galaxy-scale strong gravitational lenses in Pan-STARRS using convolutional neural networks. *A&A*, *644*, Article A163, A163. https://doi.org/10.1051/0004-6361/202038219

Cano, Z., Selsing, J., Hjorth, J., de Ugarte Postigo, A., Christensen, L., Gall, C., & Kann, D. A. (2018). A spectroscopic look at the gravitationally lensed Type Ia supernova 2016geu at z = 0.409. *MNRAS*, *473*(3), 4257–4267. https://doi.org/10.1093/mnras/stx2624

Caon, N., Capaccioli, M., & D'Onofrio, M. (1993). On the shape of the light profiles of early-type galaxies. *MNRAS*, *265*, 1013–1021. https://doi.org/10.1093/mnras/265.4.1013

Capozziello, S., D'Agostino, R., & Luongo, O. (2020). High-redshift cosmography: auxiliary variables versus Padé polynomials. *MNRAS*, *494*(2), 2576–2590. https://doi.org/10.1093/mnras/staa871

Carneiro, S., de Holanda, P. C., Pigozzo, C., & Sobreira, F. (2019). Is the H_0 tension suggesting a fourth neutrino generation? *Phys. Rev. D*, *100*(2), Article 023505, 023505. https://doi.org/10.1103/PhysRevD.100.023505

Carter, P., Beutler, F., Percival, W. J., Blake, C., Koda, J., & Ross, A. J. (2018). Low redshift baryon acoustic oscillation measurement from the reconstructed 6-degree field galaxy survey. *MNRAS*, *481*(2), 2371–2383. https://doi.org/10.1093/mnras/sty2405

Carter, P., Beutler, F., Percival, W. J., DeRose, J., Wechsler, R. H., & Zhao, C. (2020). The impact of the fiducial cosmology assumption on BAO distance scale measurements. *MNRAS*, *494*(2), 2076–2089. https://doi.org/10.1093/mnras/staa761

Cattoën, C., & Visser, M. (2007). The Hubble series: convergence properties and redshift variables. *Classical and Quantum Gravity*, *24*(23), 5985–5997. https://doi.org/10.1088/0264-9381/24/23/018

Chandrasekhar, S. (1931). The Maximum Mass of Ideal White Dwarfs. *ApJ*, *74*, 81. https://doi.org/10.1086/143324

Chao, D. C. .-Y., Chan, J. H. .-H., Suyu, S. H., Yasuda, N., More, A., Oguri, M., Morokuma, T., & Jaelani, A. T. (2020). Lensed quasar search via time variability with the HSC transient survey. *A&A*, *640*, Article A88, A88. https://doi.org/10.1051/0004-6361/201936806

Charnock, T., Lavaux, G., & Wandelt, B. D. (2018). Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, *97*(8), Article 083004, 083004. https://doi.org/10.1103/PhysRevD.97.083004

Charnock, T., & Moss, A. (2017). Deep Recurrent Neural Networks for Supernovae Classification. *ApJ*, *837*(2), Article L28, L28. https://doi.org/10.3847/2041-8213/aa603d

Charnock, T., Perreault-Levasseur, L., & Lanusse, F. (2020). Bayesian Neural Networks. *ArXiv e-prints*.

Chen, G. C. .-F., Fassnacht, C. D., Suyu, S. H., Rusu, C. E., Chan, J. H. H., Wong, K. C., Auger, M. W., Hilbert, S., Bonvin, V., Birrer, S., Millon, M., Koopmans, L. V. E., Lagattuta, D. J., McKean, J. P., Vegetti, S., Courbin, F., Ding, X., Halkola, A., Jee, I., ... Treu, T. (2019). A SHARP view of HoLiCOW: H_0 from three time-delay gravitational lens systems with adaptive optics imaging. *MNRAS*, *490*(2), 1743–1773. https://doi.org/10.1093/mnras/stz2547

Cheng, T.-Y., Li, N., Conselice, C. J., Aragón-Salamanca, A., Dye, S., & Metcalf, R. B. (2020). Identifying strong lenses with unsupervised machine learning using convolutional autoencoder. *MNRAS*, *494*(3), 3750–3765. https://doi.org/10.1093/mnras/staa1015

Chiba, T., & Nakamura, T. (1998). The Luminosity Distance, the Equation of State, and the Geometry of the Universe. *Progress of Theoretical Physics*, *100*(5), 1077–1082. https://doi.org/10.1143/PTP.100.1077

Choi, Y.-Y., Park, C., & Vogeley, M. S. (2007). Internal and Collective Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, *658*(2), 884–897. https://doi.org/10.1086/511060

Chollet, F. et al. (2015). Keras.

Coelho, R. C. V., Calvão, M. O., Reis, R. R. R., & Siffert, B. B. (2015). Standardization of type Ia supernovae. *European Journal of Physics*, *36*(1), Article 015007, 015007. https://doi.org/10.1088/0143-0807/36/1/015007

Cohn, J. D., & Battaglia, N. (2020). Multiwavelength cluster mass estimates and machine learning. *MNRAS*, *491*(2), 1575–1584. https://doi.org/10.1093/mnras/stz3087

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., ... Taylor, K. (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, *362*(2), 505–534. https://doi.org/10.1111/j.1365-2966.2005.09318.x

Collett, T., Montanari, F., & Räsänen, S. (2019). Model-Independent Determination of H_0 and Ω_{K0} from Strong Lensing and Type Ia Supernovae. *Phys. Rev. Lett.*, *123*(23), Article 231101, 231101. https://doi.org/10.1103/PhysRevLett.123.231101

Collett, T. E. (2015). The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, *811*(1), Article 20, 20. https://doi.org/10.1088/0004-637X/811/1/20

Connolly, A. J., Peterson, J., Jernigan, J. G., Abel, R., Bankert, J., Chang, C., Claver, C. F., Gibson, R., Gilmore, D. K., Grace, E., Jones, R. L., Ivezic, Z., Jee, J., Juric, M., Kahn, S. M., Krabbendam, V. L., Krughoff, S., Lorenz, S., Pizagno, J., . . . Young, M. (2010). Simulating the LSST system. In G. Z. Angeli & P. Dierickx (Eds.), *Modeling, systems engineering, and project management for astronomy iv* (77381O). https://doi.org/10.1117/12.857819

Cora, S. A. (2006). Metal enrichment of the intracluster medium: a three-dimensional picture of chemical and dynamical properties. *MNRAS*, *368*(4), 1540–1560. https://doi.org/10.1111/j.1365-2966.2006.10271.x

Cora, S. A., Vega-Martínez, C. A., Hough, T., Ruiz, A. N., Orsi, Á. A., Muñoz Arancibia, A. M., Gargiulo, I. D., Collacchioni, F., Padilla, N. D., Gottlöber, S., & Yepes, G. (2018). Semi-analytic galaxies - I. Synthesis of environmental and star-forming regulation mechanisms. *MNRAS*, *479*(1), 2–24. https://doi.org/10.1093/mnras/sty1131

Cranmer, K., Brehmer, J., & Louppe, G. (2019). The frontier of simulation-based inference. *ArXiv e-prints*, Article arXiv:1911.01429.

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., & Yoshida, N. (2006). The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, *365*(1), 11–28. https://doi.org/10.1111/j.1365-2966.2005.09675.x

Cuceu, A., Farr, J., Lemos, P., & Font-Ribera, A. (2019). Baryon Acoustic Oscillations and the Hubble constant: past, present and future. *J. Cosmology Astropart. Phys.*, *2019*(10), Article 044, 044. https://doi.org/10.1088/1475-7516/2019/10/044

Dahle, H. (2006). The Cluster Mass Function from Weak Gravitational Lensing. *ApJ*, *653*(2), 954–962. https://doi.org/10.1086/508654

Davies, A., Serjeant, S., & Bromley, J. M. (2019). Using convolutional neural networks to identify gravitational lenses in astronomical images. *MNRAS*, *487*(4), 5263–5271. https://doi.org/10.1093/mnras/stz1288

Davis, T. M., Hinton, S. R., Howlett, C., & Calcino, J. (2019). Can redshift errors bias measurements of the Hubble Constant? *MNRAS*, *490*(2), 2948–2957. https://doi.org/10.1093/mnras/stz2652

Dayal, P., & Ferrara, A. (2018). Early galaxy formation and its large-scale effects. *Phys. Rep.*, *780*, 1–64. https://doi.org/10.1016/j.physrep.2018.10.002

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. (2013). The Kilo-Degree Survey. *Experimental Astronomy*, *35*(1-2), 25–44. https://doi.org/10.1007/s10686-012-9306-1

de Sainte Agathe, V., Balland, C., du Mas des Bourboux, H., Busca, N. G., Blomqvist, M., Guy, J., Rich, J., Font-Ribera, A., Pieri, M. M., Bautista, J. E., Dawson, K., Le Goff, J.-M., de la Macorra, A., Palanque-Delabrouille, N., Percival, W. J., Pérez-Ràfols, I., Schneider, D. P., Slosar, A., & Yèche, C. (2019). Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14. *A&A*, *629*, Article A85, A85. https://doi.org/10.1051/0004-6361/201935638

Denissenya, M., Linder, E. V., & Shafieloo, A. (2018). Cosmic curvature tested directly from observations. *J. Cosmology Astropart. Phys.*, *3*, Article 041, 041. https://doi.org/10.1088/1475-7516/2018/03/041

DES Collaboration, Abbott, T. M. C., Aguena, M., Alarcon, A., Allam, S., Alves, O., Amon, A., Andrade-Oliveira, F., Annis, J., Avila, S., Bacon, D., Baxter, E., Bechtol, K., Becker, M. R., Bernstein, G. M., Bhargava, S., Birrer, S., Blazek, J., Brandao-Souza, A., ... Zuntz, J. (2021). Dark Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *arXiv e-prints*, Article arXiv:2105.13549, arXiv:2105.13549.

DESI Collaboration, Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Allende Prieto, C., Annis, J., Bailey, S., Balland, C., Ballester, O., Baltay, C., Beaufore, L., Bebek, C., Beers, T. C., Bell, E. F., Bernal, J. L., Besuner, R., Beutler, F., ... Zu, Y. (2016). The DESI Experiment Part I: Science,Targeting, and Survey Design. *ArXiv e-prints*, Article arXiv:1611.00036.

Desmond, H., Jain, B., & Sakstein, J. (2019). Local resolution of the Hubble tension: The impact of screened fifth forces on the cosmic distance ladder. *Phys. Rev. D*, *100*(4), Article 043537, 043537. https://doi.org/10.1103/PhysRevD.100.043537

Di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., Mota, D. F., Riess, A. G., & Silk, J. (2021). In the realm of the Hubble tension-a review of solutions. *Classical and Quantum Gravity*, *38*(15), Article 153001, 153001. https://doi.org/10.1088/1361-6382/ac086d

Diaferio, A. (2006). Sunyaev-Zeldovich Effect in Galaxy Clusters . *Memorie della Societa Astronomica Italiana Supplementi*, *9*, 398.

Diaferio, A. (1999). Mass estimation in the outer regions of galaxy clusters. *MNRAS*, *309*(3), 610–622. https://doi.org/10.1046/j.1365-8711.1999.02864.x

Diaferio, A., & Geller, M. J. (1997). Infall Regions of Galaxy Clusters. *ApJ*, *481*(2), 633–643. https://doi.org/10.1086/304075

Diemer, B., More, S., & Kravtsov, A. V. (2013). The Pseudo-evolution of Halo Mass. *ApJ*, *766*(1), Article 25, 25. https://doi.org/10.1088/0004-637X/766/1/25

Diggle, P. J., & Gratton, R. J. (1984). Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(2), 193–212.

Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., & Saurous, R. A. (2017). TensorFlow Distributions. *ArXiv e-prints*.

Ding, X., Liao, K., Birrer, S., Shajib, A. J., Treu, T., & Yang, L. (2021). Improved time-delay lens modelling and H_0 inference with transient sources. *ArXiv e-prints*.

Dobler, G., & Keeton, C. R. (2006). Microlensing of Lensed Supernovae. *ApJ*, *653*(2), 1391–1399. https://doi.org/10.1086/508769

Domínguez, A., Wojtak, R., Finke, J., Ajello, M., Helgason, K., Prada, F., Desai, A., Paliya, V., Marcotulli, L., & Hartmann, D. (2019). A new measurement of the Hubble constant and matter content of the Universe using extragalactic background light γ-ray attenuation. *arXiv e-prints*, Article arXiv:1903.12097, arXiv:1903.12097.

Doran, M., Stern, S., & Thommes, E. (2007). Baryon acoustic oscillations and dynamical dark energy. *J. Cosmology Astropart. Phys.*, *2007*(4), Article 015, 015. https://doi.org/10.1088/1475-7516/2007/04/015

Ducourant, C., Wertz, O., Krone-Martins, A., Teixeira, R., Le Campion, J. .-F., Galluccio, L., Klüter, J., Delchambre, L., Surdej, J., Mignard, F., Wambsganss, J., Bastian, U., Graham, M. J., Djorgovski, S. G., & Slezak, E. (2018). Gaia GraL II - Gaia DR2 Gravitational Lens Systems: The Known Multiply Imaged Quasars. *arXiv e-prints*, Article arXiv:1805.07359, arXiv:1805.07359.

Dunkley, J., Komatsu, E., Nolta, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. (2009). Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data. *ApJS*, *180*, 306–329. https://doi.org/10.1088/0067-0049/180/2/306

Efstathiou, G. (2021). To H_0 or not to H_0? *MNRAS*, *505*(3), 3866–3872. https://doi.org/10.1093/mnras/stab1588

Einstein, A. (1905a). Does the Inertia of a Body Depend Upon its Energy-Content. *Annalen der Physik*, *323*(113), 639–641.

Einstein, A. (1905b). On the electrodynamics of moving bodies [[Annalen Phys.14,194(2005)]]. *Annalen Phys.*

Einstein, A. (1915). The Field Equations of Gravitation.

Einstein, A. (1917). Cosmological Considerations in the General Theory of Relativity. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, *1917*, 142–152.

Eisenstein, D., & White, M. (2004). Theoretical uncertainty in baryon oscillations. *Phys. Rev. D*, *70*(10), Article 103523, 103523. https://doi.org/10.1103/PhysRevD.70.103523

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., … York, D. G. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, *633*(2), 560–574. https://doi.org/10.1086/466512

Faber, S. M., & Gallagher, J. S. (1979). Masses and mass-to-light ratios of galaxies.

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. (1985). On model-dependent bounds on H 0 from gravitational images : application to Q 0957+561 A, B. *ApJ*, *289*, L1–L4. https://doi.org/10.1086/184422

Falco, M., Hansen, S. H., Wojtak, R., Brinckmann, T., Lindholmer, M., & Pandolfi, S. (2014). A new method to measure the mass of galaxy clusters. *MNRAS*, *442*(2), 1887–1896. https://doi.org/10.1093/mnras/stu971

Fitzpatrick, E. L. (1999). Correcting for the Effects of Interstellar Extinction. *PASP*, *111*(755), 63–75. https://doi.org/10.1086/316293

Fixsen, D. J. (2009). The Temperature of the Cosmic Microwave Background. *ApJ*, *707*(2), 916–920. https://doi.org/10.1088/0004-637X/707/2/916

Fluri, J., Lucchi, A., Kacprzak, T., Refregier, A., & Hofmann, T. (2021). Cosmological Parameter Estimation and Inference using Deep Summaries. *arXiv e-prints*, Article arXiv:2107.09002, arXiv:2107.09002.

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, *125*, 306. https://doi.org/10.1086/670067

Foxley-Marrable, M., Collett, T. E., Vernardos, G., Goldstein, D. A., & Bacon, D. (2018). The impact of microlensing on the standardization of strongly lensed Type Ia supernovae. *MNRAS*, *478*(4), 5081–5090. https://doi.org/10.1093/mnras/sty1346

Freedman, W. L. (2021). Measurements of the Hubble Constant: Tensions in Perspective. *arXiv e-prints*, Article arXiv:2106.15656, arXiv:2106.15656.

Freedman, W. L., Madore, B. F., Hatt, D., Hoyt, T. J., Jang, I. S., Beaton, R. L., Burns, C. R., Lee, M. G., Monson, A. J., Neeley, J. R., Phillips, M. M., Rich, J. A., & Seibert, M. (2019). The Carnegie-Chicago Hubble Program. VIII. An Independent Determination of the Hubble Constant Based on the Tip of the Red Giant Branch. *ApJ*, *882*(1), Article 34, 34. https://doi.org/10.3847/1538-4357/ab2f73

Freedman, W. L., Madore, B. F., Hoyt, T., Jang, I. S., Beaton, R., Lee, M. G., Monson, A., Neeley, J., & Rich, J. (2020). Calibration of the Tip of the Red Giant Branch. *ApJ*, *891*(1), Article 57, 57. https://doi.org/10.3847/1538-4357/ab7339

Friedman, A. (1922). On the Curvature of space [[Gen. Rel. Grav.31,1991(1999)]]. *Z. Phys.*, *10*, 377–386. https://doi.org/10.1007/BF01332580

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, *10*, 377–386. https://doi.org/10.1007/BF01332580

Friedmann, A. (1924). Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift fur Physik*, *21*, 326–332. https://doi.org/10.1007/BF01328280

Fumagalli, J. (2018). *Running in the early Universe: UV sensitivity of single-field inflationary models* (Doctoral dissertation). U. Amsterdam, IHEF.

Gal, Y., & Ghahramani, Z. (2015a). Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. *ArXiv e-prints*.

Gal, Y., & Ghahramani, Z. (2015b). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *ArXiv e-prints*.

Gariazzo, S. (2016). *New Developments in Cosmology* (Doctoral dissertation).

Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. (2014). RINGFINDER: Automated Detection of Galaxy-scale Gravitational Lenses in Ground-based Multi-filter Imaging Data. *ApJ*, *785*(2), Article 144, 144. https://doi.org/10.1088/0004-637X/785/2/144

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*, 457–472. https://doi.org/10.1214/ss/1177011136

Gelmini, G. B., Kusenko, A., & Takhistov, V. (2019). Hints of Sterile Neutrinos in Recent Measurements of the Hubble Parameter. *arXiv e-prints*, Article arXiv:1906.10136, arXiv:1906.10136.

Gentile, F., Tortora, C., Covone, G., Koopmans, L. V. E., Spiniello, C., Fan, Z., Li, R., Liu, D., Napolitano, N. R., Vaccari, M., & Fu, L. (2021). Lenses In VoicE (LIVE): Searching for strong gravitational lenses in the VOICE@VST survey using Convolutional Neural Networks. *ArXiv e-prints*.

Germain, M., Gregor, K., Murray, I., & Larochelle, H. (2015). MADE: Masked Autoencoder for Distribution Estimation. *ArXiv e-prints*, Article arXiv:1502.03509.

Giusarma, E., Reyes Hurtado, M., Villaescusa-Navarro, F., He, S., Ho, S., & Hahn, C. (2019). Learning neutrino effects in Cosmology with Convolutional Neural Networks. *ArXiv e-prints*.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In G. Gordon, D. Dunson, & M. Dudík (Eds.), *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315–323). PMLR. https://proceedings.mlr.press/v15/glorot11a.html

Goldstein, D. A., & Nugent, P. E. (2017). How to Find Gravitationally Lensed Type Ia Supernovae. *ApJ*, *834*(1), Article L5, L5. https://doi.org/10.3847/2041-8213/834/1/L5

Goldstein, D. A., Nugent, P. E., & Goobar, A. (2019). Rates and Properties of Supernovae Strongly Gravitationally Lensed by Elliptical Galaxies in Time-domain Imaging Surveys. *ApJS*, *243*(1), Article 6, 6. https://doi.org/10.3847/1538-4365/ab1fe0

Goldstein, D. A., Nugent, P. E., Kasen, D. N., & Collett, T. E. (2018). Precise Time Delays from Strongly Gravitationally Lensed Type Ia Supernovae with Chromatically Microlensed Images. *ApJ*, *855*(1), Article 22, 22. https://doi.org/10.3847/1538-4357/aaa975

Goobar, A., Amanullah, R., Kulkarni, S. R., Nugent, P. E., Johansson, J., Steidel, C., Law, D., Mörtsell, E., Quimby, R., Blagorodnova, N., Brandeker, A., Cao, Y., Cooray, A., Ferretti, R., Fremling, C., Hangard, L., Kasliwal, M., Kupfer, T., Lunnan, R., . . . Yaron, O. (2017). iPTF16geu: A multiply imaged, gravitationally lensed type Ia supernova. *Science*, *356*(6335), 291–295. https://doi.org/10.1126/science.aal2729

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodman, J., & Weare, J. (2010a). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010*, *5*, 65–80. https://doi.org/10.2140/camcos.2010.5.65

Goodman, J., & Weare, J. (2010b). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Gorenstein, M. V., Falco, E. E., & Shapiro, I. I. (1988). Degeneracies in Parameter Estimates for Models of Gravitational Lens Systems. *ApJ*, *327*, 693. https://doi.org/10.1086/166226

Grillo, C., Rosati, P., Suyu, S. H., Balestra, I., Caminha, G. B., Halkola, A., Kelly, P. L., Lombardi, M., Mercurio, A., Rodney, S. A., & Treu, T. (2018). Measuring the Value of the Hubble Constant "à la Refsdal". *ApJ*, *860*(2), Article 94, 94. https://doi.org/10.3847/1538-4357/aac2c9

Grossi, M., & Springel, V. (2009). The impact of early dark energy on non-linear structure formation. *MNRAS*, *394*(3), 1559–1574. https://doi.org/10.1111/j.1365-2966.2009.14432.x

Gunn, J. E., & Gott, I., J. Richard. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, *176*, 1. https://doi.org/10.1086/151605

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, *23*(2), 347–356. https://doi.org/10.1103/PhysRevD.23.347

Guy, J., Astier, P., Baumont, S., Hardin, D., Pain, R., Regnault, N., Basa, S., Carlberg, R. G., Conley, A., Fabbro, S., Fouchez, D., Hook, I. M., Howell, D. A., Perrett, K., Pritchet, C. J., Rich, J., Sullivan, M., Antilogus, P., Aubourg, E., . . . Ruhlmann-Kleider, V. (2007).

SALT2: using distant supernovae to improve the use of type Ia supernovae as distance indicators. *A&A*, *466*(1), 11–21. https://doi.org/10.1051/0004-6361:20066930

Guy, J., Sullivan, M., Conley, A., Regnault, N., Astier, P., Balland, C., Basa, S., Carlberg, R. G., Fouchez, D., Hardin, D., Hook, I. M., Howell, D. A., Pain, R., Palanque-Delabrouille, N., Perrett, K. M., Pritchet, C. J., Rich, J., Ruhlmann-Kleider, V., Balam, D., . . . Walker, E. S. (2010). The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *A&A*, *523*, Article A7, A7. https://doi.org/10.1051/0004-6361/201014468

Hakobyan, A. A., Adibekyan, V. Z., Aramyan, L. S., Petrosian, A. R., Gomes, J. M., Mamon, G. A., Kunth, D., & Turatto, M. (2012). Supernovae and their host galaxies. I. The SDSS DR8 database and statistics. *A&A*, *544*, Article A81, A81. https://doi.org/10.1051/0004-6361/201219541

Han, D.-H., Park, C., Choi, Y.-Y., & Park, M.-G. (2010). The Properties of Type Ia Supernova Host Galaxies from the Sloan Digital Sky Survey. *ApJ*, *724*(1), 502–510. https://doi.org/10.1088/0004-637X/724/1/502

Hannestad, S. (2003). Neutrino masses and the number of neutrino species from WMAP and 2dFGRS. *J. Cosmology Astropart. Phys.*, *2003*(5), Article 004, 004. https://doi.org/10.1088/1475-7516/2003/05/004

Hanson, K. (2001). Markov chain monte carlo posterior sampling with the hamiltonian method. *Proc SPIE*, *4322*, 456–467. https://doi.org/10.1117/12.431119

Harmark, T. (2021). General Relativity and Cosmology lecture notes.

Hastings, W. K. (1970). Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

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

Heavens, A., Jimenez, R., & Verde, L. (2014). Standard Rulers, Candles, and Clocks from the Low-Redshift Universe. *Physical Review Letters*, *113*(24), Article 241302, 241302. https://doi.org/10.1103/PhysRevLett.113.241302

Heavens, A. (2009). Statistical techniques in cosmology. *arXiv e-prints*.

Heymans, C., Tröster, T., Asgari, M., Blake, C., Hildebrandt, H., Joachimi, B., Kuijken, K., Lin, C.-A., Sánchez, A. G., van den Busch, J. L., Wright, A. H., Amon, A., Bilicki, M., de Jong, J., Crocce, M., Dvornik, A., Erben, T., Fortuna, M. C., Getman, F., . . . Wolf, C. (2021). KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *A&A*, *646*, Article A140, A140. https://doi.org/10.1051/0004-6361/202039063

Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. (2017). Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, *548*(7669), 555–557. https://doi.org/10.1038/nature23463

Hicken, M., Challis, P., Jha, S., Kirshner, R. P., Matheson, T., Modjaz, M., Rest, A., Wood-Vasey, W. M., Bakos, G., Barton, E. J., Berlind, P., Bragg, A., Briceño, C., Brown, W. R., Caldwell, N., Calkins, M., Cho, R., Ciupik, L., Contreras, M., . . . Zhao, P. (2009). CfA3: 185

Type Ia Supernova Light Curves from the CfA. *ApJ*, *700*(1), 331–357. https://doi.org/10.1088/0004-637X/700/1/331

Hill, J. C., McDonough, E., Toomey, M. W., & Alexander, S. (2020). Early dark energy does not restore cosmological concordance. *Phys. Rev. D*, *102*(4), Article 043507, 043507. https://doi.org/10.1103/PhysRevD.102.043507

Hilton, M., Sifón, C., Naess, S., Madhavacheril, M., Oguri, M., Rozo, E., Rykoff, E., Abbott, T. M. C., Adhikari, S., Aguena, M., Aiola, S., Allam, S., Amodeo, S., Amon, A., Annis, J., Ansarinejad, B., Aros-Bunster, C., Austermann, J. E., Avila, S., ... Zhang, Y. (2021). The Atacama Cosmology Telescope: A Catalog of >4000 Sunyaev–Zel'dovich Galaxy Clusters. *ApJS*, *253*(1), Article 3, 3. https://doi.org/10.3847/1538-4365/abd023

Ho, M., Farahi, A., Rau, M. M., & Trac, H. (2020). Approximate Bayesian Uncertainties on Deep Learning Dynamical Mass Estimates of Galaxy Clusters. *ArXiv e-prints*, Article arXiv:2006.13231.

Ho, M., Rau, M. M., Ntampaka, M., Farahi, A., Trac, H., & Póczos, B. (2019). A Robust and Efficient Deep Learning Method for Dynamical Mass Measurements of Galaxy Clusters. *ApJ*, *887*(1), Article 25, 25. https://doi.org/10.3847/1538-4357/ab4f82

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hoekstra, H., Bartelmann, M., Dahle, H., Israel, H., Limousin, M., & Meneghetti, M. (2013). Masses of Galaxy Clusters from Gravitational Lensing. *Space Sci. Rev.*, *177*(1-4), 75–118. https://doi.org/10.1007/s11214-013-9978-5

Hortúa, H. J., Volpi, R., Marinelli, D., & Malagò, L. (2020). Parameter estimation for the cosmic microwave background with Bayesian neural networks. *Phys. Rev. D*, *102*(10), Article 103509, 103509. https://doi.org/10.1103/PhysRevD.102.103509

Hotokezaka, K., Nakar, E., Gottlieb, O., Nissanke, S., Masuda, K., Hallinan, G., Mooley, K. P., & Deller, A. T. (2019). A Hubble constant measurement from superluminal motion of the jet in GW170817. *Nature Astronomy*, *3*, 940–944. https://doi.org/10.1038/s41550-019-0820-1

Houlsby, N., Huszár, F., Ghahramani, Z., & Lengyel, M. (2011). Bayesian Active Learning for Classification and Preference Learning. *ArXiv e-prints*.

Hu, W. (2001). Intermediate level CMB tutorial.

Hu, W., & Sugiyama, N. (1996). Small-Scale Cosmological Perturbations: an Analytic Approach. *ApJ*, *471*, 542. https://doi.org/10.1086/177989

Hu, W., & White, M. (1996). Acoustic Signatures in the Cosmic Microwave Background. *ApJ*, *471*, 30. https://doi.org/10.1086/177951

Huang, C. D., Riess, A. G., Yuan, W., Macri, L. M., Zakamska, N. L., Casertano, S., Whitelock, P. A., Hoffmann, S. L., Filippenko, A. V., & Scolnic, D. (2020). Hubble Space Telescope Observations of Mira Variables in the SN Ia Host NGC 1559: An Alternative Candle to Measure the Hubble Constant. *ApJ*, *889*(1), Article 5, 5. https://doi.org/10.3847/1538-4357/ab5dbd

Huang, C.-W., Krueger, D., Lacoste, A., & Courville, A. (2018). Neural Autoregressive Flows. *ArXiv e-prints*, Article arXiv:1804.00779.

Huang, N., Bleem, L. E., Stalder, B., Ade, P. A. R., Allen, S. W., Anderson, A. J., Austermann, J. E., Avva, J. S., Beall, J. A., Bender, A. N., Benson, B. A., Bianchini, F., Bocquet, S., Brodwin, M., Carlstrom, J. E., Chang, C. L., Chiang, H. C., Citron, R., Moran, C. C., ... Yefremenko, V. (2020). Galaxy Clusters Selected via the Sunyaev-Zel'dovich Effect in the SPTpol 100-square-degree Survey. *AJ*, *159*(3), Article 110, 110. https://doi.org/10.3847/1538-3881/ab6a96

Huang, X., Storfer, C., Gu, A., Ravi, V., Pilon, A., Sheu, W., Venguswamy, R., Banka, S., Dey, A., Landriau, M., Lang, D., Meisner, A., Moustakas, J., Myers, A. D., Sajith, R., Schlafly, E. F., & Schlegel, D. J. (2021). Discovering New Strong Gravitational Lenses in the DESI Legacy Imaging Surveys. *ApJ*, *909*(1), Article 27, 27. https://doi.org/10.3847/1538-4357/abd62b

Huang, X., Storfer, C., Ravi, V., Pilon, A., Domingo, M., Schlegel, D. J., Bailey, S., Dey, A., Gupta, R. R., Herrera, D., Juneau, S., Landriau, M., Lang, D., Meisner, A., Moustakas, J., Myers, A. D., Schlafly, E. F., Valdes, F., Weaver, B. A., ... Yèche, C. (2020). Finding Strong Gravitational Lenses in the DESI DECam Legacy Survey. *ApJ*, *894*(1), Article 78, 78. https://doi.org/10.3847/1538-4357/ab7ffb

Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae.

Huber, S., Suyu, S. H., Ghoshdastidar, D., Taubenberger, S., Bonvin, V., Chan, J. H. H., Kromer, M., Noebauer, U. M., Sim, S. A., & Leal-Taixé, L. (2021). HOLISMOKES - VII. Time-delay measurement of strongly lensed SNe Ia using machine learning. *ArXiv e-prints*.

Huber, S., Suyu, S. H., Noebauer, U. M., Bonvin, V., Rothchild, D., Chan, J. H. H., Awan, H., Courbin, F., Kromer, M., Marshall, P., Oguri, M., Ribeiro, T., & LSST Dark Energy Science Collaboration. (2019). Strongly lensed SNe Ia in the era of LSST: observing cadence for lens discoveries and time-delay measurements. *A&A*, *631*, Article A161, A161. https://doi.org/10.1051/0004-6361/201935370

Huber, S., Suyu, S. H., Noebauer, U. M., Chan, J. H. H., Kromer, M., Sim, S. A., Sluse, D., & Taubenberger, S. (2021). HOLISMOKES. III. Achromatic phase of strongly lensed Type Ia supernovae. *A&A*, *646*, Article A110, A110. https://doi.org/10.1051/0004-6361/202039218

Humphrey, P. J., & Buote, D. A. (2010). The slope of the mass profile and the tilt of the Fundamental Plane in early-type galaxies. *MNRAS*, *403*(4), 2143–2151. https://doi.org/10.1111/j.1365-2966.2010.16257.x

Hyde, J. B., & Bernardi, M. (2009). Curvature in the scaling relations of early-type galaxies. *MNRAS*, *394*(4), 1978–1990. https://doi.org/10.1111/j.1365-2966.2009.14445.x

Iben, J., I., & Tutukov, A. V. (1984). Supernovae of type I as end products of the evolution of binaries with components of moderate initial mass. *ApJS*, *54*, 335–372. https://doi.org/10.1086/190932

Ichikawa, K., Sekiguchi, T., & Takahashi, T. (2008). Probing the effective number of neutrino species with the cosmic microwave background. *Phys. Rev. D*, *78*(8), Article 083526, 083526. https://doi.org/10.1103/PhysRevD.78.083526

Ishiyama, T., Prada, F., Klypin, A. A., Sinha, M., Metcalf, R. B., Jullo, E., Altieri, B., Cora, S. A., Croton, D., de la Torre, S., Millán-Calero, D. E., Oogi, T., Ruedas, J., & Vega-Martínez,

C. A. (2020). The Uchuu Simulations: Data Release 1 and Dark Matter Halo Concentrations. *ArXiv e-prints*, Article arXiv:2007.14720.

Ivanov, M. M., Ali-Haïmoud, Y., & Lesgourgues, J. (2020). H_0 tension or T_0 tension? *Phys. Rev. D*, *102*(6), Article 063515, 063515. https://doi.org/10.1103/PhysRevD.102.063515

Ivezic, Z., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, R., Angeli, G., Ansari, R., Antilogus, P., Arndt, K. T., Astier, P., Aubourg, E., Axelrod, T., Bard, D. J., Barr, J. D., Barrau, A., ... for the LSST Collaboration. (2008). LSST: from Science Drivers to Reference Design and Anticipated Data Products. *ArXiv e-prints*.

Jackson, J. C. (1972). A critique of Rees's theory of primordial gravitational radiation. *MNRAS*, *156*, 1P. https://doi.org/10.1093/mnras/156.1.1P

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review*, *106*(4), 620.

Jeans, J. H. (1902). The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, *199*, 1–53. https://doi.org/10.1098/rsta.1902.0012

Jedamzik, K., Pogosian, L., & Zhao, G.-B. (2021). Why reducing the cosmic sound horizon alone can not fully resolve the Hubble tension. *Communications Physics*, *4*(1), Article 123, 123. https://doi.org/10.1038/s42005-021-00628-x

Jee, I., Komatsu, E., Suyu, S. H., & Huterer, D. (2016). Time-delay cosmography: increased leverage with angular diameter distances. *J. Cosmology Astropart. Phys.*, *2016*(4), Article 031, 031. https://doi.org/10.1088/1475-7516/2016/04/031

Jee, I., Suyu, S. H., Komatsu, E., Fassnacht, C. D., Hilbert, S., & Koopmans, L. V. E. (2019). A measurement of the Hubble constant from angular diameter distances to two gravitational lenses. *Science*, *365*(6458), 1134–1138. https://doi.org/10.1126/science.aat7371

Jennings, E., & Madigan, M. (2017). astroABC : An Approximate Bayesian Computation Sequential Monte Carlo sampler for cosmological parameter estimation. *Astronomy and Computing*, *19*, 16–22. https://doi.org/10.1016/j.ascom.2017.01.001

Jimenez, R., Cimatti, A., Verde, L., Moresco, M., & Wandelt, B. (2019). The local and distant Universe: stellar ages and H_0. *J. Cosmology Astropart. Phys.*, *2019*(3), Article 043, 043. https://doi.org/10.1088/1475-7516/2019/03/043

Jimenez Rezende, D., & Mohamed, S. (2015). Variational Inference with Normalizing Flows. *ArXiv e-prints*, Article arXiv:1505.05770.

Jones, D. O., Foley, R. J., Narayan, G., Hjorth, J., Huber, M. E., Aleo, P. D., Alexander, K. D., Angus, C. R., Auchettl, K., Baldassare, V. F., Bruun, S. H., Chambers, K. C., Chatterjee, D., Coppejans, D. L., Coulter, D. A., DeMarchi, L., Dimitriadis, G., Drout, M. R., Engel, A., ... Young Supernova Experiment. (2021). The Young Supernova Experiment: Survey Goals, Overview, and Operations. *ApJ*, *908*(2), Article 143, 143. https://doi.org/10.3847/1538-4357/abd7f5

Jones, R. L., Yoachim, P., Ivezic, Z., Neilsen, E. H., & Ribeiro, T. (2020). Survey Strategy and Cadence Choices for the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST). https://doi.org/10.5281/zenodo.4048838

Kaiser, N. (1987). Clustering in real space and in redshift space. *MNRAS*, *227*, 1–21. https://doi.org/10.1093/mnras/227.1.1

Karwal, T., & Kamionkowski, M. (2016). Dark energy at early times, the Hubble parameter, and the string axiverse. *Physical Review D*, *94*(10), Article 103523, 103523. https://doi.org/10.1103/PhysRevD.94.103523

Kazin, E. A., Koda, J., Blake, C., Padmanabhan, N., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Croton, D. J., Davis, T. M., Drinkwater, M. J., Forster, K., Gilbank, D., Gladders, M., Glazebrook, K., Jelliffe, B., Jurek, R. J., Li, I. .-h., . . . Yee, H. K. C. (2014). The WiggleZ Dark Energy Survey: improved distance measurements to z = 1 with reconstruction of the baryonic acoustic feature. *MNRAS*, *441*(4), 3524–3542. https://doi.org/10.1093/mnras/stu778

Kelly, P. L., Rodney, S. A., Treu, T., Foley, R. J., Brammer, G., Schmidt, K. B., Zitrin, A., Sonnenfeld, A., Strolger, L.-G., Graur, O., Filippenko, A. V., Jha, S. W., Riess, A. G., Bradac, M., Weiner, B. J., Scolnic, D., Malkan, M. A., von der Linden, A., Trenti, M., . . . Tucker, B. E. (2015). Multiple images of a highly magnified supernova formed by an early-type cluster galaxy lens. *Science*, *347*(6226), 1123–1126. https://doi.org/10.1126/science.aaa3350

Kendall, A., & Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *ArXiv e-prints*.

Kenworthy, W. D., Scolnic, D., & Riess, A. (2019). The Local Perspective on the Hubble Tension: Local Structure Does Not Impact Measurement of the Hubble Constant. *ApJ*, *875*(2), Article 145, 145. https://doi.org/10.3847/1538-4357/ab0ebf

Khetan, N., Izzo, L., Branchesi, M., Wojtak, R., Cantiello, M., Murugeshan, C., Agnello, A., Cappellaro, E., Della Valle, M., Gall, C., Hjorth, J., Benetti, S., Brocato, E., Burke, J., Hiramatsu, D., Howell, D. A., Tomasella, L., & Valenti, S. (2021). A new measurement of the Hubble constant using Type Ia supernovae calibrated with surface brightness fluctuations. *A&A*, *647*, Article A72, A72. https://doi.org/10.1051/0004-6361/202039196

Killestein, T. L., Lyman, J., Steeghs, D., Ackley, K., Dyer, M. J., Ulaczyk, K., Cutter, R., Mong, Y. .-L., Galloway, D. K., Dhillon, V., O'Brien, P., Ramsay, G., Poshyachinda, S., Kotak, R., Breton, R. P., Nuttall, L. K., Pallé, E., Pollacco, D., Thrane, E., . . . Williams, S. C. (2021). Transient-optimised real-bogus classification with Bayesian Convolutional Neural Networks – sifting the GOTO candidate stream. *ArXiv e-prints*.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *ArXiv e-prints*.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., & Welling, M. (2016). Improving Variational Inference with Inverse Autoregressive Flow. *ArXiv e-prints*, Article arXiv:1606.04934.

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. (2016). MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, *457*(4), 4340–4359. https://doi.org/10.1093/mnras/stw248

Knebe, A., Stoppacher, D., Prada, F., Behrens, C., Benson, A., Cora, S. A., Croton, D. J., Padilla, N. D., Ruiz, A. N., Sinha, M., Stevens, A. R. H., Vega-Martínez, C. A., Behroozi, P., Gonzalez-Perez, V., Gottlöber, S., Klypin, A. A., Yepes, G., Enke, H., Libeskind, N. I., . . . Steinmetz, M. (2018). MULTIDARK-GALAXIES: data release and first results. *MNRAS*, *474*(4), 5206–5231. https://doi.org/10.1093/mnras/stx2662

Knox, L., & Millea, M. (2020). Hubble constant hunter's guide. *Phys. Rev. D*, *101*(4), Article 043533, 043533. https://doi.org/10.1103/PhysRevD.101.043533

Kochanek, C. S. (2020). Overconstrained gravitational lens models and the Hubble constant. *MNRAS*, *493*(2), 1725–1735. https://doi.org/10.1093/mnras/staa344

Kochanek, C. S., Mochejska, B., Morgan, N. D., & Stanek, K. Z. (2006). A Simple Method to Find All Lensed Quasars. *ApJ*, *637*(2), L73–L76. https://doi.org/10.1086/500559

Kodi Ramanah, D. (2019). *Bayesian statistical inference and deep learning for primordial cosmology and cosmic acceleration* (Theses 2019SORUS169). Sorbonne Université. https://tel.archives-ouvertes.fr/tel-02969186

Kodi Ramanah, D., Arendse, N., & Wojtak, R. (2021). AI-driven spatio-temporal engine for finding gravitationally lensed supernovae. *ArXiv e-prints*.

Kodi Ramanah, D., Charnock, T., & Lavaux, G. (2019). Painting halos from cosmic density fields of dark matter with physically motivated neural networks. *Phys. Rev. D*, *100*(4), Article 043515, 043515. https://doi.org/10.1103/PhysRevD.100.043515

Kodi Ramanah, D., Charnock, T., Villaescusa-Navarro, F., & Wandelt, B. D. (2020). Super-resolution emulator of cosmological simulations using deep physical models. *MNRAS*, *495*(4), 4227–4236. https://doi.org/10.1093/mnras/staa1428

Kodi Ramanah, D., Wojtak, R., Ansari, Z., Gall, C., & Hjorth, J. (2020). Dynamical mass inference of galaxy clusters with neural flows. *MNRAS*, *499*(2), 1985–1997. https://doi.org/10.1093/mnras/staa2886

Kodi Ramanah, D., Wojtak, R., & Arendse, N. (2021). Simulation-based inference of dynamical galaxy cluster masses with 3D convolutional neural networks. *MNRAS*, *501*(3), 4080–4091. https://doi.org/10.1093/mnras/staa3922

Kolatt, T. S., & Bartelmann, M. (1998). Gravitational lensing of type IA supernovae by galaxy clusters. *MNRAS*, *296*(3), 763–772. https://doi.org/10.1046/j.1365-8711.1998.01466.x

Koopmans, L. V. E. (2004). Gravitational Lensing & Stellar Dynamics: Dark-Matter and Baryons in Early-type Galaxies to z=1. *arXiv e-prints*, Article astro-ph/0412596, astro–ph/0412596.

Koopmans, L. V. E., Bolton, A., Treu, T., Czoske, O., Auger, M. W., Barnabè, M., Vegetti, S., Gavazzi, R., Moustakas, L. A., & Burles, S. (2009). The Structure and Dynamics of Massive Early-Type Galaxies: On Homology, Isothermality, and Isotropy Inside One Effective Radius. *ApJ*, *703*(1), L51–L54. https://doi.org/10.1088/0004-637X/703/1/L51

Kormann, R., Schneider, P., & Bartelmann, M. (1994). Isothermal elliptical gravitational lens models. *A&A*, *284*, 285–299.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Lacki, B. C., Kochanek, C. S., Stanek, K. Z., Inada, N., & Oguri, M. (2009). Difference Imaging of Lensed Quasar Candidates in the Sloan Digital Sky Survey Supernova Survey Region. *ApJ*, *698*(1), 428–438. https://doi.org/10.1088/0004-637X/698/1/428

Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C.-L., Ravanbakhsh, S., Mandelbaum, R., & Póczos, B. (2018). CMU DeepLens: deep learning for automatic image-based galaxy-

galaxy strong lens finding. *MNRAS*, *473*(3), 3895–3906. https://doi.org/10.1093/mnras/stx1665
Leavitt, H. S. (1908). 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, *60*, 87–108.3.
Leclercq, F. (2018). Bayesian optimization for likelihood-free cosmological inference. *Phys. Rev. D*, *98*(6), Article 063511, 063511. https://doi.org/10.1103/PhysRevD.98.063511
Leclercq, F., Pisani, A., & Wandelt, B. D. (2014). Cosmology: from theory to data, from data to theory. *arXiv e-prints*.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444. https://doi.org/10.1038/nature14539
LeCun, Y., Bengio, Y. et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.
Lee, H. (2017). *High-Energy Aspects of Inflationary Cosmology (Doctoral thesis)* (Doctoral dissertation). https://doi.org/10.17863/CAM.20391
Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, *47*, 49–59.
Lemaître, G. (1931). Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *MNRAS*, *91*, 483–490. https://doi.org/10.1093/mnras/91.5.483
Lemaître, G. (1933). L'Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, *53*.
Lemos, P., Jeffrey, N., Whiteway, L., Lahav, O., Libeskind, N., & Hoffman, Y. (2021). Sum of the masses of the Milky Way and M31: A likelihood-free inference approach. *Phys. Rev. D*, *103*(2), Article 023009, 023009. https://doi.org/10.1103/PhysRevD.103.023009
L'Huillier, B., & Shafieloo, A. (2017). Model-independent test of the FLRW metric, the flatness of the Universe, and non-local estimation of H_0 r_d. *Journal of Cosmology and Astro-Particle Physics*, *2017*(1), Article 015, 015. https://doi.org/10.1088/1475-7516/2017/01/015
Li, E.-K., Du, M., & Xu, L. (2020). General cosmography model with spatial curvature. *MNRAS*, *491*(4), 4960–4972. https://doi.org/10.1093/mnras/stz3308
Li, X., & Shafieloo, A. (2019). A Simple Phenomenological Emergent Dark Energy Model can Resolve the Hubble Tension. *ApJ*, *883*(1), Article L3, L3. https://doi.org/10.3847/2041-8213/ab3e09
Li, Y., Ni, Y., Croft, R. A. C., Di Matteo, T., Bird, S., & Feng, Y. (2021). AI-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Science*, *118*(19), Article 2022038118, 2022038118. https://doi.org/10.1073/pnas.2022038118
Liao, K., Treu, T., Marshall, P., Fassnacht, C. D., Rumbaugh, N., Dobler, G., Aghamousa, A., Bonvin, V., Courbin, F., Hojjati, A., Jackson, N., Kashyap, V., Rathna Kumar, S., Linder, E., Mandel, K., Meng, X.-L., Meylan, G., Moustakas, L. A., Prabhu, T. P., . . . van Dyk, D. (2015). Strong Lens Time Delay Challenge. II. Results of TDC1. *ApJ*, *800*, Article 11, 11. https://doi.org/10.1088/0004-637X/800/1/11

Lin, Y.-C., & Proty Wu, J.-H. (2020). Detection of Gravitational Waves Using Bayesian Neural Networks. *ArXiv e-prints*.

Linde, A. D. (1982). A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Physics Letters B*, *108*(6), 389–393. https://doi.org/10.1016/0370-2693(82)91219-9

Linder, E. V. (2005). Cosmic growth history and expansion history. *Phys. Rev. D*, *72*(4), Article 043529, 043529. https://doi.org/10.1103/PhysRevD.72.043529

Linder, E. V., & Robbers, G. (2008). Shifting the Universe: early dark energy and standard rulers. *J. Cosmology Astropart. Phys.*, *2008*(6), Article 004, 004. https://doi.org/10.1088/1475-7516/2008/06/004

Lintusaari, J., Vuollekoski, H., Kangasrääsiö, A., Skytén, K., Järvenpää, M., Marttinen, P., Gutmann, M. U., Vehtari, A., Corander, J., & Kaski, S. (2017). ELFI: Engine for Likelihood-Free Inference. *ArXiv e-prints*, Article arXiv:1708.00707.

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *ArXiv e-prints*.

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. (2016). Photometric Supernova Classification with Machine Learning. *ApJS*, *225*(2), Article 31, 31. https://doi.org/10.3847/0067-0049/225/2/31

Łokas, E. L., & Mamon, G. A. (2003). Dark matter distribution in the Coma cluster from galaxy kinematics: breaking the mass-anisotropy degeneracy. *MNRAS*, *343*(2), 401–412. https://doi.org/10.1046/j.1365-8711.2003.06684.x

Loureiro, A., Cuceu, A., Abdalla, F. B., Moraes, B., Whiteway, L., McLeod, M., Balan, S. T., Lahav, O., Benoit-Lévy, A., Manera, M., Rollins, R. P., & Xavier, H. S. (2019). Upper Bound of Neutrino Masses from Combined Cosmological Observations and Particle Physics Experiments. *Phys. Rev. Lett.*, *123*(8), Article 081301, 081301. https://doi.org/10.1103/PhysRevLett.123.081301

Luzzi, G., Génova-Santos, R. T., Martins, C. J. A. P., De Petris, M., & Lamagna, L. (2015). Constraining the evolution of the CMB temperature with SZ measurements from Planck data. *J. Cosmology Astropart. Phys.*, *2015*(9), Article 011, 011. https://doi.org/10.1088/1475-7516/2015/09/011

Macaulay, E., Nichol, R. C., Bacon, D., Brout, D., Davis, T. M., Zhang, B., Bassett, B. A., Scolnic, D., Möller, A., D'Andrea, C. B., Hinton, S. R., Kessler, R., Kim, A. G., Lasker, J., Lidman, C., Sako, M., Smith, M., Sullivan, M., Abbott, T. M. C., . . . Wiseman, P. (2019). First cosmological results using Type Ia supernovae from the Dark Energy Survey: measurement of the Hubble constant. *MNRAS*, *486*, 2184–2196. https://doi.org/10.1093/mnras/stz978

Mack, K. (2020). *The End of Everything (Astrophysically Speaking)*.

MacKay, D. J. (1992). A practical bayesian framework for backpropagation networks. *Neural computation*, *4*(3), 448–472.

Madireddy, S., Li, N., Ramachandra, N., Butler, J., Balaprakash, P., Habib, S., & Heitmann, K. (2019). A Modular Deep Learning Pipeline for Galaxy-Scale Strong Gravitational Lens Detection and Modeling. *ArXiv e-prints*.

Maguire, K. (2017). Type Ia Supernovae. In A. W. Alsabti & P. Murdin (Eds.), *Handbook of supernovae* (p. 293). https://doi.org/10.1007/978-3-319-21846-5_36

Makinen, T. L., Charnock, T., Alsing, J., & Wandelt, B. D. (2021). Lossless, Scalable Implicit Likelihood Inference for Cosmological Fields. *arXiv e-prints*, Article arXiv:2107.07405, arXiv:2107.07405.

Maltoni, F., & Maccio, A. (2010). Advanced simulation techniques in particle physics and cosmology: Linear theory and perturbation growth.

Mancarella, M., Kennedy, J., Bose, B., & Lombriser, L. (2020). Seeking New Physics in Cosmology with Bayesian Neural Networks I: Dark Energy and Modified Gravity. *ArXiv e-prints*.

Mantz, A. B., Allen, S. W., Morris, R. G., Rapetti, D. A., Applegate, D. E., Kelly, P. L., von der Linden, A., & Schmidt, R. W. (2014). Cosmology and astrophysics from relaxed galaxy clusters - II. Cosmological constraints. *MNRAS*, *440*(3), 2077–2098. https://doi.org/10.1093/mnras/stu368

Maresca, J., Dye, S., & Li, N. (2021). Auto-identification of unphysical source reconstructions in strong gravitational lens modelling. *MNRAS*, *503*(2), 2229–2241. https://doi.org/10.1093/mnras/stab387

Maxwell, J. C. (1865). A dynamical theory of the electromagnetic field. *Philosophical Transactions of the Royal Society of London*, *155*, 459–513.

Mcculloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 127–147.

Merloni, A., Predehl, P., Becker, W., Böhringer, H., Boller, T., Brunner, H., Brusa, M., Dennerl, K., Freyberg, M., Friedrich, P., Georgakakis, A., Haberl, F., Hasinger, G., Meidinger, N., Mohr, J., Nandra, K., Rau, A., Reiprich, T. H., Robrade, J., . . . German eROSITA Consortium, t. (2012). eROSITA Science Book: Mapping the Structure of the Energetic Universe. *ArXiv e-prints*, Article arXiv:1209.3114.

Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decencière, E., Flamary, R., Gavazzi, R., Geiger, M., Hartley, P., Huertas-Company, M., Jackson, N., Jacobs, C., Jullo, E., Kneib, J. .-P., . . . Vernardos, G. (2019). The strong gravitational lens finding challenge. *A&A*, *625*, Article A119, A119. https://doi.org/10.1051/0004-6361/201832797

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Millon, M., Galan, A., Courbin, F., Treu, T., Suyu, S. H., Ding, X., Birrer, S., Chen, G. C. .-F., Shajib, A. J., Wong, K. C., Agnello, A., Auger, M. W., Buckley-Geer, E. J., Chan, J. H. H., Collett, T., Fassnacht, C. D., Hilbert, S., Koopmans, L. V. E., Motta, V., . . . Van de Vyvere, L. (2019). TDCOSMO. I. An exploration of systematic uncertainties in the inference of H_0 from time-delay cosmography. *arXiv e-prints*, Article arXiv:1912.08027, arXiv:1912.08027.

Molnar, C. (2019). *Interpretable machine learning: A guide for making black box models explainable* [https://christophm.github.io/interpretable-ml-book/].

Montanari, F., & Durrer, R. (2012). New method for the Alcock-Paczyński test. *Phys. Rev. D*, *86*(6), Article 063503, 063503. https://doi.org/10.1103/PhysRevD.86.063503

More, A., Suyu, S. H., Oguri, M., More, S., & Lee, C.-H. (2017). Interpreting the Strongly Lensed Supernova iPTF16geu: Time Delay Predictions, Microlensing, and Lensing Rates. *ApJ*, *835*(2), Article L25, L25. https://doi.org/10.3847/2041-8213/835/2/L25

Morgan, R., Nord, B., Birrer, S., Lin, J., & Poh, J. (2021). deeplenstronomy: A dataset simulation package for strong gravitational lensing. *The Journal of Open Source Software*, *6*(58), Article 2854, 2854. https://doi.org/10.21105/joss.02854

Mörtsell, E., & Dhawan, S. (2018). Does the Hubble constant tension call for new physics? *J. Cosmology Astropart. Phys.*, *2018*(9), Article 025, 025. https://doi.org/10.1088/1475-7516/2018/09/025

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning*, 807–814. http://dl.acm.org/citation.cfm?id=3104322.3104425

Narayan, R., & Bartelmann, M. (1996). Lectures on Gravitational Lensing. *arXiv e-prints*, Article astro-ph/9606001, astro–ph/9606001.

Neal, R. M. (2012). *Bayesian learning for neural networks* (Vol. 118). Springer Science & Business Media.

Nomoto, K. (1982). Accreting white dwarf models for type I supernovae. I - Presupernova evolution and triggering mechanisms. *ApJ*, *253*, 798–810. https://doi.org/10.1086/159682

Norman, M. L. (2010). Simulating Galaxy Clusters. *arXiv e-prints*, Article arXiv:1005.1100, arXiv:1005.1100.

Ntampaka, M., Trac, H., Sutherland, D. J., Battaglia, N., Póczos, B., & Schneider, J. (2015). A Machine Learning Approach for Dynamical Mass Measurements of Galaxy Clusters. *ApJ*, *803*(2), Article 50, 50. https://doi.org/10.1088/0004-637X/803/2/50

Ntampaka, M., Trac, H., Sutherland, D. J., Fromenteau, S., Póczos, B., & Schneider, J. (2016). Dynamical Mass Measurements of Contaminated Galaxy Clusters Using Machine Learning. *ApJ*, *831*(2), Article 135, 135. https://doi.org/10.3847/0004-637X/831/2/135

Nugent, P., Kim, A., & Perlmutter, S. (2002). K-Corrections and Extinction Corrections for Type Ia Supernovae. *PASP*, *114*(798), 803–819. https://doi.org/10.1086/341707

Oguri, M. (2006). The image separation distribution of strong lenses: halo versus subhalo populations. *MNRAS*, *367*(3), 1241–1250. https://doi.org/10.1111/j.1365-2966.2006.10043.x

Oguri, M. (2010). The Mass Distribution of SDSS J1004+4112 Revisited. *PASJ*, *62*, 1017. https://doi.org/10.1093/pasj/62.4.1017

Oguri, M., & Kawano, Y. (2003). Gravitational lens time delays for distant supernovae: breaking the degeneracy between radial mass profiles and the Hubble constant. *MNRAS*, *338*(4), L25–L29. https://doi.org/10.1046/j.1365-8711.2003.06290.x

Oguri, M., & Marshall, P. J. (2010). Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *MNRAS*, *405*(4), 2579–2593. https://doi.org/10.1111/j.1365-2966.2010.16639.x

Old, L., Wojtak, R., Mamon, G. A., Skibba, R. A., Pearce, F. R., Croton, D., Bamford, S., Behroozi, P., de Carvalho, R., Muñoz-Cuartas, J. C., Gifford, D., Gray, M. E., von der Linden, A., Merrifield, M. R., Muldrew, S. I., Müller, V., Pearson, R. J., Ponman, T. J., Rozo, E., . . . Tempel, E. (2015). Galaxy Cluster Mass Reconstruction Project - II. Quantifying scatter and bias using contrasting mock catalogues. *MNRAS*, *449*(2), 1897–1920. https://doi.org/10.1093/mnras/stv421

Old, L., Wojtak, R., Pearce, F. R., Gray, M. E., Mamon, G. A., Sifón, C., Tempel, E., Biviano, A., Yee, H. K. C., de Carvalho, R., Müller, V., Sepp, T., Skibba, R. A., Croton, D., Bamford, S. P., Power, C., von der Linden, A., & Saro, A. (2018). Galaxy Cluster Mass Reconstruction Project - III. The impact of dynamical substructure on cluster mass estimates. *MNRAS*, *475*(1), 853–866. https://doi.org/10.1093/mnras/stx3241

O'Raifeartaigh, C., & Mitton, S. (2018). Einstein's "biggest blunder" - interrogating the legend. *ArXiv e-prints*.

Padilla, N. D., & Strauss, M. A. (2008). The shapes of galaxies in the Sloan Digital Sky Survey. *MNRAS*, *388*(3), 1321–1334. https://doi.org/10.1111/j.1365-2966.2008.13480.x

Padin, S., Staniszewski, Z., Keisler, R., Joy, M., Stark, A. A., Ade, P. A. R., Aird, K. A., Benson, B. A., Bleem, L. E., Carlstrom, J. E., Chang, C. L., Crawford, T. M., Crites, A. T., Dobbs, M. A., Halverson, N. W., Heimsath, S., Hills, R. E., Holzapfel, W. L., Lawrie, C., . . . Vieira, J. D. (2008). South Pole Telescope optics. *Appl. Opt.*, *47*(24), 4418–4428. https://doi.org/10.1364/AO.47.004418

Papamakarios, G., & Murray, I. (2016). Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *ArXiv e-prints*, Article arXiv:1605.06376.

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked Autoregressive Flow for Density Estimation. *ArXiv e-prints*, Article arXiv:1705.07057.

Papamakarios, G., Sterratt, D. C., & Murray, I. (2018). Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows. *ArXiv e-prints*, Article arXiv:1805.07226.

Paraficz, D., & Hjorth, J. (2009). Gravitational lenses as cosmic rulers: Ω_m, Ω_Λ from time delays and velocity dispersions. *A&A*, *507*(3), L49–L52. https://doi.org/10.1051/0004-6361/200913307

Park, J. W., Wagner-Carena, S., Birrer, S., Marshall, P. J., Lin, J. Y.-Y., Roodman, A., & LSST Dark Energy Science Collaboration. (2021). Large-scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant. *ApJ*, *910*(1), Article 39, 39. https://doi.org/10.3847/1538-4357/abdfc4

Patel, B., McCully, C., Jha, S. W., Rodney, S. A., Jones, D. O., Graur, O., Merten, J., Zitrin, A., Riess, A. G., Matheson, T., Sako, M., Holoien, T. W. .-S., Postman, M., Coe, D., Bartelmann, M., Balestra, I., Benítez, N., Bouwens, R., Bradley, L., . . . Zheng, W. (2014). Three Gravitationally Lensed Supernovae behind CLASH Galaxy Clusters. *ApJ*, *786*(1), Article 9, 9. https://doi.org/10.1088/0004-637X/786/1/9

Peacock, J. (1998). *Cosmological Physics*. Cambridge University Press. https://books.google.nl/books?id=EvPoAwAAQBAJ

Pearson, J., Li, N., & Dye, S. (2019). The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples. *MNRAS*, *488*(1), 991–1004. https://doi.org/10.1093/mnras/stz1750

Pearson, J., Maresca, J., Li, N., & Dye, S. (2021). Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting. *ArXiv e-prints*.

Peebles, P. J. E., & Yu, J. T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, *162*, 815. https://doi.org/10.1086/150713

Penzias, A. A., & Wilson, R. W. (1965a). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, *142*, 419–421. https://doi.org/10.1086/148307

Penzias, A. A., & Wilson, R. W. (1965b). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, *142*, 419–421. https://doi.org/10.1086/148307

Pereira, R., Thomas, R. C., Aldering, G., Antilogus, P., Baltay, C., Benitez-Herrera, S., Bongard, S., Buton, C., Canto, A., Cellier-Holzem, F., Chen, J., Childress, M., Chotard, N., Copin, Y., Fakhouri, H. K., Fink, M., Fouchez, D., Gangler, E., Guy, J., . . . Wu, C. (2013). Spectrophotometric time series of SN 2011fe from the Nearby Supernova Factory. *A&A*, *554*, Article A27, A27. https://doi.org/10.1051/0004-6361/201221008

Perlmutter, S., Aldering, G., della Valle, M., Deustua, S., Ellis, R. S., Fabbro, S., Fruchter, A., Goldhaber, G., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Knop, R. A., Lidman, C., McMahon, R. G., Nugent, P., Pain, R., Panagia, N., Pennypacker, C. R., . . . Walton, N. (1998). Discovery of a supernova explosion at half the age of the universe.

Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. (2017). Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *ApJ*, *850*(1), Article L7, L7. https://doi.org/10.3847/2041-8213/aa9704

Philcox, O. H. E., Ivanov, M. M., Simonović, M., & Zaldarriaga, M. (2020). Combining full-shape and BAO analyses of galaxy power spectra: a 1.6% CMB-independent constraint on H_0. *J. Cosmology Astropart. Phys.*, *2020*(5), Article 032, 032. https://doi.org/10.1088/1475-7516/2020/05/032

Phillips, M. M. (1993). The Absolute Magnitudes of Type IA Supernovae. *ApJ*, *413*, L105. https://doi.org/10.1086/186970

Phung, V. H., & Rhee, E. J. (2018). A Deep Learning Approach for Classification of Cloud Image Patches on Small Datasets. *Journal of Information and Communication Convergence Engineering*, *16*(3), 173–178.

Pierel, J. D. R., & Rodney, S. (2019). Turning Gravitationally Lensed Supernovae into Cosmological Probes. *ApJ*, *876*(2), Article 107, 107. https://doi.org/10.3847/1538-4357/ab164a

Pierel, J. D. R., Rodney, S., Vernardos, G., Oguri, M., Kessler, R., & Anguita, T. (2020). Projected Cosmological Constraints from Strongly Lensed Supernovae with the Roman Space Telescope. *arXiv e-prints*, Article arXiv:2010.12399, arXiv:2010.12399.

Pierpaoli, E., Scott, D., & White, M. (2000). How Flat is the Universe? *Science*, *287*, 2171. https://doi.org/10.1126/science.287.5461.2171

Pietrzyński, G., Graczyk, D., Gallenne, A., Gieren, W., Thompson, I. B., Pilecki, B., Karczmarek, P., Górski, M., Suchomska, K., Taormina, M., Zgirski, B., Wielgórski, P., Kołaczkowski, Z., Konorski, P., Villanova, S., Nardetto, N., Kervella, P., Bresolin, F., Kudritzki, R. P., . . . Narloch, W. (2019). A distance to the Large Magellanic Cloud that is precise to one per cent. *Nature*, *567*(7747), 200–203. https://doi.org/10.1038/s41586-019-0999-4

Planck Collaboration, Adam, R., Ade, P. A. R., Aghanim, N., Akrami, Y., Alves, M. I. R., Argüeso, F., Arnaud, M., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Bartolo, N., Basak, S., Battaglia, P., . . . Zonca, A. (2016). Planck 2015 results. I. Overview of products and scientific results. *A&A*, *594*, Article A1, A1. https://doi.org/10.1051/0004-6361/201527101

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barand ela, F., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., Battaner, E., Benabed, K., Benoît, A., Benoit-Lévy, A., Bernard, J. .-P., Bersanelli, M., Bielewicz, P., . . . Zonca, A. (2014). Planck 2013 results. XVI. Cosmological parameters. *A&A*, *571*, Article A16, A16. https://doi.org/10.1051/0004-6361/201321591

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., & et al. (2016). Planck 2015 results. XIII. Cosmological parameters. *A&A*, *594*, Article A13, A13. https://doi.org/10.1051/0004-6361/201525830

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. .-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., . . . Zonca, A. (2018). Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, Article arXiv:1807.06209, arXiv:1807.06209.

Planck Collaboration, Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., . . . Zonca, A. (2018). Planck 2018 results. I. Overview and the cosmological legacy of Planck. *ArXiv e-prints*.

Poulin, V., Smith, T. L., Grin, D., Karwal, T., & Kamionkowski, M. (2018). Cosmological implications of ultralight axionlike fields. *Physical Review D*, *98*(8), Article 083525, 083525. https://doi.org/10.1103/PhysRevD.98.083525

Poulin, V., Smith, T. L., Karwal, T., & Kamionkowski, M. (2019). Early Dark Energy can Resolve the Hubble Tension. *Phys. Rev. Lett.*, *122*(22), Article 221301, 221301. https://doi.org/10.1103/PhysRevLett.122.221301

Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. (2021). Machine learning galaxy properties from 21 cm lightcones: impact of network architectures and signal contamination. *ArXiv e-prints*.

Press, W. H., & Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *187*, 425–438. https://doi.org/10.1086/152650

Quimby, R. M., Oguri, M., More, A., More, S., Moriya, T. J., Werner, M. C., Tanaka, M., Folatelli, G., Bersten, M. C., Maeda, K., & Nomoto, K. (2014). Detection of the Gravitational Lens Magnifying a Type Ia Supernova. *Science*, *344*(6182), 396–399. https://doi.org/10.1126/science.1250903

Racca, G. D., Laureijs, R., Stagnaro, L., Salvignol, J.-C., Lorenzo Alvarez, J., Saavedra Criado, G., Gaspar Venancio, L., Short, A., Strada, P., Bönke, T., Colombo, C., Calvi, A., Maiorano, E., Piersanti, O., Prezelus, S., Rosato, P., Pinel, J., Rozemeijer, H., Lesna, V., . . . Schmidt, M. (2016). The Euclid mission design. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, *9904*, Article 99040O, 99040O. https://doi.org/10.1117/12.2230762

Ravanbakhsh, S., Oliva, J., Fromenteau, S., Price, L. C., Ho, S., Schneider, J., & Poczos, B. (2017). Estimating Cosmological Parameters from the Dark Matter Distribution. *ArXiv e-prints*.

Refsdal, S. (1964). On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect. *MNRAS*, *128*, 307. https://doi.org/10.1093/mnras/128.4.307

Reid, D. D., Kittell, D. W., Arsznov, E. E., & Thompson, G. B. (2002). The picture of our universe: A view from modern cosmology. *arXiv e-prints*, Article astro-ph/0209504, astro–ph/0209504.

Reid, M. J., Pesce, D. W., & Riess, A. G. (2019). An Improved Distance to NGC 4258 and Its Implications for the Hubble Constant. *ApJ*, *886*(2), Article L27, L27. https://doi.org/10.3847/2041-8213/ab552d

Rest, A., Scolnic, D., Foley, R. J., Huber, M. E., Chornock, R., Narayan, G., Tonry, J. L., Berger, E., Soderberg, A. M., Stubbs, C. W., Riess, A., Kirshner, R. P., Smartt, S. J., Schlafly, E., Rodney, S., Botticella, M. T., Brout, D., Challis, P., Czekala, I., . . . Waters, C. (2014). Cosmological Constraints from Measurements of Type Ia Supernovae Discovered during the First 1.5 yr of the Pan-STARRS1 Survey. *ApJ*, *795*(1), Article 44, 44. https://doi.org/10.1088/0004-637X/795/1/44

Riess, A. G., Casertano, S., Yuan, W., Macri, L., Bucciarelli, B., Lattanzi, M. G., MacKenty, J. W., Bowers, J. B., Zheng, W., Filippenko, A. V., Huang, C., & Anderson, R. I. (2018). Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant.

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. *ApJ*, *876*, Article 85, 85. https://doi.org/10.3847/1538-4357/ab1422

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., & Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant.

Riess, A. G., Casertano, S., Yuan, W., Bowers, J. B., Macri, L., Zinn, J. C., & Scolnic, D. (2021). Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM. *ApJ*, *908*(1), Article L6, L6. https://doi.org/10.3847/2041-8213/abdbaf

Riess, A. G., Casertano, S., Yuan, W., Macri, L., Anderson, J., MacKenty, J. W., Bowers, J. B., Clubb, K. I., Filippenko, A. V., Jones, D. O., & Tucker, B. E. (2018). New Parallaxes of Galactic Cepheids from Spatially Scanning the Hubble Space Telescope: Implications for the Hubble Constant. *ApJ*, *855*(2), Article 136, 136. https://doi.org/10.3847/1538-4357/aaadb7

Riess, A. G., Press, W. H., & Kirshner, R. P. (1996). A Precise Distance Indicator: Type IA Supernova Multicolor Light-Curve Shapes. *ApJ*, *473*, 88. https://doi.org/10.1086/178129

Rigault, M., Aldering, G., Kowalski, M., Copin, Y., Antilogus, P., Aragon, C., Bailey, S., Baltay, C., Baugh, D., Bongard, S., Boone, K., Buton, C., Chen, J., Chotard, N., Fakhouri, H. K., Feindt, U., Fagrelius, P., Fleury, M., Fouchez, D., ... Weaver, B. A. (2015). Confirmation of a Star Formation Bias in Type Ia Supernova Distances and its Effect on the Measurement of the Hubble Constant. *ApJ*, *802*(1), Article 20, 20. https://doi.org/10.1088/0004-637X/802/1/20

Rines, K., Geller, M. J., Kurtz, M. J., & Diaferio, A. (2003). CAIRNS: The Cluster and Infall Region Nearby Survey. I. Redshifts and Mass Profiles. *AJ*, *126*(5), 2152–2170. https://doi.org/10.1086/378599

Risaliti, G., & Lusso, E. (2018). Cosmological constraints from the Hubble diagram of quasars at high redshifts. *arXiv e-prints*, Article arXiv:1811.02590, arXiv:1811.02590.

Robertson, H. P. (1935). Kinematics and World-Structure. *ApJ*, *82*, 284. https://doi.org/10.1086/143681

Robertson, H. P. (1936a). Kinematics and World-Structure II. *ApJ*, *83*, 187. https://doi.org/10.1086/143716

Robertson, H. P. (1936b). Kinematics and World-Structure III. *ApJ*, *83*, 257. https://doi.org/10.1086/143726

Rodney, S. A., Strolger, L. .-G., Kelly, P. L., Bradač, M., Brammer, G., Filippenko, A. V., Foley, R. J., Graur, O., Hjorth, J., Jha, S. W., McCully, C., Molino, A., Riess, A. G., Schmidt, K. B., Selsing, J., Sharon, K., Treu, T., Weiner, B. J., & Zitrin, A. (2016). SN Refsdal: Photometry and Time Delay Measurements of the First Einstein Cross Supernova. *ApJ*, *820*(1), Article 50, 50. https://doi.org/10.3847/0004-637X/820/1/50

Rodney, S., Brammer, G. B., Pierel, J. D. R., Richard, J., Toft, S., O'Connor, K. F., Akhshik, M., & Whitaker, K. (2021). A Gravitationally Lensed Supernova with an Observable Two-Decade Time Delay. *ArXiv e-prints*.

Rodney, S. A., Patel, B., Scolnic, D., Foley, R. J., Molino, A., Brammer, G., Jauzac, M., Bradač, M., Broadhurst, T., Coe, D., Diego, J. M., Graur, O., Hjorth, J., Hoag, A., Jha, S. W., Johnson, T. L., Kelly, P., Lam, D., McCully, C., ... Zitrin, A. (2015). Illuminating a Dark Lens : A Type Ia Supernova Magnified by the Frontier Fields Galaxy Cluster Abell 2744. *ApJ*, *811*(1), Article 70, 70. https://doi.org/10.1088/0004-637X/811/1/70

Rodney, S. A., Riess, A. G., Strolger, L.-G., Dahlen, T., Graur, O., Casertano, S., Dickinson, M. E., Ferguson, H. C., Garnavich, P., Hayden, B., Jha, S. W., Jones, D. O., Kirshner, R. P., Koekemoer, A. M., McCully, C., Mobasher, B., Patel, B., Weiner, B. J., Cenko, S. B., ... Kocevski, D. (2014). Type Ia Supernova Rate Measurements to Redshift 2.5 from CANDELS: Searching for Prompt Explosions in the Early Universe. *AJ*, *148*(1), Article 13, 13. https://doi.org/10.1088/0004-6256/148/1/13

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.

Rozo, E., Wechsler, R. H., Rykoff, E. S., Annis, J. T., Becker, M. R., Evrard, A. E., Frieman, J. A., Hansen, S. M., Hao, J., Johnston, D. E., Koester, B. P., McKay, T. A., Sheldon, E. S., & Weinberg, D. H. (2010). Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog. *ApJ*, *708*(1), 645–660. https://doi.org/10.1088/0004-637X/708/1/645

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2014). ImageNet Large Scale Visual Recognition Challenge. *arXiv e-prints*, Article arXiv:1409.0575, arXiv:1409.0575.

Rusu, C. E., Wong, K. C., Bonvin, V., Sluse, D., Suyu, S. H., Fassnacht, C. D., Chan, J. H. H., Hilbert, S., Auger, M. W., Sonnenfeld, A., Birrer, S., Courbin, F., Treu, T., Chen, G. C.-F., Halkola, A., Koopmans, L. V. E., Marshall, P. J., & Shajib, A. J. (2019). HoLiCOW XII. Lens mass model of WFI2033-4723 and blind measurement of its time-delay distance and H_0. *arXiv e-prints*.

Rusu, C. E., Fassnacht, C. D., Sluse, D., Hilbert, S., Wong, K. C., Huang, K.-H., Suyu, S. H., Collett, T. E., Marshall, P. J., Treu, T., & Koopmans, L. V. E. (2017). HoLiCOW - III. Quantifying the effect of mass along the line of sight to the gravitational lens HE 0435-1223 through weighted galaxy counts. *MNRAS*, *467*(4), 4220–4242. https://doi.org/10.1093/mnras/stx285

Ryden, B. (2003). *Introduction to cosmology*. Addison-Wesley. https://books.google.nl/books?id=z27vAAAAMAAJ

Saha, P. (2000). Lensing Degeneracies Revisited. *AJ*, *120*(4), 1654–1659. https://doi.org/10.1086/301581

Sako, M., Bassett, B., Becker, A. C., Brown, P. J., Campbell, H., Wolf, R., Cinabro, D., D'Andrea, C. B., Dawson, K. S., De Jongh, F., Depoy, D. L., Dilday, B., Doi, M., Filippenko, A. V., Fischer, J. A., Foley, R. J., Frieman, J. A., Galbany, L., Garnavich, P. M., … Zheng, C. (2018). The Data Release of the Sloan Digital Sky Survey-II Supernova Survey. *PASP*, *130*(988), 064002. https://doi.org/10.1088/1538-3873/aab4e0

Salvati, L., Douspis, M., & Aghanim, N. (2018). Constraints from thermal Sunyaev-Zel'dovich cluster counts and power spectrum combined with CMB. *A&A*, *614*, Article A13, A13. https://doi.org/10.1051/0004-6361/201731990

Sánchez, A. G., Scoccimarro, R., Crocce, M., Grieb, J. N., Salazar-Albornoz, S., Dalla Vecchia, C., Lippich, M., Beutler, F., Brownstein, J. R., Chuang, C.-H., Eisenstein, D. J., Kitaura, F.-S., Olmstead, M. D., Percival, W. J., Prada, F., Rodríguez-Torres, S., Ross, A. J., Samushia, L., Seo, H.-J., … Zhao, G.-B. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the configuration-space clustering wedges. *MNRAS*, *464*(2), 1640–1658. https://doi.org/10.1093/mnras/stw2443

Satpathy, S., Alam, S., Ho, S., White, M., Bahcall, N. A., Beutler, F., Brownstein, J. R., Chuang, C.-H., Eisenstein, D. J., Grieb, J. N., Kitaura, F., Olmstead, M. D., Percival, W. J., Salazar-Albornoz, S., Sánchez, A. G., Seo, H.-J., Thomas, D., Tinker, J. L., & Tojeiro, R. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: on the measurement of growth rate using galaxy correlation functions. *MNRAS*, *469*(2), 1369–1382. https://doi.org/10.1093/mnras/stx883

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. .-P. (2018). Deep convolutional neural networks as strong gravitational lens detectors. *A&A*, *611*, Article A2, A2. https://doi.org/10.1051/0004-6361/201731201

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., . . . Ciardullo, R. (1998). The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type IA Supernovae. *ApJ*, *507*(1), 46–63. https://doi.org/10.1086/306308

Schneider, P., Ehlers, J., & Falco, E. E. (1992). *Gravitational Lenses*. Springer. https://doi.org/10.1007/978-3-662-03758-4

Schneider, P., & Sluse, D. (2013). Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing. *A&A*, *559*, Article A37, A37. https://doi.org/10.1051/0004-6361/201321882

Schuldt, S., Suyu, S. H., Meinhardt, T., Leal-Taixé, L., Cañameras, R., Taubenberger, S., & Halkola, A. (2021). HOLISMOKES. IV. Efficient mass modeling of strong lenses through deep learning. *A&A*, *646*, Article A126, A126. https://doi.org/10.1051/0004-6361/202039574

Scolnic, D., & Kessler, R. (2016). Measuring Type Ia Supernova Populations of Stretch and Color and Predicting Distance Biases. *ApJ*, *822*(2), Article L35, L35. https://doi.org/10.3847/2041-8205/822/2/L35

Scolnic, D. M., Jones, D. O., Rest, A., Pan, Y. C., Chornock, R., Foley, R. J., Huber, M. E., Kessler, R., Narayan, G., Riess, A. G., Rodney, S., Berger, E., Brout, D. J., Challis, P. J., Drout, M., Finkbeiner, D., Lunnan, R., Kirshner, R. P., Sand ers, N. E., . . . Smith, K. W. (2018). The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *ApJ*, *859*(2), Article 101, 101. https://doi.org/10.3847/1538-4357/aab9bb

Shafieloo, A., L'Huillier, B., & Starobinsky, A. A. (2018). Falsifying Λ CDM : Model-independent tests of the concordance model with eBOSS DR14Q and Pantheon. *Phys. Rev. D*, *98*(8), Article 083526, 083526. https://doi.org/10.1103/PhysRevD.98.083526

Shah, P., Lemos, P., & Lahav, O. (2021). A buyer's guide to the Hubble Constant. *arXiv e-prints*, Article arXiv:2109.01161, arXiv:2109.01161.

Shajib, A. J., Birrer, S., Treu, T., Auger, M. W., Agnello, A., Anguita, T., Buckley-Geer, E. J., Chan, J. H. H., Collett, T. E., Courbin, F., Fassnacht, C. D., Frieman, J., Kayo, I., Lemon, C., Lin, H., Marshall, P. J., McMahon, R., More, A., Morgan, N. D., . . . Walker, A. R. (2019). Is every strong lens model unhappy in its own way? Uniform modelling of a sample of 13 quadruply+ imaged quasars. *MNRAS*, *483*, 5649–5671. https://doi.org/10.1093/mnras/sty3397

Shajib, A. J., Treu, T., & Agnello, A. (2018). Improving time-delay cosmography with spatially resolved kinematics. *MNRAS*, *473*(1), 210–226. https://doi.org/10.1093/mnras/stx2302

Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. (2021). Dark matter haloes of massive elliptical galaxies at z ∼ 0.2 are well described by the Navarro-Frenk-White profile. *MNRAS*, *503*(2), 2380–2405. https://doi.org/10.1093/mnras/stab536

Shapiro, I. I. (1964). Fourth Test of General Relativity. *Phys. Rev. Lett.*, *13*(26), 789–791. https://doi.org/10.1103/PhysRevLett.13.789

Sheather, S. J. (2004). Density estimation. *Statistical Science*, *19*(4), 588–597. https://doi.org/10.1214/088342304000000297

Sherwin, B. D., & White, M. (2019). The impact of wrong assumptions in BAO reconstruction. *J. Cosmology Astropart. Phys.*, *2019*(2), Article 027, 027. https://doi.org/10.1088/1475-7516/2019/02/027

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-k., & Woo, W.-c. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *arXiv e-prints*.

Shihavuddin, A., Basu, S., Rexhepaj, E., Delestro, F., Menezes, N., Sigoillot, S. M., Del Nery, E., Selimi, F., Spassky, N., & Genovesio, A. (2017). Smooth 2d manifold extraction from 3d image stack. *Nature Communications*, *8*(1), 1–8.

Silk, J. (1968). Cosmic Black-Body Radiation and Galaxy Formation. *ApJ*, *151*, 459. https://doi.org/10.1086/149449

Sluse, D., Rusu, C. E., Fassnacht, C. D., Sonnenfeld, A., Richard, J., Auger, M. W., Coccato, L., Wong, K. C., Suyu, S. H., Treu, T., Agnello, A., Birrer, S., Bonvin, V., Collett, T., Courbin, F., Hilbert, S., Koopmans, L. V. E., Tihhanova, O., Marshall, P. J., . . . Tarle, G. (2019). HoLiCOW - X. Spectroscopic/imaging survey and galaxy-group identification around the strong gravitational lens system WFI 2033-4723. *MNRAS*, *490*(1), 613–633. https://doi.org/10.1093/mnras/stz2483

Smith, A. F. M., & Roberts, G. O. (1993). *Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods*. Journal of the Royal Statistical Society.

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., . . . Wilkinson, D. T. (1992). Structure in the COBE differential microwave radiometer first-year maps. *ApJ*, *396*, L1–L5. https://doi.org/10.1086/186504

Sonnenfeld, A. (2018). On the choice of lens density profile in time delay cosmography. *MNRAS*, *474*(4), 4648–4659. https://doi.org/10.1093/mnras/stx3105

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., . . . Zhao, F. (2015). Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *ArXiv e-prints*.

Sridhar, S. (2016). *Statistical analysis of the galaxy cluster distribution from next generation cluster surveys* (Doctoral dissertation).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, *15*(1), 1929–1958.

Storm, J., Gieren, W., Fouqué, P., Barnes, T. G., Pietrzyński, G., Nardetto, N., Weber, M., Granzer, T., & Strassmeier, K. G. (2011). Calibrating the Cepheid period-luminosity relation from the infrared surface brightness technique. I. The p-factor, the Milky Way relations, and a universal K-band relation. *A&A*, *534*, Article A94, A94. https://doi.org/10.1051/0004-6361/201117155

Strauss, M. A., Weinberg, D. H., Lupton, R. H., Narayanan, V. K., Annis, J., Bernardi, M., Blanton, M., Burles, S., Connolly, A. J., Dalcanton, J., Doi, M., Eisenstein, D., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Kim, R. S. J., Knapp, G. R., . . . Zehavi, I. (2002). Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *AJ*, *124*(3), 1810–1824. https://doi.org/10.1086/342343

Stritzinger, M. D., Phillips, M. M., Boldt, L. N., Burns, C., Campillay, A., Contreras, C., Gonzalez, S., Folatelli, G., Morrell, N., Krzeminski, W., Roth, M., Salgado, F., DePoy, D. L., Hamuy, M., Freedman, W. L., Madore, B. F., Marshall, J. L., Persson, S. E., Rheault, J.-P., . . . Filippenko, A. V. (2011). The Carnegie Supernova Project: Second Photometry Data Release of Low-redshift Type Ia Supernovae. *AJ*, *142*(5), Article 156, 156. https://doi.org/10.1088/0004-6256/142/5/156

Sunyaev, R. A., & Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, *4*, 173.

Sutherland, D. J., Xiong, L., Póczos, B., & Schneider, J. (2012). Kernels on Sample Sets via Nonparametric Divergence Estimates. *ArXiv e-prints*, Article arXiv:1202.0302.

Suyu, S. H., Bonvin, V., Courbin, F., Fassnacht, C. D., Rusu, C. E., Sluse, D., Treu, T., Wong, K. C., Auger, M. W., Ding, X., Hilbert, S., Marshall, P. J., Rumbaugh, N., Sonnenfeld, A., Tewes, M., Tihhonova, O., Agnello, A., Blandford, R. D., Chen, G. C.-F., . . . Spiniello, C. (2017). HoLiCOW - I. H_0 Lenses in COSMOGRAIL's Wellspring: program overview. *MNRAS*, *468*, 2590–2604. https://doi.org/10.1093/mnras/stx483

Suyu, S. H., Marshall, P. J., Auger, M. W., Hilbert, S., Blandford, R. D., Koopmans, L. V. E., Fassnacht, C. D., & Treu, T. (2010). Dissecting the Gravitational lens B1608+656. II. Precision Measurements of the Hubble Constant, Spatial Curvature, and the Dark Energy Equation of State. *ApJ*, *711*(1), 201–221. https://doi.org/10.1088/0004-637X/711/1/201

Suyu, S. H., Treu, T., Hilbert, S., Sonnenfeld, A., Auger, M. W., Blandford, R. D., Collett, T., Courbin, F., Fassnacht, C. D., Koopmans, L. V. E., Marshall, P. J., Meylan, G., Spiniello, C., & Tewes, M. (2014). Cosmology from Gravitational Lens Time Delays and Planck Data. *ApJ*, *788*, Article L35, L35. https://doi.org/10.1088/2041-8205/788/2/L35

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI Conference on Artificial Intelligence*, *4*, 12.

Takey, A. (2014). *The XMM-Newton/SDSS Galaxy Cluster Survey (PhD thesis)* (Doctoral dissertation).

Tananbaum, H., Avni, Y., Branduardi, G., Elvis, M., Fabbiano, G., Feigelson, E., Giacconi, R., Henry, J. P., Pye, J. P., Soltan, A., & Zamorani, G. (1979). X-ray studies of quasars with the Einstein Observatory. *ApJ*, *234*, L9–L13. https://doi.org/10.1086/183100

Tasca, L. A. M., & White, S. D. M. (2011). Quantitative morphology of galaxies from the SDSS. I. Luminosity in bulges and discs. *A&A*, *530*, Article A106, A106. https://doi.org/10.1051/0004-6361/200913625

Taubenberger, S., Suyu, S. H., Komatsu, E., Jee, I., Birrer, S., Bonvin, V., Courbin, F., Rusu, C. E., Shajib, A. J., & Wong, K. C. (2019). The Hubble constant determined through an inverse distance ladder including quasar time delays and Type Ia supernovae. *A&A*, *628*, Article L7, L7. https://doi.org/10.1051/0004-6361/201935980

The Dark Energy Survey Collaboration. (2005). The Dark Energy Survey. *arXiv e-prints*, Article astro-ph/0510346, astro–ph/0510346.

Tihhonova, O., Courbin, F., Harvey, D., Hilbert, S., Rusu, C. E., Fassnacht, C. D., Bonvin, V., Marshall, P. J., Meylan, G., Sluse, D., Suyu, S. H., Treu, T., & Wong, K. C. (2018). HoLiCOW VIII. A weak-lensing measurement of the external convergence in the field of the lensed quasar HE 0435-1223. *MNRAS*, *477*(4), 5657–5669. https://doi.org/10.1093/mnras/sty1040

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. (2008). Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. *ApJ*, *688*(2), 709–728. https://doi.org/10.1086/591439

Tojeiro, R. (2006). Understanding the Cosmic Microwave Background Temperature Power Spectrum.

Tonry, J. L., Stubbs, C. W., Lykke, K. R., Doherty, P., Shivvers, I. S., Burgett, W. S., Chambers, K. C., Hodapp, K. W., Kaiser, N., Kudritzki, R. .-P., Magnier, E. A., Morgan, J. S., Price, P. A., & Wainscoat, R. J. (2012). The Pan-STARRS1 Photometric System. *ApJ*, *750*(2), Article 99, 99. https://doi.org/10.1088/0004-637X/750/2/99

Treu, T., & Marshall, P. J. (2016). Time delay cosmography. *A&A Rev.*, *24*(1), Article 11, 11. https://doi.org/10.1007/s00159-016-0096-8

Tripp, R. (1998). A two-parameter luminosity correction for Type IA supernovae. *A&A*, *331*, 815–820.

Tucker, E., Walker, M. G., Mateo, M., Olszewski, E. W., Geringer-Sameth, A., & Miller, C. J. (2018). Galaxy Cluster Mass Estimates in the Presence of Substructure. *ArXiv e-prints*, Article arXiv:1810.10474.

Umetsu, K. (2020). Cluster-galaxy weak lensing. *A&A Rev.*, *28*(1), Article 7, 7. https://doi.org/10.1007/s00159-020-00129-w

Uria, B., Côté, M.-A., Gregor, K., Murray, I., & Larochelle, H. (2016). Neural Autoregressive Distribution Estimation. *ArXiv e-prints*, Article arXiv:1605.02226.

Vagnozzi, S. (2020). New physics in light of the H_0 tension: An alternative view. *Phys. Rev. D*, *102*(2), Article 023518, 023518. https://doi.org/10.1103/PhysRevD.102.023518

Valcin, D., Bernal, J. L., Jimenez, R., Verde, L., & Wandelt, B. D. (2020). Inferring the age of the universe with globular clusters. *J. Cosmology Astropart. Phys.*, *2020*(12), Article 002, 002. https://doi.org/10.1088/1475-7516/2020/12/002

van de Weygaert, R. (2009). Large Scale Structure of the Universe course, Chapter 4: Linear Perturbation Theory.

Vega-Ferrero, J., Diego, J. M., Miranda, V., & Bernstein, G. M. (2018). The Hubble Constant from SN Refsdal. *ApJ*, *853*(2), Article L31, L31. https://doi.org/10.3847/2041-8213/aaa95f

Velten, H., & Gomes, S. (2020). Is the Hubble diagram of quasars in tension with concordance cosmology? *Phys. Rev. D*, *101*(4), Article 043502, 043502. https://doi.org/10.1103/PhysRevD.101.043502

Verde, L., Bernal, J. L., Heavens, A. F., & Jimenez, R. (2017). The length of the low-redshift standard ruler. *MNRAS*, *467*, 731–736. https://doi.org/10.1093/mnras/stx116

Verde, L., Protopapas, P., & Jimenez, R. (2013). Planck and the local Universe: Quantifying the tension. *Physics of the Dark Universe*, *2*(3), 166–175. https://doi.org/10.1016/j.dark.2013.09.002

Vernardos, G. (2021). Simulating time-varying strong lenses. *ArXiv e-prints*.

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., Ebeling, H., Forman, W. R., Hornstrup, A., Jones, C., Murray, S. S., Nagai, D., Quintana, H., & Voevodkin, A. (2009). Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. *ApJ*, *692*(2), 1060–1074. https://doi.org/10.1088/0004-637X/692/2/1060

Villaescusa-Navarro, F., Hahn, C., Massara, E., Banerjee, A., Delgado, A. M., Ramanah, D. K., Charnock, T., Giusarma, E., Li, Y., Allys, E., Brochard, A., Uhlemann, C., Chiang, C.-T., He, S., Pisani, A., Obuljen, A., Feng, Y., Castorina, E., Contardo, G., … Spergel, D. N. (2020). The Quijote Simulations. *ApJS*, *250*(1), Article 2, 2. https://doi.org/10.3847/1538-4365/ab9d82

Visser, M. (2004). Jerk, snap and the cosmological equation of state. *Classical and Quantum Gravity*, *21*(11), 2603–2615. https://doi.org/10.1088/0264-9381/21/11/006

W Wilson, R., & A Penzias, A. (1967). Isotropy of cosmic background radiation at 4080 megahertz. *156*, 1100–1.

Wagner-Carena, S., Park, J. W., Birrer, S., Marshall, P. J., Roodman, A., Wechsler, R. H., & LSST Dark Energy Science Collaboration. (2021). Hierarchical Inference with Bayesian Neural Networks: An Application to Strong Gravitational Lensing. *ApJ*, *909*(2), Article 187, 187. https://doi.org/10.3847/1538-4357/abdf59

Walker, A. G. (1937). On Milne's theory of world-structure. *Proceedings of the London Mathematical Society*, *42*, 90. https://doi.org/10.1112/plms/s2-42.1.90

Wallisch, B. (2018). *Cosmological Probes of Light Relics* (Doctoral dissertation).

Walsh, D., Carswell, R. F., & Weymann, R. J. (1979). 0957+561 A, B: twin quasistellar objects or gravitational lens? *Nature*, *279*, 381–384. https://doi.org/10.1038/279381a0

Wand, M. P., & Jones, M. C. (1994). *Kernel smoothing*. CRC press.

Wang, Y.-C., Xie, Y.-B., Zhang, T.-J., Huang, H.-C., Zhang, T., & Liu, K. (2020). Likelihood-free Cosmological Constraints with Artificial Neural Networks: An Application on Hubble Parameters and SN Ia. *ArXiv e-prints*, Article arXiv:2005.10628.

Webbink, R. F. (1984). Double white dwarfs as progenitors of R Coronae Borealis stars and type I supernovae. *ApJ*, *277*, 355–360. https://doi.org/10.1086/161701

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

Weinberg, S. (2008). *Cosmology*.

Wheeler, J. A., & Ford, K. W. (1998). Geons, Black Holes, and Quantum Foam: A Life in Physics.

Whelan, J., & Iben, J., Icko. (1973). Binaries and Supernovae of Type I. *ApJ*, *186*, 1007–1014. https://doi.org/10.1086/152565

Wojtak, R., & Agnello, A. (2019). The Hubble-Lemaître constant and sound horizon from low-redshift probes. *MNRAS*, *486*, 5046–5051. https://doi.org/10.1093/mnras/stz1163

Wojtak, R., Old, L., Mamon, G. A., Pearce, F. R., de Carvalho, R., Sifón, C., Gray, M. E., Skibba, R. A., Croton, D., Bamford, S., Gifford, D., von der Linden, A., Muñoz-Cuartas, J. C., Müller, V., Pearson, R. J., Rozo, E., Rykoff, E., Saro, A., Sepp, T., & Tempel, E. (2018). Galaxy Cluster Mass Reconstruction Project - IV. Understanding the effects of imperfect membership on cluster mass estimation. *MNRAS*, *481*(1), 324–340. https://doi.org/10.1093/mnras/sty2257

Wojtak, R., Hjorth, J., & Gall, C. (2019). Magnified or multiply imaged? - Search strategies for gravitationally lensed supernovae in wide-field surveys. *MNRAS*, *487*(3), 3342–3355. https://doi.org/10.1093/mnras/stz1516

Wolf, C., Meisenheimer, K., Rix, H. .-W., Borch, A., Dye, S., & Kleinheinrich, M. (2003). The COMBO-17 survey: Evolution of the galaxy luminosity function from 25 000 galaxies with 0.2< z <1.2. *A&A*, *401*, 73–98. https://doi.org/10.1051/0004-6361:20021513

Wong, K. C., Suyu, S. H., Auger, M. W., Bonvin, V., Courbin, F., Fassnacht, C. D., Halkola, A., Rusu, C. E., Sluse, D., Sonnenfeld, A., Treu, T., Collett, T. E., Hilbert, S., Koopmans, L. V. E., Marshall, P. J., & Rumbaugh, N. (2017). HoLiCOW - IV. Lens mass model of HE 0435-1223 and blind measurement of its time-delay distance for cosmology. *MNRAS*, *465*, 4895–4913. https://doi.org/10.1093/mnras/stw3077

Wong, K. C., Suyu, S. H., Chen, G. C. .-F., Rusu, C. E., Millon, M., Sluse, D., Bonvin, V., Fassnacht, C. D., Taubenberger, S., Auger, M. W., Birrer, S., Chan, J. H. H., Courbin, F., Hilbert, S., Tihhonova, O., Treu, T., Agnello, A., Ding, X., Jee, I., . . . Meylan, G. (2020). HoLiCOW – XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *MNRAS*, *498*(1), 1420–1439. https://doi.org/10.1093/mnras/stz3094

Wu, H.-Y. (2011). *Precision Cosmology with Galaxy Cluster Surveys* (Doctoral dissertation).

Xu, L., & Wang, Y. (2011). Cosmography: Supernovae Union2, Baryon Acoustic Oscillation, observational Hubble data and Gamma ray bursts. *Physics Letters B*, *702*(2), 114–120. https://doi.org/10.1016/j.physletb.2011.06.091

Yan, Z., Mead, A. J., Van Waerbeke, L., Hinshaw, G., & McCarthy, I. G. (2020). Galaxy cluster mass estimation with deep learning and hydrodynamical simulations. *MNRAS*, *499*(3), 3445–3458. https://doi.org/10.1093/mnras/staa3030

Yang, T., Banerjee, A., & Colgáin, E. Ó. (2019). On cosmography and flat ΛCDM tensions at high redshift. *arXiv e-prints*, Article arXiv:1911.01681, arXiv:1911.01681.

Yıldırım, A., Suyu, S. H., & Halkola, A. (2020). Time-delay cosmographic forecasts with strong lensing and JWST stellar kinematics. *MNRAS*, *493*(4), 4783–4807. https://doi.org/10.1093/mnras/staa498

Yuan, W. (2017). *Period-luminosity Relations of Cepheids and Mira variables and their Application to the Extragalactic Distance Scale* (Doctoral dissertation).

Yuan, W., Riess, A. G., Macri, L. M., Casertano, S., & Scolnic, D. M. (2019). Consistent Calibration of the Tip of the Red Giant Branch in the Large Magellanic Cloud on the Hubble Space Telescope Photometric System and a Redetermination of the Hubble Constant. *ApJ*, *886*(1), Article 61, 61. https://doi.org/10.3847/1538-4357/ab4bc9

Zamorani, G., Henry, J. P., Maccacaro, T., Tananbaum, H., Soltan, A., Avni, Y., Liebert, J., Stocke, J., Strittmatter, P. A., Weymann, R. J., Smith, M. G., & Condon, J. J. (1981). X-ray studies of quasars with the Einstein Observatory II. *ApJ*, *245*, 357–374. https://doi.org/10.1086/158815

Zhang, M.-J., Li, H., & Xia, J.-Q. (2017). What do we know about cosmography. *European Physical Journal C*, *77*(7), Article 434, 434. https://doi.org/10.1140/epjc/s10052-017-5005-4

Zhang, X., Wang, Y., Zhang, W., Sun, Y., He, S., Contardo, G., Villaescusa-Navarro, F., & Ho, S. (2019). From Dark Matter to Galaxies with Convolutional Networks. *ArXiv e-prints*.

Zhao, M.-M., Li, Y.-H., Zhang, J.-F., & Zhang, X. (2017). Constraining neutrino mass and extra relativistic degrees of freedom in dynamical dark energy models using Planck 2015 data in combination with low-redshift cosmological probes: basic extensions to ΛCDM cosmology. *MNRAS*, *469*(2), 1713–1724. https://doi.org/10.1093/mnras/stx978